

Identification and involvement of G-quadruplex structures in cancer cell proliferation and stemness

Thesis submitted for the degree of
Doctor of Philosophy (Science)

In

Life science and Biotechnology

By

Ananya Roy

(Registration Number: SLSBT1105319)

Department of Life Science and
Biotechnology

Jadavpur University

India

2023



मुख्य कैम्पस / Main Campus :

93/1, आचार्य प्रफुल्ल चन्द्र रोड, कोलकाता-700 009
93/1, Acharya Prafulla Chandra Road, Kolkata-700 009
फोन / Phone : 2350-7073 (निदेशक / Director)
इपीएबीएक्स / EPABX : 2350-6619/6702/2402/2403, 2303-0000/1111
फैक्स / Fax : 91-33-2350-6790

शतवार्षिकी भवन / Centenary Building :

पी-1/12, सी.आई.टी. स्कीम-VII एम, कोलकाता - 700 054
P-1/12, C.I.T. Scheme VII-M, Kolkata-700054
फोन / Phone : 2355-7434 (निदेशक / Director), 2355-0595 (रजिस्ट्रार / Registrar)
इपीएबीएक्स / EPABX : 2355-9416/9219/9544, 2569-3271, फैक्स / Fax : 91-33-2355-3886

समन्वित शैक्षिक परिसर / Unified Academic Campus :

ब्लॉक-ईएन, प्लॉट नं-80, सेक्टर-V, सॉल्ट लेक सिटी, कोलकाता-700 091
Block-EN, Plot No.-80, Sector-V, Salt Lake City, Kolkata-700 091
फोन / Phone : 2355-7434 (निदेशक / Director)
इपीएबीएक्स / EPABX : 2569-3123/28, फैक्स / Fax : 91-33-2569-3127

संदर्भ सं. / Ref. No. _____

दिनांक / Date : 04/05/23

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled **“Identification and involvement of G-quadruplex structures in cancer cell proliferation and stemness”** Submitted by Smt. **Ananya Roy** who got her name registered on **30th August, 2019** for the award of Ph.D. (Science) Degree of Jadavpur University, is absolutely based upon her own work under the supervision of **Prof. Subhrangsu Chatterjee** and that neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.

Signature of the Supervisor
04/05/23

Prof. SUBHRANGSU CHATTERJEE
Chairman
Dept. of Biophysics
Bose Institute, Kolkata-54

Dedicated to my parents

Mr. Tapan Kumar Roy

&

Mrs. Suktí Roy



Acknowledgement

I would like to express my sincere gratitude to my mentor **Prof. Subhrangsu Chatterjee**, Professor, Department of Biological Sciences (Former Biophysics department), Bose Institute, for his invaluable guidance throughout my PhD tenure. His immense knowledge, scientific visions, motivation, and enthusiasm has guided me throughout my research. He has always been by my side through the peaks and valleys of time. I remember how he provided mental support to me and my family when I lost my father during the third wave of COVID-19. He is the kindest person I have ever met who is so close to God. Prof. Subhrangsu Chatterjee is the best advisor and mentor I could have ever asked for.

I am thankful to **CSIR** for providing me the research fellowship during my Ph.D. tenure. I shall be ever grateful to **Bose Institute** for providing me the chance and platform to work in such a holy place from where stalwart scientists have emerged. I have a special affection for the central instrument facility of Bose institute (**C.I.F.**) where I learned the research basics. I heartily thank the members of the **C.I.F.** and **Department of Biophysics** of Bose Institute India. It would not have been possible for me to work alone without their assistance.

I would like to thank my lab's super-seniors **Dr. Jagannath Jana, Dr. Soma Mondal, Dr. Meghomukta Mukherjee, Dr. Pallabi Sengupta**, and **Dr. Gopa Dhar** for enlightening me with the glimpse of research work construction and execution during my early days of my Ph.D. life. I am indebted to my lab seniors, **Dr. Anindya Dutta and Dr. Nilanjan Banerjee**, and **Dr. Suman Panda**, who are more like my elder brothers, for their immense support and assistance throughout my PhD tenure.

I am extremely grateful and privileged enough to have **Ms. Debopriya Bose, Ms. Laboni Roy** and **Ms. Oishika Chatterjee** as my lab mates. In all these 6 years of my PhD life, we have grown a special bond that I shall treasure all my life. From scientific discussions, weird

hypothetical analyses of some data to conversations on our future, we have spent more than thousand days together. Long hours of biophysical experiments with ‘chai breaks’ in between, design and successful execution of molecular biology experiments was fun, the memorable cell culture contaminations followed by series of fumigations, accidental breaking of DMS gel casting plates and the long-awaited delivery of a new casting plate will not be forgotten. The informative conferences we attended together, the late-night conversations (building connections), Delhi to Kolkata flight miss, adventurous rides, to surviving low oxygen level in Gurudongmar (Sikkim), we have many such memories together to recollect.

Last but not the least, I am thankful to my parents **Mr. Tapan Kumar Roy** and **Mrs. Sukti Roy**, and my brother **Mr. Aritra Roy** without their immense love, support and sacrifice, my boat would not have sailed this far. I wish my father could see her daughter getting a Ph.D. degree. Although he is in heaven now, his blessings have always been and will be there for me. Ph.D. has not only sculpted my vision of science, but also made me a stronger woman with higher endurance to meet the upcoming challenges in life.

Ananya Roy

2023

Abbreviations

ΔH	Change in enthalpy
ΔS	Change in entropy
A/G ladder	Adenine guanine ladder
AML	Acute myeloid leukemia
BRC	BRACO-19
CD	Circular Dichroism
ChIP	Chromatin immunoprecipitation
c-MYC	Cellular homolog of myelocytomatosis
CNBP	Cellular nucleic acid binding protein
CSC	Cancer stem cell
CUR	Curcumin
CYP	Cyclophosphamide
ddH ₂ O	Double-distilled water
DMEM	Dulbecco's modified Eagle's medium
DMS	Dimethyl sulphate
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DPR	Distal promoter region
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-to-mesenchymal transition
EtBr	Ethidium bromide
ES	Embryonic stem cells
FACs	Flow cytometry
FAM	Fluorescein amidite
FBS	Fetal Bovine Serum
FITC	Fluorescein isothiocyanate
FP	Forward primer
G4/GQ	G-quadruplex
gDNA	genomic DNA
Gs	Guanines
hnRNPK	Heterogenous nuclear ribonucleoprotein K
H-bonds	Hydrogen bonds

iM	i-motif
ITC	Isothermal Titration Calorimetry
ITZ	Itraconazole
K _d	Dissociation constant
KP	Potassium phosphate buffer
KP-KCl	Potassium phosphate supplemented with potassium chloride buffer
lincRNA	Long intergenic RNA
lncRNA	Long non-coding RNA
MD	Molecular dynamics
miRNA	Micro-RNA
NA	Nucleic acid
NCL	Nucleolin
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NHE	Nuclear hypersensitive element
NM23-H2	Non-metastatic cell 2 protein
NMR	Nuclear magnetic resonance
NSCLC	Non-small cell lung cancer
NT	Nucleotide
ORF	Open reading frame
PBS	Phosphate buffer saline
PDB	Protein Data Bank
PI	Propidium Iodide
PLB	Passive Lysis Buffer
Pol II	Polymerase II
PPR	Proximal Promoter Region
PQS	Putative G-quadruplex forming sequence
Pu-27	Wild-type G-quadruplex at NHE _{III} element of c-MYC promoter
PVDF	Polyvinylidene fluoride
QGRS	Quadruplex forming G Rich Sequences
qPCR	quantitative PCR
R1mut1	REX1 mutant G quadruplex type 1 of R1WT
R1mut2	REX1 mutant G quadruplex type 2 of R1WT

R1WT	REX1 wild-type G quadruplex
RMSD	Root mean square deviation
RP	Reverse primer
RT	Room Temperature
Sp1	Specificity Protein 1
sqPCR	Semi quantitative PCR
TBS	Tris buffer saline
TF	Transcription factor
T _m	Melting temperature
TMP	TMPyP4
TNBC	Triple-negative breast cancer
TSP	3-(trimethylsilyl)-2, 2', 3, 3'-tetradeuteropropionic acid
TSS	Transcription start site
UTR	Untranslated region
ZEB	Zinc finger E-box-binding homeobox protein
ZFP	Zinc Finger Protein

CONTENTS

I. Abstract	1
--------------------	----------

II. Chapter 1

Insights into the administrative role of G-quadruplex in cancer and the road map for developing G-quadruplex targeted ligands

1.1	Introduction	2
1.2	Structural diversity of G-tetrads	4
1.3	Relevance of G-quadruplex in human genome	7
1.4	G-quadruplex in oncogenic promoters	9
1.5	Experimental proof of GQs inside cells	10
1.6	GQ-protein coordination in genetic and epigenetic regulation	11
1.7	G-quadruplex targeted anti-cancer therapeutics	16
1.8	GQ ligands in the clinical arena	22
1.9	Unresolved issues, future perspectives and research drive	25
1.10	Conclusion	26
1.11	References	27

III. Chapter 2

Identification and characterization of a flexible G-quadruplex based regulatory switch at the distal promoter region of the stemness gene

2.1	Introduction	32
2.2	Results and discussions	36
2.2.1.	Mining of G rich stretches in the gene promoter of REX1 (ZFP42)	36
2.2.2.	Analysis and structural clarification of the promoter REX1 GQ (R1WT)	37

2.2.3.	Molecular modelled structures of R1WT and its mutated conformers	40
2.2.4.	Comparative structural stability of the REX1 GQ sequences	42
2.2.5.	UV-Visible spectroscopic investigation suggested the interaction of TMPyP4 or BRACO-19 with REX1 GQ to be remarkably different	43
2.2.6.	Fluorescence spectroscopic study revealed substantially distinct mode of interaction between R1mut2 and TMPyP4 or BRACO-19	45
2.2.7.	Structural stabilization or destabilization of R1mut2 upon interaction with small molecules (TMP and BRC)	46
2.2.8.	Corroboration of the interactive behavior of TMP or BRC with REX1 GQ	48
2.2.9.	Computational scrutiny of the interaction between REX1 G-quadruplex and TMP or BRC	50
2.2.10.	In vitro existence of REX1 G-quadruplex	52
2.2.11.	Endogenous R1WT sequence serves as a docking platform for multiple gene regulatory proteins	53
2.3	Conclusion	57
2.4	Materials and methods	58
2.5	References	67

IV. Chapter 3

Curcumin mediated oncogene downregulation via targeting G-quadruplex in the Nuclear Hypersensitive Element III1 of c-MYC leading to apoptotic death in TNBC cells

3.1	Introduction	72
3.1.1.	The mechanosensor mechanism controlling GQ/iM dynamicity at the NHE _{III1} element of <i>c-MYC</i> gene promoter	75
3.2	Results and discussions	79
3.2.1.	Significant binding interaction between curcumin (CUR) and c-MYC GQ	79
3.2.2.	Curcumin and Pu-27 interacts via stacking mode of interaction	82
3.2.3.	Enhancement of thermal stability of c-MYC G-quadruplex conformation upon complexation with curcumin	82

3.2.4. Thermodynamic feasibility of curcumin-c-MYC GQ interaction	83
3.2.5. Atomic level investigation of curcumin-c-MYC GQ interaction by Molecular Dynamics (MD) simulation	85
3.2.6. Curcumin transcriptionally downregulates c-MYC in metastatic cancer	85
3.2.7. Curcumin-Pu-27 GQ interaction mediates transcriptional repression of c-MYC	86
3.2.8. Curcumin mediated repression of c-Myc protein in metastatic cancer	88
3.2.9. Alteration of TF recruitment and stabilization of endogenous Pu-27 by curcumin triggers c-MYC downregulation in metastatic cancer	88
3.2.10. In silico investigation of the interaction of NCL with CUR-MYC-GQ Complex	91
3.2.11. Activation of an alternative cell death pathway by curcumin mediated c-MYC downregulation	93
3.3 Conclusion	97
3.4 Materials and methods	98
3.5 References	109

V. Chapter 4

Itraconazole and Cyclophosphamide combinational treatment enhances cancer cell death *via* c-MYC-GQ- ITZ interaction at the NHE_{III1} element in MDA-MB-231

4.1 Introduction	114
4.2 Results and discussions	116
4.2.1. ITZ suppresses cell proliferation and reduces the adherence property of MDA-MB-231 cells	116
4.2.2. ITZ downregulates c-MYC gene expression at the RNA level in MDA-MB-231 cell line	118
4.2.3. Thermodynamically favorable interaction of ITZ with c-MYC G-quadruplex	118

4.2.4. ITZ specifically targets c-MYC G-quadruplex and downregulates its promoter activity	119
4.2.5. ITZ stabilizes endogenous c-MYC GQ by facilitating the recruitment to NCL	120
4.2.6. ITZ and CYP arrest cells at different phases that synergizes to enhance cancer cell death	120
4.2.7. ITZ and CYP combinational treatment enhances cancer cell death	121
4.3 Conclusion	125
4.4 Materials and methods	125
4.5 References	129

VI. Chapter 5

Insights into the role of lncRNAs and miRNAs in cancer progression and metastasis

5.1 Introduction	132
5.1.1 Role of non-coding RNAs in cancer	135
• Long non-coding RNAs (lncRNAs)	135
• microRNAs (miRNAs)	137
5.2 lncRNA-miRNA interaction in regulating gene expression	139
5.3 Potential anti-metastatic lncRNA-based therapeutics	142
5.4 Future perspectives	143
5.5 References	144

VII. Chapter 6

M2 mediated LINC00273 downregulation modulates lncRNA-miRNA interaction to suppresses cancer metastasis

6.1 Introduction	148
6.2 Results and discussions	151

6.2.1.	M2 downregulates LINC00273 in metastatic cancer cell line MDA-MB-231 and A549	151
6.2.2.	M2 mediated downregulation of LINC00273 supresses cancer cell migration and tumor formation capacity of highly metastatic cells	152
6.2.3.	A comparative analysis of the oncogenes at the transcript level in MDA-MB-231 and A549 metastatic cell lines upon M2 treatment	153
6.2.4.	Differential transcriptome scenario of MDA-MB-231 tumorsphere in presence of LINC00273 and M2 treatment	154
6.2.5.	Antimetastatic property of M2 validated in TNBC xenograft nude mice model	156
6.2.6.	Downstream targets and the plausible mechanism of LINC00273 mediated oncogenic regulation	157
6.2	Conclusion	160
6.4	Materials and methods	162
6.5	References	172

VIII. Publications

Identification and involvement of G- quadruplex structures in cancer cell proliferation and stemness

In the myriad of non-canonical DNA secondary structures, the study on G-quadruplexes (GQs) has taken precedence due to their skewed distribution across the human genome. Deregulation of GQ-based regulatory switches accelerate genomic instability, leading to cancer initiation and progression. We have identified a new G-quadruplex (R1WT) in the distal promoter region of the stemness gene REX1, the activation of which adds pluripotent characteristic to the cancer cells. We have structurally characterized R1WT GQ, and with the help of two well-established GQ binder molecules: TMPyP4 and BRACO-19, we have revealed the flexibility of this newly identified G-quadruplex. Our results elucidate the importance of the promoter region of REX1 gene, encompassing R1WT GQ sequence, in regulation of cancer stemness.

For long, MYC gene has been a challenge to the anticancer research community due to its association with uncontrolled cell proliferation and malignancy. The promoter region of *c-MYC* contains a nuclear hypersensitive element (NHE_{III}) with a G-quadruplex motif (*Pu-27*) that controls MYC gene expression. We have studied the mechanism occurring at this region in metastatic cancer cell model with the help of a natural compound curcumin (CUR). Through a series of biophysical, molecular biology experiments, and MD simulation, we reported the stable binding interaction of CUR with *c-MYC*-GQ. Chromatin immunoprecipitation assay helped us reveal how curcumin modulates the transcription factors binding at the NHE_{III} element while stabilizing *Pu-27* GQ which results in the downregulation of *c-MYC* gene expression. We have also shown the stable binding interaction of Itraconazole (ITZ) with *c-MYC* GQ and studied its modulatory effect on the NHE_{III} element that downregulates *c-MYC* expression. ITZ induced *c-MYC* downregulation sensitizes the cells to apoptotic cell death. That's why the combinatorial treatment of metastatic cells with ITZ and cyclophosphamide (CYP); a cytostatic chemotherapy drug resulted in drastic enhancement of cancer cell death. Thus, ITZ was observed to enhance the efficiency of CYP in killing the metastatic cancer cells, thereby shedding light on the potential of (ITZ+CYP) combinational treatment.

Cancer cells seeded with stemness, and high proliferative potential are the most threatening cells in a tumor mass. We further investigated the global effect of an antimetastatic drug M2 and how its specific targeting and downregulation of LINC00273 gene modulates the transcriptomic scenario (mRNAs and miRNAs) in metastatic cell and animal models. By executing a series of cell and molecular biology experiments, we confirmed the pro-metastatic role of LINC00273 and the anti-metastatic effect of the M2 drug in metastatic cancer cell model. Our findings shed light on the crucial role of G-quadruplex as a regulatory switch controlling gene expression and how we can modulate its function *via* small molecules to combat cancer metastasis and stemness.

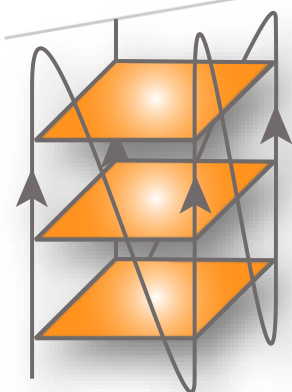
Prof. SUBHRANGSU CHATTERJEE
PROFESSOR
Dept. of Biophysics
Bose Institute, Kolkata-54

Subhrangsu Chatterjee 04/05/23

Guide: Prof. Subhrangsu Chatterjee

Ananya Roy 4/5/23

Candidate: Ms Ananya Roy



Chapter 1

Insights into the administrative role of G-quadruplex in cancer and the road map for developing G-quadruplex targeted ligands

1.1 Introduction

G-quadruplexes (GQs) are not just a mere non-canonical curious structure; it is evidenced to be involved in key genome regulatory functions¹. The history of G-rich sequence adopting higher-order secondary structure dates back to 1910 when the caliber of G-rich DNA to form aggregates in solution was pioneered². The actual structural nature of GQ was discovered in 1962 by Gellert and co-workers², while the crystal structure of GQ was studied much later in the noughties. Especially after the discovery of GQ in the telomere with crystal structure by 2002, this field of research gained enormous attention. The distribution of GQs in the human genome is not random; instead, they inhabit crucial regulatory areas such as telomeres, replication origins, gene promoters, the untranslated regions of nucleic acids, and non-coding RNAs³⁻⁵. According to research, G-quadruplexes (GQs) are important in maintaining cellular balance and stability⁶. It is estimated that there are approximately 376,000 G-rich sequences in the human genome that are capable of forming GQs⁷. However, recent advancements in decoding the human genome have led to an increase in the abundance of identified GQ motifs⁸. Further recent high-resolution sequencing techniques have led to the identification of at least

716000 potential G-quadruplex forming sequences^{8,9}. However, our understanding is currently limited regarding which fraction executes cellular functions¹⁰.

Discoveries have advanced stimulated research probing GQ structure and function. G-quadruplexes are sculpted from 4 strands of nucleic acids by particular DNA and RNA G rich sequences⁵. Four guanine bases of a G-rich form a square co-planar array through Hoogsteen hydrogen bonding, the guanine bases interact with nitrogen atoms N1, N7, and N2 as well as oxygen atom O6^{11,12} wherein each guanine residue may act as a donor as well as an acceptor for 2 H-bonds, resulting in a total of 8 H-bonds between guanines in each GQ quartet plate [Figure 1]. Two, three, or four such quartets stack upon one another *via* π - π interaction, forming the tetrad core of the GQ structure, connected through varying loop length and topology¹³ [Figure 1]. The O6 carbonyl group coordinates the central cation (K⁺/Na⁺/Pb²⁺) through a central ring of H-bonds, enhancing quadruplex stability, while the tetrad plates are stacked on top of each other *via* strong Van der Waal force of attraction¹⁴. The coordination of monovalent or bivalent cations with the eight electronegative O6 atoms of the adjacently stacked G-tetrads is crucial in stabilizing the GQ structure. It effectively neutralizes the electrostatic repulsion of the negatively charged oxygen atom located at the C6 position¹⁵. The K⁺-GQ harmony is well preferred and more biologically relevant because of better coordination and the higher intracellular concentration of K⁺ ions (~140mM) compared to Na⁺ (5-15mM)². The thermodynamic stability of GQ depends on the base composition and the length of the connecting loops, the number of Gs per repeat, and the final topology it procures¹⁶. The predictive typical G4 forming motif follows $G_{\geq 3}-N_{1-7}-G_{\geq 3}-N_{1-7}-G_{\geq 3}-N_{1-7}-G_{\geq 3}$, which has been included in algorithms by various GQ predictor software's¹⁷. Thus, several factors such as base stacking interaction, hydrogen bonding, and electrostatic interaction help in GQ structural stabilization^{6,18}. The hydration of the sugar-phosphate chain also influences the stability of the conformation. Water molecules create an ordered hydration shell by forming a network of H-

bonds, which interact with the bases, sugar molecules, and charged phosphates on the periphery of the GQ. This network of hydrogen bonds and interactions contributes to the overall stability of the structure¹⁹. Still, the complete prediction of GQ conformation is difficult. Structural characterization is necessary for a complete understanding of G-quadruplex biology. Certain guanine-rich sequences in the genome can form G-quadruplexes, crucial in regulating replication, transcription, and gene repair processes. These processes are essential events in building the monumental architecture of the biological system²⁰. However, further exploration is needed to fully understand the regulatory role of G-quadruplexes within the human genome. In-depth atomic-level structural elucidation of GQ shall help us understand G-quadruplex biology and utilize it as a target or weapon to defeat deadly diseases like cancer.

1.2 Structural diversity in G-tetrads

Although GQs are related to the primary nucleotide sequence, they can fold into varying topologies due to strand polarities and interconnecting loop orientation²¹. GQs can be monomolecular or intermolecular, contingent on the involvement of the number of NA strands with wide topological diversity¹. G4s comprise one, two, or four G-rich strands. Tetramolecular GQs engage four G-rich strands to adopt 3D structures with the participating guanines in anti-glycosidic conformation and NA strands in parallel orientation. Intramolecular GQs involve one G-rich strand that conform steadily into a 3D structure with greater complexity and conformational diversity compared to tetramolecular G-quadruplexes. Monomolecular

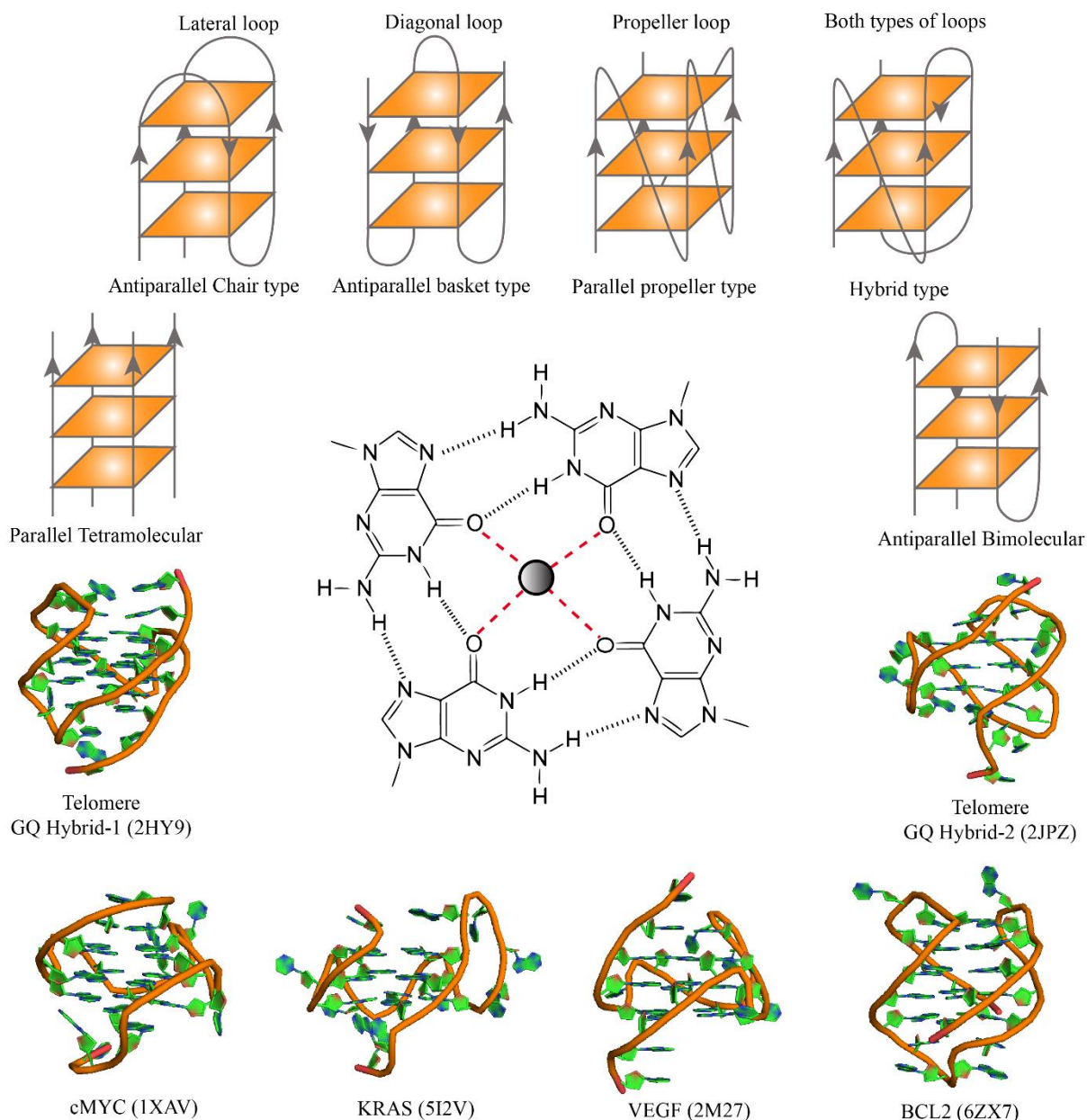


Figure1. Different types of G-quadruplex with variable loop architecture; PDB structures of selected G-quadruplexes. In the middle: atomic arrangement of a single quartet plate comprising eight hydrogen bonds and coordination with the central cation.

scaffolds earned maximum attention due to the relevance of their presence in the biological system¹⁸.

Based on the alignment of the G-tracts, GQs are widely characterized into three topologies: parallel, antiparallel, and hybrid [Figure 1]. In the parallel conformation, all four guanine main chains are parallel to each other, with the sugar in *syn* glycosidic conformation²². In hybrid and

antiparallel topologies, one and two G-stretches orient antiparallely to the remaining G-tracts respectively⁶. Loops that connect the G-quartet plates are basically of three types: propeller type, lateral type, and diagonal type²³. Propeller loops are found in parallel GQs connecting the bottom quartet plate to the top tetrad plate between adjacent strands, while lateral loops are observed in antiparallel GQs where adjacent guanines in the tetrad plate are connected. Diagonal loops evolve, when an antiparallel and a parallel strand, located diagonally in the GQ structure, need to be connected⁶. Diagrammatic representations of the various types of GQs are provided in *Figure 1*. The telomeric DNA is a tandem repeat of d(TTAGGG)_n, yet the TTA loop confers structural polymorphism to the telomeric GQ²². The telomeric sequence conforms into a basket-type GQ in Na⁺ environment, while it forms a parallel GQ in the abundance of K⁺ ions in the crystalline state. A few years later, it was reported that telomeric GQ rather conforms to a hybrid structure that predominates in K⁺ solution²². G-tracts in gene promoters, unlike telomeric GQ are often composed of disparate repeats of Gs and varied number of nucleotides in-between G-tract thereby, increasing the number of possible combinations of quadruplex structures²⁴. Different G-rich sequences can adopt distinct topologies, while there are sequences that possess the flexibility to fold into various different conformations²⁴. The available algorithms are inept at accounting such structural variations, like long loops, bulges, flanking sequences, two-tetrad GQs, and/or possible higher order assemblies²¹. Computational methods resolving these gaps as well as identification of GQs in the chromatin context is incipient.

In double stranded DNA (dsDNA) with putative G-quadruplex forming sequence, when GQ conforms, the complementary DNA strand fails to connect *via* Watson-Crick base pair bonding. The C-rich stretch is either maintained in the single stranded form with the help of interacting ssDNA binding proteins or the C-rich strand folds to form a higher order secondary structure termed intercalated motif (i-motif)^{25,26}. Hemi-protonated C-neutral C base pairs

(C⁺:C) interact *via* intercalation to enhance the stability of the i-motif (iM) at low pH¹. The dynamic evolution and resolution of GQs and iMs is provided in details in the next chapter.

1.3 Relevance of G-quadruplex in the human genome

For long the biological relevance of G-quadruplex remained questionable. Human chromosomal telomeres possess the tandem sequence d(TTAGGG)_n with 50 to 600 nt ssDNA 3' overhangs²². In the year 1987, the importance of G-quadruplex was first identified in the eukaryotic telomeric DNA. The formation of GQs in the human telomere can prevent telomerase function, the enzyme responsible for maintaining genomic integrity and conferring immortality to cancer cells. This discovery highlighted the biological significance of G-quadruplex in the human genome and marked GQ as a primary target in the development of anticancer drug²². The pioneering of the existence of GQ *in vivo* advanced the GQ research field toward understanding the biological functions of GQs. Almost 80% of replication origins harbour putative G-quadruplex sequences (PGQS)¹. GQ structures might act as 'roadblocks' impeding replication fork progression, thereby controlling the replication rate [Figure 2]. Small molecule targeted GQ stabilization has been observed to induce replication-mediated disruption of epigenetic information, indicating the role of GQs as blockades in the path of the replication machinery¹. While contrariwise, the enrichment of PGQS in replication origins suggests GQ's supportive role during DNA replication¹. Genes with higher quantity of GQs on the antisense NA strand are associated with higher Pol II occupancy and higher rate of transcription, suggesting the role of GQ sequences in maintaining DNA in an open state to facilitate active transcription²¹.

Computational scanning identified clusters of PGQS in the promoter region of around 20,000 genes²⁷. The overrepresentation of PGQS in several promoter region of genes is suggestive of a transcriptional repressive function of GQ in maintaining a threshold level of transcripts.

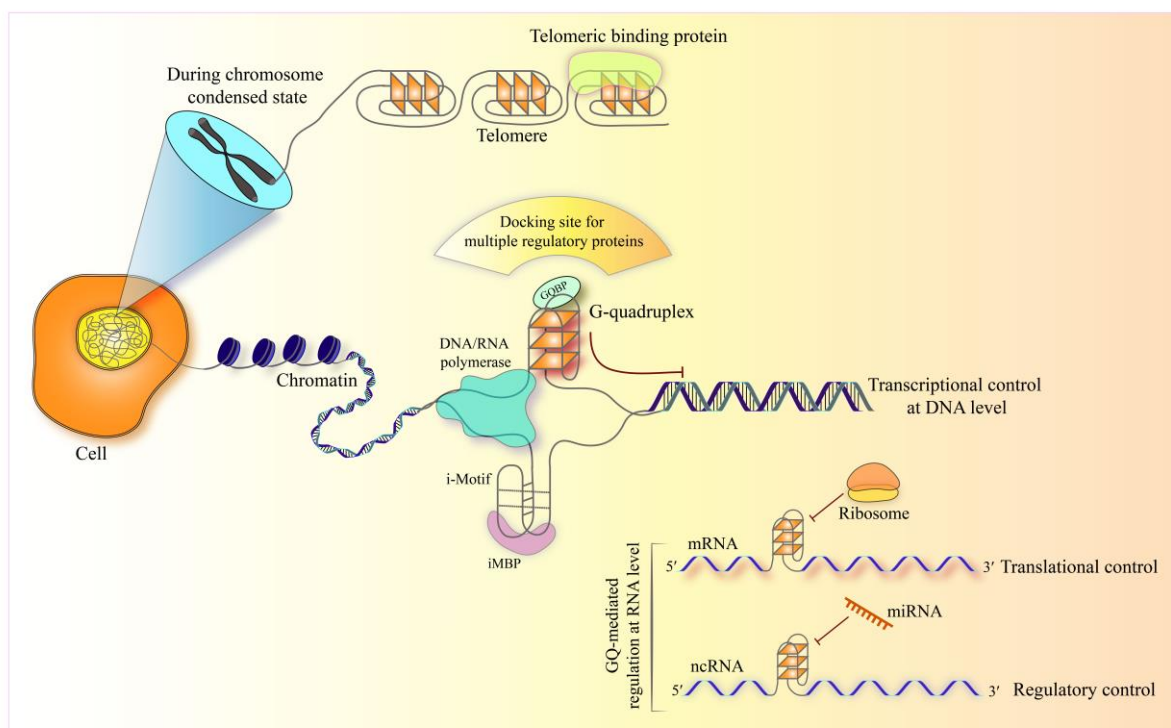


Figure 2. Illustrative depiction of the role of G-quadruplex inside the cell. GQ located at the gene promoter, 5'UTR of RNA acts as a regulatory switch controlling gene expression at the RNA and protein level. GQs at gene promoters have evolved as the recognition elements for various transcription factors, that coordinate to regulate gene expression.

Among the other vital cellular functions, rDNA and ribosome biogenesis are observed to be meticulously controlled by the coordinated functioning of multiple signalling pathways with GQ switching dynamicity⁶. Putative RNA GQs are rich in the 5'UTR, 3'-UTR and intronic regions of pre-mRNAs, mRNAs, and ncRNAs indicating their importance in processing and gene regulation like RNA splicing and polyadenylation, turnover, RNA stability maintenance, and other RNA mediated gene regulation^{6,28}. Discovery of G-rich telomeric repeat RNA (TERRA), Guanine stretches found in the 5'-UTR of NRAS transcript, 3'-UTR of FXR1 and LRP5 mRNAs have been correlated with the involvement of GQ in alternative polyadenylation and mRNA stability, indicating its pivotal role in controlling cellular machinery⁶. There are several reports of GQ present in the 5'-UTR of RNAs to be involved in translation inhibition²¹ [Figure 2]. Loss of the helicase eIF4A to unwind GQs results in reduced translational efficiency indicating the role of G4s in RNAs in influencing ribosome-mediated translation²¹. GQs found

in the 5'-UTR of mRNAs have potential translational regulatory function. The ribosome pauses until a helicase resolves the GQ element, thereby controlling the rate of protein translation. GQ-ribosome interaction is yet to be determined²⁹. *In vivo* probing of cell-type-specific GQ landscape is required for a detailed understanding of the G-quadruplex biology. Recent studies have suggested the role of GQs in regulating higher-order genomic architecture such as chromosome looping, promoter-enhancer interaction, etc. These findings are preliminary and require direct pieces of evidence to determine the role of GQ in higher-order chromatin architecture²¹.

1.4 G-quadruplex in oncogenic promoters

The close association between GQs, DNA breakage, and genomic instability in cancer-associated genes have focused the investigation into the role of GQs in cancer research field with the inspection of GQ targeted ligands as potential therapeutic agents^{20,21}. Promoters of oncogenes have high potential GQ forming motifs²⁴. Upstream elements of oncogenes are reported to have strong correlation with nuclear hypersensitive elements, indicative of the importance of the non-canonical DNA higher order conformations in gene regulation at transcriptional and translational levels and RNA processing as well²⁴. Bioinformatic studies revealed enriched localization of GQ forming NA tracts in promoter region of genes. GQs are experimentally reported in the promoter of human c-MYC³⁰, VEGF³¹, HIF-1 α ³², BCL2³³, KRAS³⁴, c-KIT³⁵, RET³⁶, and PDGF^{2,30} [Table 1]. GQs in the regulatory region of oncogenes vary in number and length of the G tract and the intervening base sequence with the capability of forming multiple possible quadruplex structures and loop isomers². The most common motif observed is an intramolecular parallel stranded GQ structure with 1 or 2 nucleotide double chain reversal loop². However, extensive investigation is required to unveil the detail mechanism of promoter GQ-mediated transcriptional regulation.

List of important G-quadruplex sequences in the genome

Genes	GQ Sequence	Conformation	Ref.
Telomere	GGGTTAGGGTTAGGGTTAGGGTTAGGG	Parallel or anti-parallel or hybrid	22
c-MYC	TGGGGAGGGTGGGGAGGGTGGGGAAGG	Parallel	22,37
VEGF	GGGGCGGGCCGGGGCGGGGTCCCGGCGGGGCGGAG	Parallel	22,31
c-KIT	CCCGGGCGGGCGCGAGGGAGGGGAGG	Variant parallel with snapback	22,38
BCL2Mid	GGGCGCGGGAGGAAGGGGGCGGGA	Mixed parallel-anti-parallel hybrid	22
KRAS	AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG	Parallel	34,39
HIF1α	GGGAGGGAGAGGGGGCGGG	Parallel	22
hTERT	AGGGGAGGGGCTGGGAGGGC	Hybrid	22

Table 1. Reported GQs in the promoter and telomeric region of important genes involved in cancer.

1.5 Experimental proof of GQs inside cells

The *in vivo* existence of GQ was detected for the first time in the telomeres using GQ structure-specific antibody, whereby the telomeric GQ structure was observed to be dynamically controlled by regulatory proteins^{40,41}. Subsequently, research on the *in vivo* existence and regulatory function of GQs got accelerated. In order of a quadruplex to evolve within the chromatin context in the genome, duplex unwinding is needed for ssDNA to be available for self-folding with or without the assistance of proteins inside the cell [Figure 2]. The mechanism superintended by G-quadruplexes engages the recruitment of protein factors that stabilize or destabilize GQ formation and serves as a regulator for the functional transduction of a

molecular event [Figure 2]. Negative torsional stress and molecular crowding favor GQ formation, both of which are crucial for the functionality of the gDNA²¹. High throughput sequencing and chromatin immunoprecipitation with GQ-specific antibodies (1H6 and BG4) have confirmed the existence and location of GQs in the genome²¹. Polymerase stalling experiment and G4-sequencing have identified more than 700000 DNA G-quadruplexes in the human genome¹. G4 ChIP sequencing using BG4 G4 antibody divulged around 10,000 high-confidence GQ elements in cell-lines. More recently a synthetic G4 protein probe (6.7kDa) that binds to G-quadruplexes more strongly has been used to identify GQs in live cells utilizing high throughput DNA sequencing [ChIP-seq] that encountered more than 123,000 G4P peaks⁴². The majority of GQs were found in the regulatory regions, specifically in the nucleosome-depleted chromatin areas¹. BG4 antibody has also been employed in the detection of RNA GQs in human cell cytoplasm²¹. Reverse transcriptase stalling (rGQ-sequencing) on poly(A)-enriched RNAs has mapped RNA GQs in above 3000 mRNAs²¹. Thus, endogenous GQs in live cells are encountered and disclosed using antibody- or small-molecule-based imaging, chemical-mapping, and sequencing techniques²¹. Although the complete repertoire of GQs in cells remains unelucidated, these demonstrations offer a relevant reflection of the biological scenario.

1.6 GQ-protein coordination in genetic and epigenetic regulation

Understanding how G-quadruplex formation takes place inside the cell is fundamental to *in vivo* G-quadruplex biology. Computational predictions and experimental analysis unravelled the enrichment of GQ-motifs in gene promoter, suggesting the functional role of GQs in transcriptional regulation²¹. The binding motifs of several transcription factors have been computed to be enriched with certain endogenous GQ elements, suggesting relevance of protein-GQ interaction²¹. The dynamism of the G-quadruplex must coordinate with different

proteins to regulate gene expression. G-quadruplex interacting proteins identified to date are classified according to their activity of G4s-stabilization or destabilization. Topo1, Nucleolin, thrombin, nucleophosmin proteins are associated with stabilizing GQ structure, while shelterin complex, Pif1, DEAH-box helicase (RHAU/DHX36), Fanconi Anemia Complementation Group J [FANCI], Bloom syndrome protein (BLM), Werner syndrome ATP-dependent helicase (WRN), RTEL1, ATRX, eIF4A, hnRNP are associated with resolving GQ structure. GQN1, Mre11 are reported to degrade GQ structure^{6,15,43}. Inside cells G4s are unfolded by these helicase family proteins that preferentially unwind G4 structure. Deficiencies of WRN-DNA-G4 and BLM helicases, leads to autosomal recessive syndromes¹⁵. BLM mutated cells derived from patients with Bloom syndrome showed high rates of sister chromatid exchange at GQ motif containing sites²¹. Aberrant functioning of WRN and BLM with GQ motifs in their promoter region are reported with altered transcription indicating the regulatory role of GQ in gene transcription¹. POT1 and RPA are 2 constituents of the shelterin complex that bind and unwind telomeric GQ, facilitating the functioning of telomerase¹ [Figure 3A]. Single molecule imaging revealed the GQ unfolding mechanism of certain helicases like DHX36, BLM and WRN with different substrate specificity to displace GQ-stabilizing ligands¹. DHX36, upon interaction with MYC-GQ partially unfolds it, the DHX36-specific motif (DSM) stacks on the top of the partially unfolded GQ *via* a non-polar hydrophobic surface interaction and modulates further downstream activities¹. DHX36 has a positively charged pocket that binds and destabilises the G4^{1,44}. These discoveries suggest that absolute planarity of ligands may not be required, thereby providing us with ideas for designing unique GQ-specific ligands. Overall, we observed strand orientation, loops and grooves, and core GQ structural features to likely

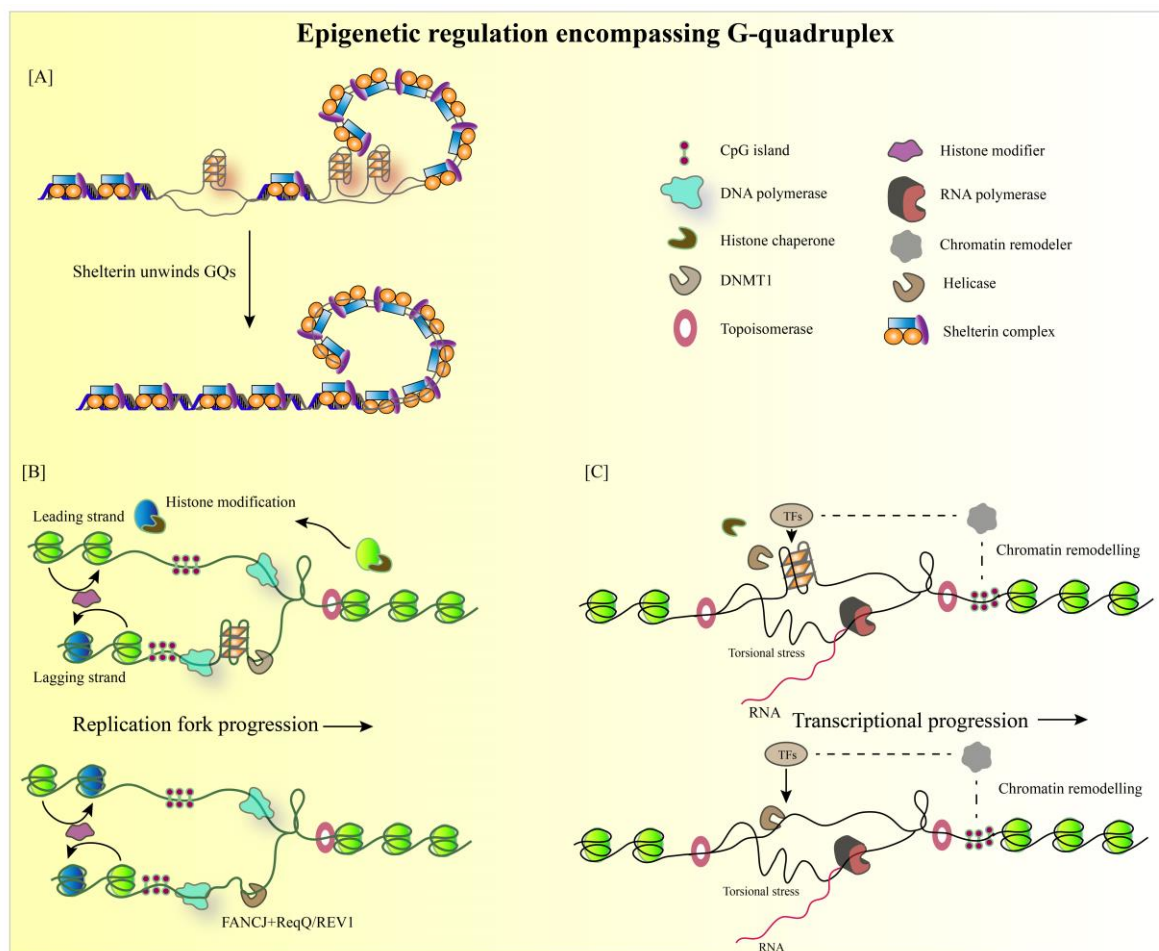


Figure 3. [A] Reduced shelterin binding results in less compact chromatin at the telomeric end which in turn facilitates the evolution of GQs. The binding of shelterin complex unwinds GQ and compacts the telomeric end. [B] During replication, GQs are resolved by helicases for smooth progression of DNA polymerase. GQs participate in histone modification. [C] Helicase mediated GQ unfolding coordinates with TFs and chromatin and histone modulators to regulate transcription of gene.

play vital roles in providing the necessary selectivity and specificity for GQ-GQ-specific binding protein interactions, subsequently leading to specific cellular signaling outcomes²⁹.

Epigenetic and chromatin remodelling enzymes are also identified to interact with DNA GQ sequences²¹ [Figure 3B and 3C]. GQ present in the telomeric lncRNA TERRA and telomeric DNA interact with several proteins to recruit histone methyltransferases and maintain telomere and sub-telomere in heterochromatin state²¹. ATR-X is an example of a chromatin remodelling protein that binds to CpG islands and GC-rich tandem repeats with

G-quadruplex interacting proteins	Target	Function	Ref.
Protection of telomeres 1 (POT1) Replication Protein A (RPA) CTC1-STN1-TEN1 (CST complex)	Telomeric GQ	Unwinds GQ	29,46
Heterogenous nuclear ribonucleoprotein (hnRNP A1)	Telomeric GQ KRAS promoter GQ	Unwinds GQ	29,47,48
RecQ-like helicase Bloom protein (BLM)	Telomeric GQ and replication origin	Unwinds GQ in 3' to 5' direction	29,49
RecQ-like helicase Werner's syndrome protein (WRN)	Telomeric GQ and replication origin	Unwinds GQ in 3' to 5' direction	29,50
Preimplantation factor-1 (Pif1)	Telomeric GQ and other GQs	Unwinds GQ in 5' to 3' direction	29
DEAH box protein 36 (DHX36) helicase	Telomere GQ c-MYC promoter GQ DNA and RNA GQs	Unwinds GQ in ATP independent manner.	29,51
D-E-A-H box protein 9 (DHX9)	Multiple targets	Preferentially resolves RNA G-quadruplexes (rG4s)	52
Non-metastatic factor (NM23-H2); A nucleoside diphosphate kinase	c-MYC promoter GQ	Unwinds GQ	25,29,53
Nucleolin phosphoprotein (NCL)	c-MYC promoter GQ Other GQ promoters bcl-2 VEGF, hTERT, PDGF-A, c-KIT, and RET.	Stabilizes GQ	25,29
MAZ Zinc finger protein	Promoter GQs of c-MYC, HRAS, KRAS.	Unfolds GQ	29

Kruppel like family: Specificity protein (Sp1) with zinc finger motif	Linear DNA, Promoter GQs of c-KIT, HRAS, VEGF	Binds to conformed GQ.	29,54,55
Poly (ADP-ribose) polymerase 1 (PARP1) ZFP	Promoter GQs of c-KIT, c-MYC and KRAS	GQ interaction	29
Fanconi Anemia Complementation Group J (FANCI) helicase	Multiple GQs	Unwinds GQ DNA in 5' to 3' direction	29
G-quartet nuclease 1 (GQN1) human nuclease	Multiple DNA GQs	Cleaves and degrades GQs	29

Table2: List o important G-quadruplex interacting proteins.

putative GQ forming structures¹. With its helicase like activity ATR-X protects the genome that are prone to DNA breakage by unfolding GQs at those regions²¹. Loss of ATR-X promotes GQ formation and is correlated with GQ-dependent replication stress and DNA damage^{1,21}. DNMTs responsible for CpG island methylation are reported to lose enzyme activity upon binding to G-quadruplex. This suggests the contribution of GQ in DNA hypomethylation at the gene promoters by sequestration of DNMTs¹. Conversely, DNA methylation induced GQ formation was reported at hTERT gene promoter CpG island, thereby elevating the level of transcription¹. Thus, GQ structures are associated with both genomic and epigenetic instability^{20,45}. Pif1 helicase is identified to be recruited to G-rich PGQS sites with DNA double strand break, whereby it promotes homologous recombination, indicating GQs as potentials sensors for ROS induced oxidative DNA damage¹. The GQ scenario is dynamic and varies with the functional status of cells where protein factors play an important role in regulating GQ structural dynamics throughout the genome. The biological relevance of protein-GQ interactions and the functional mechanism remains elusive and requires further exploration. A list of protein-GQ interaction is provided in *Table 2*. An *in-depth* study involving the dynamics

of chromatin-associated protein complexes in the understanding of the GQ interactome for the successful development of drug molecules targeting deregulated interactions in diseased cells is elaborated in the subsequent chapters.

1.7 G-quadruplex targeted anti-cancer therapeutics

Ever since the discovery of GQ ligands targeting telomeric GQ and inhibiting telomerase-mediated telomere extension leading to cancer cell death, the design, and development of GQ-targeted ligand molecules have become an attractive area of research in the field of anticancer drug discovery^{21,24}. G-quadruplex structure exhibits an ideal platform for ligand interaction. Among the class of small molecule GQ binders, the majority of them possess aromatic surface available to interact *via* π - π stacking interaction with the quartets, positively charged ligands interact with loops or grooves of the GQs^{1,6}. Ligands can be designed to target the terminal tetrad plates and the G-plates in-between, the grooves and loops, or the central channel of the GQ skeleton⁶ [Figure 4]. The ordered structure of GQ is crucial for the design and development of GQ-interactive compounds. New GQ structures are better understood through GQ-interactive compounds. A similar characteristic among the GQ-targeted drug molecules is a fused-ring system that can interact with the end-tetrad plates, resulting in stacking interactions. The molecular structures of different GQ-targeted ligand molecules are shown in Figure 4. Some of the ligands possess positively charged side chains or basic groups that interact with the trenches in GQ through electrostatic interaction along with the central ring participating in stacking interaction². Positive charges on ligand molecule also increase hydrogen bond formation, thereby strengthening receptor-ligand interaction as well as water solubility⁵⁶. To enhance the binding efficiency of the ligand, the aromatic rings, positive charge, and H-bond donors are needed to be increased for gaining better pharmacokinetic properties¹. Unlike the well-recognized GQ-targeted ligand molecules with symmetric cyclic fused rings showing stacking interaction, we also anticipate a new paradigm ligand with smaller asymmetric

stacking moiety and appropriate functional groups like quindoline derivatives having higher specificity for GQs²². The discovery of novel GQ-specific ligands shall open the realm of targeted cancer therapy at the gene level.

Ligand	Chemical Compound	Target GQs of genes	GQ-Ligand interaction/mechanism of interaction	Ref.
Quarfloxin (CX-3543)	Fluoroquinolone derivative	c-MYC, rDNA-GQ	Quarfloxin was initially designed to target c-MYC promoter GQ. It shows high selectivity for c-MYC GQ. Quarfloxin also targets and binds to the GQ in rDNA and disrupt GQ-NCL interaction, thereby inhibiting ribosome biogenesis and inducing apoptosis in cancer cells.	^{2,12}
Pidnarulex (CX-5461)	Chemically related to quarfloxin (CX-3543)	-	Stabilizes G-quadruplex structure. It blocks replication forks and induces ssDNA strand breaks. Specifically toxic against BRCA1/2 deficient tumors.	^{12,57}
Ligand 12459	2,4,6-triamino-1,3,5-triazine derivative	hTERT GQ	12459 induces the G-rich region of hTERT to form GQ, thereby altering the splicing pattern of hTERT and leading to higher production of inactive transcripts over active transcripts.	⁶

APTO-253	2-indolyl imidazole [4,5-d] phenanthroline derivative	Potent stabilizer of multiple GQs including c-MYC, c-KIT	Downregulates c-MYC, induces G0/G1 cell cycle arrest and apoptosis in AML cancer cells.	35,58
Quindoline	Derivative of natural product cryptolepine	c-MYC GQ	Stabilizes MYC GQ and inhibits c-MYC expression.	22,59
IZCZ-3	A triaryl-substituted imidazole/carbazole conjugate, with two 1-methylpiperazine side chains and a phenyl moiety in the imidazole ring.	c-MYC GQ	1-methylpiperazine increases the electrostatic interaction, while the phenyl moiety reduces its binding affinity with other GQ conformations, thereby enhancing the specificity of the molecule for cMYC GQ. IZCZ-3 stacks on the terminal GQ plate via π - π interaction. IZCZ-3 arrest cancer cells at G0/G1 phase of the cell cycle, thereby, showing its significant anticancer property.	6
FP3	A derivative of furo [2,3-d] pyridazin-4 (5H)-one	BCL2 GQ	Pyridazonone possess anti-inflammatory and anti-cancerous properties. FP3 selectively bind and stabilize BCL2 GQ leading to downregulation of BCL2 gene.	6
MM41	Naphthalenediimide derivative	BCL2, KRAS GQ, Telomeric GQ	The compound's four arms maintain intimate, non-bonded contact with the GQ loops, while the chromophore engages in π - π stacking interaction with the terminal plate of the GQ.	60

IQc:3e	one regioisomers of 5-methyl-indolo[3,2-c] quinoline derivatives with a range of alkyl diamine side chains	KRAS GQ	The IQc:3e molecule has a heteroaromatic surface that interacts with the quartets through π - π stacking. Additionally, the side chains of the molecule form H-bonds, establishing electrostatic interactions with the phosphate backbone and the connecting nucleotides of the GQ structure. This interaction helps in stabilizing the KRAS GQ.	^{6,61}
AQ1	Anthraquinone derivative with phenylalanine derivative.	c-KIT GQ	Both the ligands bind and stabilizes c-KIT GQ	⁶
AQ7	Anthraquinone derivative without phenylalanine derivative.	c-KIT GQ		
SYUIQ-FM05	Quindoline derivatives with 5- <i>N</i> -methyl and 7-fluoro groups.	VEGF GQ	Electron donating alkyl group, Fluorine and <i>N,N</i> -dimethyl propane-1,3-Diamine side chain interacts with VEGF GQ and inhibits transcription.	⁶
BRACO-19	3, 6, 9-tri-substituted acridine 9-[4-(<i>N,N</i> -dimethylamino) phenylamino]-3,6-bis(3-pyrrolidinopropionamido) acridine	Telomeric GQ and also multiple other GQ sequences.	BRACO-19 is a trisubstituted acridine derivative which interacts with the grooves of GQ and the terminal plates. BRACO-19 is reported to target telomere where it destabilizes telomere capping and inhibit telomerase function in cancer.	^{2,35,62}

TMPyP4	5, 10, 15, 20-tetra (N-methyl-4-pyridyl) porphine	Multiple GQs	The fused planar ring system carrying positive charge stacks externally with the G tetrads and reported to stabilize the GQs. It is a telomere targeting agent where it inhibits telomerase in cancer cells. TMPyP4 is also reported to stabilize the silencer GQ element of c-MYC gene promoter, thereby downregulating c-MYC transcription.	2,35
TMPyPz	Tetramethylpyridinium Porphyrazines, non-symmetrical phthalocyanine azo derivatives	Telomeric GQ, but targets other GQs as well	It showed 100-fold higher affinity for telomeric GQ compared to TMPyP4. Further studies are yet to be reported.	63
Se2SAP	5,10,15,20- [tetra (N-methyl-3-pyridyl)]-26,28-di selenasapphyrin chloride (Se2SAP)	c-MYC	Se2SAP binds selectively to c-MYC GQ. It also binds to BCL2 GQ. Effect or role in cell biology is not yet reported in literature.	6
Telomastatin	It is a natural product isolated from Streptomyces anulatus 3533-SV4. Oxazole derivative	Telomeric GQ	It is reported to preferentially interact with intramolecular GQ structures. It is suppress telomerase function, induces telomere shortening and apoptotic death in cancer cells. Anti-proliferative role of Telomastatin is also reported.	2,35

307A	2, 6-pyridine-dicarboxamide derivative	c-MYC	This molecule is effective against both c-MYC and telomeric GQs. It works by stabilizing the GQ structure, which then inhibits c-MYC expression in tumor cells. As a result, it prevents cell proliferation and promotes apoptosis.	²
RHPS4	Five-ring acridine, 3, 11-difluoro-6,8,13-trimethyl-8H-quinol[4,3,2-kl]acridinium methosulfate	Telomeric GQ	It is shown to function by end stacking mode of interaction and induces telomere disfunctioning and inhibits cell proliferation.	²
BOQ1	Cyclic bis-quinacridine (BOQ1), scaffold is mainly composed of two quinacridines connected by polyamine linkers.	-	It has greater quadruplex selectivity than mono-quinacridine compounds.	⁶
Pyridostatin (PDS)	Pyridostatin, 4-(2-Aminoethoxy)-N2, N6-bis[4-(2-aminoethoxy)-2-quinoliny]-2,6 pyridinedicarboxamide	Multi-targeted or specific-target is unclear.	PDS induces telomere abnormality and DNA double strand breaks. PDS downregulate BRCA1 gene transcription. Pyridostatin is specifically toxic to BRCA1/2 deficient tumors.	^{35,64,65}
PhenDC3	Bis-aryl compound derivative	c-MYC GQ and telomeric GQ	Inhibits oncogene transcription and induces telomere disfunctioning	⁶⁵

PIPER	N, N'-Bis (2-piperidinoethyl) perylene-3,4:9,10-bisdicarbimide	Stabilizes telomeric DNA GQ	Due to its flat geometry and electron-rich polyaromaticity, this ligand is one of the most suitable for G4 binding.	^{6,66}
--------------	--	-----------------------------	---	-----------------

Table 3. List of G- quadruplex targeted molecules and their mechanism of action.

1.8 GQ ligands in the clinical arena

CX-3543 (Quarfloxin) small molecule ligand that is thought to target c-MYC promoter GQ *in vitro*. It possesses potent antitumor activity, although the exact mechanism of CX-3543 is not known. There are reports where CX-3543 has been shown to inhibit ribosome biogenesis by targeted stabilization of rDNA G-quadruplexes followed by induction of apoptosis in cancer cells²⁴ [Table 3]. Chemically similar to quarfloxin: **CX-5461**, a DNA double strand break causing agent has entered human clinical trials for HR-deficient breast cancer patients¹. The cytotoxicity of **PDS** drug is found to have a synergistic effect with NU7441, leading to the suppression of DNA-dependent protein kinase. This kinase is a critical factor in the non-homologous end-joining mechanism⁶⁷. **APTO-253** is an example of a small molecule phenanthroline derivative that has entered clinical development for the treatment of AML disease. It acts by downregulating c-MYC followed by G0/G1 phase arrest, upregulates p21, thereby triggering apoptosis in AML cancer cells⁵⁸. **AS1411**, a 26 nt GQ forming oligonucleotide with nuclease resistance has also entered clinical trial. AS1411 is thought to be preferentially up-taken/internalized by cancer cells expressing high level of cell surface nucleolin. Anti-cancerous property of AS1411 depends on its binding to nucleolin protein and affecting its function, thereby inhibiting multiple cancer cell survival pathways²⁴. **MM41** ligand is documented to reduce the transcription of BCL2 and KRAS genes by stabilizing the

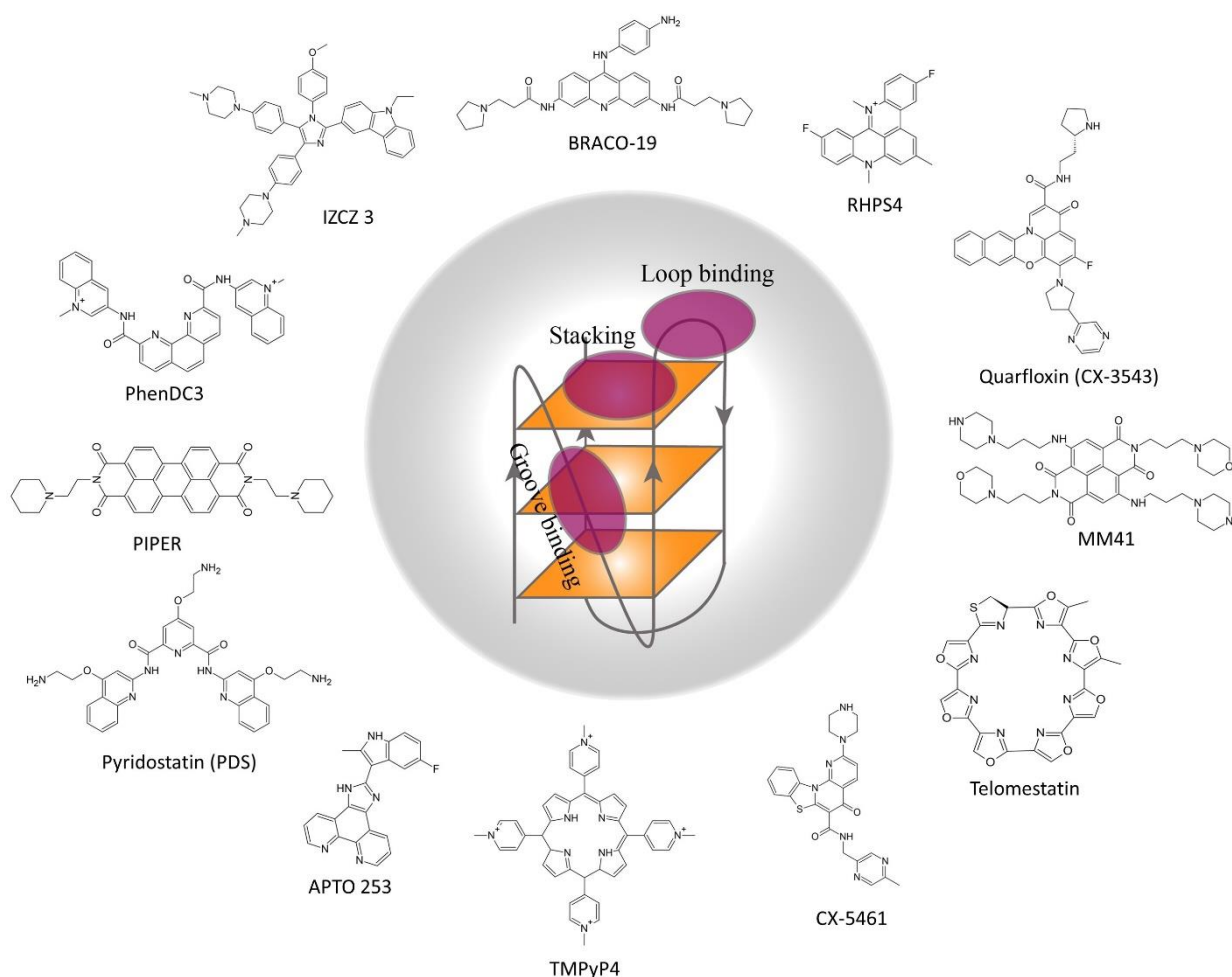


Figure 4: Molecular structures of important G quadruplex ligands

G4 in their promoters. It has a stronger efficacy for the BCL2 gene. **CM03** (2,7-bis(3-morpholinopropyl)-4-((2-(pyrrolidin-1-yl) ethyl) amino) benzo [lmn], phenanthroline-1,3,6,8(2*H*,7*H*)-tetraone) is a derivative of MM41 with increased tolerability and specificity. In pancreatic ductal adenocarcinomas (PDACs), the mentioned genes have high transcriptional activity. Experimental investigation on animal models confirmed the efficacy of CM03 as a tumor suppressing agent⁶⁷. Other telomere G4 stabilizers include **telomestatin** and **BMSG-SH-3**, which are reported to block cell cycle progression and induce apoptosis in glioblastoma, uterus, gastrointestinal, and prostate cancers⁶⁷. **Pyridostatin derivatives (PDS)** targeting GQs represses expression of several major oncogenes. In cancer cells, GQ ligands **RHPS4** and **Pyridostatin** cause dsDNA breaks at GQ sites with defective DNA repair mechanisms, thereby

inducing cell death. BRCA2 deficient cells are sensitive to PDS and RHPS4 treatment. The G4-targeting ligand **CM03**, naphthalene diimide, reduces the expression of G4-rich genes linked to drug resistance pathways, metastasis, and cancer cell survival⁶⁷. The conjugate of imidazole and benzothiazole, **IZTZ-1**, stabilizes the c-MYC promoter GQ and suppresses transcription. Additionally, IZTZ-1 blocks cell cycle progression and induces apoptosis, and hence retards cell proliferation rate in cancer, indicating the possibility of IZTZ-1 to be a promising alternative drug to specifically target c-MYC-GQ⁶⁷. Several molecules have been developed targeting GQ but few of them entered the clinical trial phase. **Quarflorin (CX-3543)** and **CX-5461** are two promising therapeutic candidates for treating several types of tumors. CX-3543 has already reached Phase II clinical trials, while CX-5461 is currently in advanced Phase I clinical trials for patients with BRCA1/2-deficient tumors. Another important ligand triarylpyridine **20A** also showed high binding affinity toward telomere repeats in vitro which further modulates multiple biological altered functions in cancer.

Very few drugs with efficient pharmacokinetic properties have been identified to date. Limitations of these drugs that need to be addressed are lack of specificity, low membrane permeability, small therapeutic window, and cytotoxicity. Ligands with aromatic ring structures providing increased planarity, and strong π - π stacking interaction might have better GQ targeting properties but at the same time, it leads to off-target effects. Certain structural features of GQ ligands like groove binding, and phosphate backbone interaction with reduced planarity can be exploited for increasing the selectivity of the drug. The planarity, polarity, lipophilicity, and rotatable bonds must be stringently considered for optimizing the GQ binding capacity of the ligand, its solubility, and permeability. The enhancement of GQ ligand specificity is still being researched. Incorporation of extra structural motifs such as oligonucleotides, glycosides, and peptides are being tried to solve the issues of recognizing capability, efficiency and solubility.

1.9 Unresolved issues, future perspectives and research drive

The majority of identified GQs are reported to be highly stable, understating GQs as very stiff structures. The flexibility of GQs between folding and unfolding and their capability to transit between different conformations are yet to be explored *in-depth*. The kinetics of quadruplex dynamicity and the persistent length of time of its existence inside cell is another unexplored area of the GQ research field. Alongside, a database of quadruplex structures with folding topology and loop conformation shall be an invaluable asset to GQ researchers²⁴. Commercially available GQ ligands can be repurposed as reference molecules for biochemical, biophysical, and cell biology studies. The clinical potential of new ligands with enhanced selectivity towards specific GQ topology can be studied using these reference GQ ligand molecules. The primary mechanism of how the existing quadruplex anticancer agents function in different cell lines still remain questionable. Is it possible to design a drug molecule that can perceive different GQ topologies remains a question that needs to be addressed. Most GQ ligands have the potential to regulate multiple gene via their interaction with GQs at the nucleic acid level, resulting in polypharmacology¹. Global transcriptional profiling of GQ ligand treated cancer tissues may lead us to the identification of the affected genes and signaling pathways, while suggesting the best suited drug for tissue specific cancer treatment. This may be a way of repurposing the multitargeted drugs in combination therapy with other drugs¹. High-throughput G4 Fluorescent-Intercalator Displacement technique (HT-G4-FID) is a recently documented screening tool to identify new GQ ligand scaffolds as lead compounds for drug development⁶⁸.

The development of targeted drugs can be challenging due to the existence of GQs in various genomic regions. However, if we conduct an in-depth study of the structure of GQs, we can potentially develop a ligand molecule that can distinguish a specific GQ from a pool of GQs with diverse conformations. Since GQ DNA secondary structures come with variable folding

patterns, unique loop conformations, and specific capping structures. With the characterization of unique drug binding pockets in GQs, the design of successful ligand molecule with target specificity can be procured.

1.10 Conclusion

The study of GQ has extended from biophysical-structural studies to *in-vitro* and *in-vivo* studies i.e., in biological models with increased sophistication. Extended research has revealed the link between GQ structure and biological function. G-quadruplex playing critical role in regulating the expression of oncogenes, is a striking anti-cancer drug target. It is an attractive area in the drug discovery research field for the development of GQ targeted ligand. The high polymorphism and flexibility of GQ conformation is the main barrier in the development of specific ligand-molecule that can detect and interact with the selected GQ. *In-depth* investigation is essential to address the structural characterization of GQs and its folding and unfolding mechanism that would help in designing selective GQ binder molecules. The GQ scenario within the cell is dynamic and depends on the working status of the cell, where proteins play an important role in regulating GQ structural dynamics throughout the genome. *In depth* study involving the dynamics of the chromatin architecture and its interaction with protein factors is essential to understand the GQ interactome for successful development of drug molecules targeting deregulated interactions in diseased cells. Structure-based drug design and ligand screening along with atomic-level structural investigation by using bioinformatics and computational biology can fade the limitations that make GQ an elusive target in the successful development of GQ specific ligand. Ongoing research on GQs is revealing abundant information with the incorporation of new techniques. We are very near to finding next generation GQ-targeted ligand with higher therapeutic value to treat cancer.

1.11 References

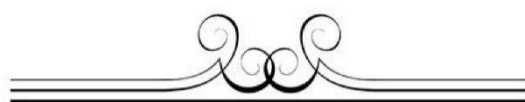
1. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem* **2**, 123–136 (2020).
2. Yang, D. & Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med Chem* **2**, 619–646 (2010).
3. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research* **43**, 8627–8637 (2015).
4. Prioleau, M.-N. G-Quadruplexes and DNA Replication Origins. *Adv Exp Med Biol* **1042**, 273–286 (2017).
5. Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* **40**, 4727–4741 (2012).
6. Banerjee, N., Panda, S. & Chatterjee, S. Frontiers in G-Quadruplex therapeutics in cancer: Selection of small molecules, peptides and aptamers. *Chem Biol Drug Des* **99**, 1–31 (2022).
7. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**, 2908–2916 (2005).
8. Rezzoug, F., Thomas, S. D., Rouchka, E. C. & Miller, D. M. Discovery of a Family of Genomic Sequences Which Interact Specifically with the c-MYC Promoter to Regulate c-MYC Expression. *PLOS ONE* **11**, e0161588 (2016).
9. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential | Nature Reviews Molecular Cell Biology. <https://www.nature.com/articles/nrm.2017.3>.
10. Puig Lombardi, E. *et al.* Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Research* **47**, 6098–6113 (2019).
11. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90**, 1149–1171 (2008).
12. Onel, B., Lin, C. & Yang, D. DNA G-quadruplex and its potential as anticancer drug target. *Sci. China Chem.* **57**, 1605–1614 (2014).
13. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **34**, 5402–5415 (2006).
14. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research* **36**, 5482–5515 (2008).
15. Nakanishi, C. & Seimiya, H. G-quadruplex in cancer biology and drug discovery. *Biochemical and Biophysical Research Communications* **531**, 45–50 (2020).
16. Zyner, K. G. *et al.* G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun* **13**, 142 (2022).

17. G-quadruplexes: a promising target for cancer therapy | Molecular Cancer | Full Text.
<https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-021-01328-4>.
18. Jana, J. & Weisz, K. Thermodynamic Stability of G-Quadruplexes: Impact of Sequence and Environment. *ChemBioChem* **22**, 2848–2856 (2021).
19. Reshetnikov, R. V., Kopylov, A. M. & Golovin, A. V. Classification of G-Quadruplex DNA on the Basis of the Quadruplex Twist Angle and Planarity of G-Quartets. *Acta Naturae* **2**, 72–81 (2010).
20. Linke, R., Limmer, M., Juranek, S. A., Heine, A. & Paeschke, K. The Relevance of G-Quadruplexes for DNA Repair. *Int J Mol Sci* **22**, 12599 (2021).
21. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* **21**, 459–474 (2020).
22. Chen, Y. & Yang, D. Sequence, Stability, Structure of G-Quadruplexes and Their Drug Interactions. *Curr Protoc Nucleic Acid Chem* **CHAPTER**, Unit17.5 (2012).
23. Jana, J. & Weisz, K. Thermodynamic Stability of G-Quadruplexes: Impact of Sequence and Environment. *Chembiochem* **22**, 2848–2856 (2021).
24. Bates, P., Mergny, J.-L. & Yang, D. Quartets in G-major. The First International Meeting on Quadruplex DNA. *EMBO Rep* **8**, 1003–1010 (2007).
25. Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* **138**, 14138–14151 (2016).
26. Sengupta, P., Bose, D. & Chatterjee, S. The Molecular Tête-à-Tête between G-Quadruplexes and the i-motif in the Human Genome. *Chembiochem* **22**, 1517–1537 (2021).
27. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406–413 (2007).
28. Tosoni, E. *et al.* Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res* **43**, 8884–8897 (2015).
29. Meier-Stephenson, V. G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys Rev* **14**, 635–654 (2022).
30. Wang, W. *et al.* Human MYC G-quadruplex: From discovery to a cancer therapeutic target. *Biochim Biophys Acta Rev Cancer* **1874**, 188410 (2020).
31. Agrawal, P., Hatzakis, E., Guo, K., Carver, M. & Yang, D. Solution structure of the major G-quadruplex formed in the human VEGF promoter in K⁺: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res* **41**, 10584–10592 (2013).

32. Chen, H. *et al.* Exploring the Formation and Recognition of an Important G-Quadruplex in a HIF1 α Promoter and Its Transcriptional Inhibition by a Benzo[c]phenanthridine Derivative. *J. Am. Chem. Soc.* **136**, 2583–2591 (2014).
33. Agrawal, P., Lin, C., Mathad, R. I., Carver, M. & Yang, D. The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K⁺ solution. *J Am Chem Soc* **136**, 1750–1753 (2014).
34. Marquevielle, J. *et al.* Structure of two G-quadruplexes in equilibrium in the KRAS promoter. *Nucleic Acids Research* **48**, 9336–9345 (2020).
35. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Molecular Cancer* **20**, 40 (2021).
36. Shin, Y.-J., Kumarasamy, V., Camacho, D. & Sun, D. Involvement of G-quadruplex structures in regulation of human RET gene expression by small molecules in human medullary thyroid carcinoma TT cells. *Oncogene* **34**, 1292–1299 (2015).
37. Mathad, R. I., Hatzakis, E., Dai, J. & Yang, D. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res* **39**, 9023–9033 (2011).
38. Hsu, S.-T. D. *et al.* A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *J. Am. Chem. Soc.* **131**, 13399–13409 (2009).
39. Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* **34**, 2536–2549 (2006).
40. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc Natl Acad Sci U S A* **98**, 8572–8577 (2001).
41. Paeschke, K. *et al.* Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol* **15**, 598–604 (2008).
42. Hou, Y., Guo, Y., Dong, S. & Yang, T. Novel Roles of G-quadruplexes on Enhancers in human chromatin. 2021.07.12.451993 Preprint at <https://doi.org/10.1101/2021.07.12.451993> (2021).
43. Scognamiglio, P. L. *et al.* G-quadruplex DNA recognition by nucleophosmin: new insights from protein dissection. *Biochim Biophys Acta* **1840**, 2050–2059 (2014).
44. Chen, W.-F. *et al.* Molecular Mechanistic Insights into *Drosophila* DHX36-Mediated G-Quadruplex Unfolding: A Structure-Based Model. *Structure* **26**, 403-415.e4 (2018).
45. Schiavone, D. *et al.* Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J* **33**, 2507–2520 (2014).

46. Ray, S., Bandaria, J. N., Qureshi, M. H., Yildiz, A. & Balci, H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc Natl Acad Sci U S A* **111**, 2990–2995 (2014).
47. Krüger, A. C. *et al.* Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level. *Eur Biophys J* **39**, 1343–1350 (2010).
48. Ferino, A. *et al.* hnRNPA1/UP1 Unfolds KRAS G-Quadruplexes and Feeds a Regulatory Axis Controlling Gene Expression. *ACS Omega* **6**, 34092–34106 (2021).
49. Budhathoki, J. B. *et al.* RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res* **42**, 11528–11545 (2014).
50. Tang, W. *et al.* The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum Mol Genet* **25**, 2060–2069 (2016).
51. Antcliff, A., McCullough, L. D. & Tsvetkov, A. S. G-Quadruplexes and the DNA/RNA helicase DHX36 in health, disease, and aging. *Aging (Albany NY)* **13**, 25578–25587 (2021).
52. Murat, P. *et al.* RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biology* **19**, 229 (2018).
53. Sengupta, P., Bhattacharya, A., Sa, G., Das, T. & Chatterjee, S. Truncated G-Quadruplex Isomers Cross-Talk with the Transcription Factors To Maintain Homeostatic Equilibria in c-MYC Transcription. *Biochemistry* **58**, 1975–1991 (2019).
54. Da Ros, S. *et al.* G-Quadruplex Modulation of SP1 Functional Binding Sites at the KIT Proximal Promoter. *Int J Mol Sci* **22**, 329 (2020).
55. Fedeles, B. I. G-quadruplex-forming promoter sequences enable transcriptional activation in response to oxidative stress. *Proceedings of the National Academy of Sciences* **114**, 2788–2790 (2017).
56. Chen, S.-B. *et al.* The role of positive charges on G-quadruplex binding small molecules: learning from bisaryldiketene derivatives. *Biochim Biophys Acta* **1830**, 5006–5013 (2013).
57. Xu, H. *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat Commun* **8**, 14432 (2017).
58. Local, A. *et al.* APTO-253 Stabilizes G-quadruplex DNA, Inhibits MYC Expression, and Induces DNA Damage in Acute Myeloid Leukemia Cells. *Mol Cancer Ther* **17**, 1177–1186 (2018).
59. Dai, J., Carver, M., Hurley, L. H. & Yang, D. Solution Structure of a 2:1 Quindoline–c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **133**, 17673–17680 (2011).
60. Ohnmacht, S. A. *et al.* A G-quadruplex-binding compound showing anti-tumour activity in an in vivo model for pancreatic cancer. *Sci Rep* **5**, 11385 (2015).

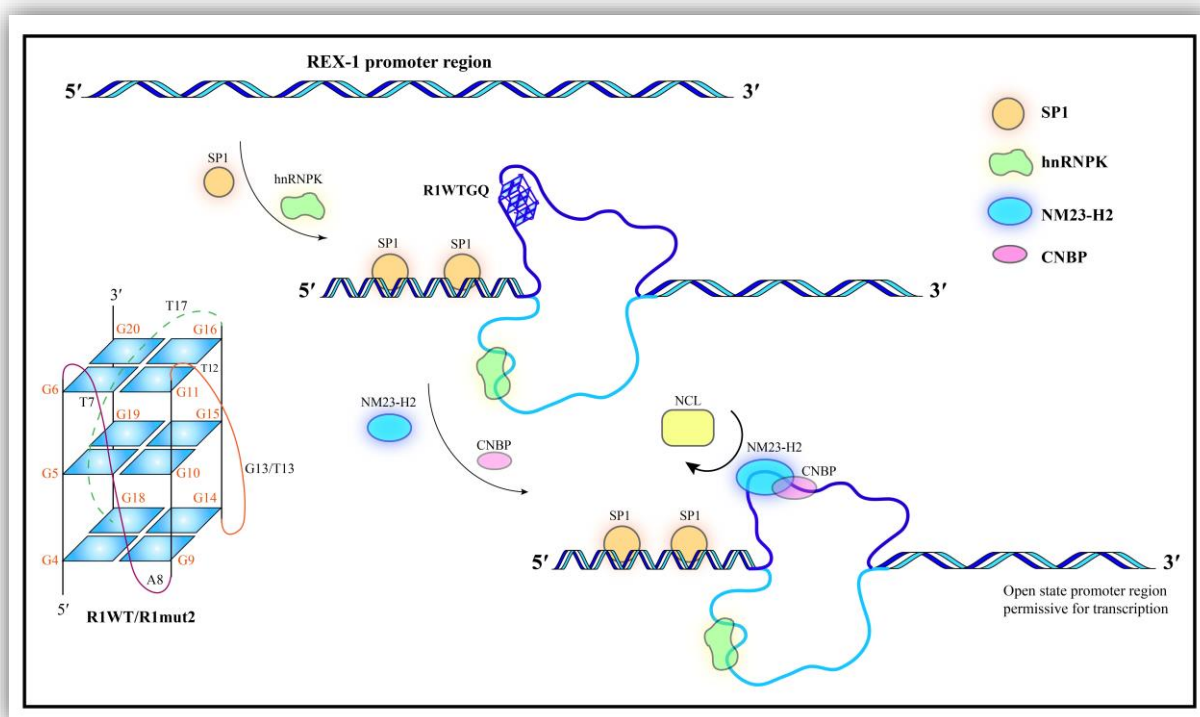
61. Lavrado, J. *et al.* KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines. *Sci Rep* **5**, 9696 (2015).
62. Burger, A. M. *et al.* The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res* **65**, 1489–1496 (2005).
63. Gonçalves, D. P. N., Rodriguez, R., Balasubramanian, S. & Sanders, J. K. M. Tetramethylpyridiniumporphyrazines—a new class of G-quadruplex inducing and stabilising ligands. *Chem. Commun.* 4685–4687 (2006) doi:10.1039/B611731G.
64. Groelly, F. J. *et al.* Anti-tumoural activity of the G-quadruplex ligand pyridostatin against BRCA1/2-deficient tumours. *EMBO Molecular Medicine* **14**, e14501 (2022).
65. Sanchez-Martin, V., Soriano, M. & Garcia-Salcedo, J. A. Quadruplex Ligands in Cancer Therapy. *Cancers (Basel)* **13**, 3156 (2021).
66. Kerwin, S. M., Chen, G., Kern, J. T. & Thomas, P. W. Perylene diimide G-quadruplex DNA binding selectivity is mediated by ligand aggregation. *Bioorg Med Chem Lett* **12**, 447–450 (2002).
67. Awadasseid, A., Ma, X., Wu, Y. & Zhang, W. G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. *Biomedicine & Pharmacotherapy* **139**, 111550 (2021).
68. Alessandrini, I., Recagni, M., Zaffaroni, N. & Folini, M. On the Road to Fight Cancer: The Potential of G-Quadruplex Ligands as Novel Therapeutic Agents. *Int J Mol Sci* **22**, 5947 (2021).



Chapter 2

Identification and characterization of a flexible G-quadruplex based regulatory switch at the distal promoter region of the stemness gene REX1

This chapter is based on the article 'Identification and characterization of a flexible G-Quadruplex in the distal promoter region of stemness gene REX1' *Int J Biol Macromol* 2023 Jan 14; 231: 123263. doi: 10.1016/j.ijbiomac.2023.123263



2.1 Introduction

Preservation of the genomic architecture and cellular homeostasis essentially requires the maintenance and proper functioning of the higher-order NA secondary structures¹. G-quadruplexes (GQs) have drawn a lot of interest amidst the multitude of different nucleic acid topologies due to their skewed distribution across the human genome, which clearly suggests their crucial regulatory function in various domains of cell biology². Cell biology experiments

and advanced techniques have confirmed the evolution of GQ structures and their regulatory functions inside the cells³. Each of the four guanines of a G-rich stretch arranges in a square coplanar orientation through Hoogsteen hydrogen bonding to form the tetrad plates of the G-quadruplex structure³⁻⁵. The G tetrad plates assemble *via* stacking interaction while being connected by loops of different orientations. The Hoogsteen hydrogen bonding network forms a central ring involving the O6 carbonyl group of guanine. The ring coordinates with cations (Na^+/K^+) positioned centrally and imparts greater stability to the quadruplex^{6,7}. The structural stability also hinge upon the total number of Gs present, the architecture of the connecting loops and the overall folding pattern^{1,4}. Several prediction algorithms employ the standardized sequence: $\text{G}_{\geq 3}\text{-N}_{1-7}\text{-G}_{\geq 3}\text{-N}_{1-7}\text{-G}_{\geq 3}\text{-N}_{1-7}\text{-G}_{\geq 3}$ as postulation to identify plausible GQ forming sequences⁸. The regulatory role of GQs is reported in crucial cellular processes like DNA replication, gene transcription, repair pathways and several other processes. Deregulation in the proper folding and dissolution of GQs is correlated with several disease prognosis⁹⁻¹¹. GQ structural disfunctioning is directly linked to transcriptional dysregulation, which result in the genesis and advancement of cancer. As a result, GQs have attracted huge research interest in the field of cancer biology⁸. GQs has evolved as molecular scaffolds that are recognized by specific transcription factors (TFs). The binding proteins and the GQ dynamicity modulate each other's function to coordinately execute a molecular program^{1,12}. A literature survey revealed reports of several GQs and their regulatory roles in the PPR of oncogenes. However, the relevance of GQs in the DPR (Distal Promoter Region) of genes is rarely documented^{12,13}.

Pluripotency is a state acquired by a cell through the joint venture of biomolecules. SOX2, OCT4, NANOG, and REX1 genes are the master controllers of multipotency of a cell¹⁴⁻¹⁶. Cancer cells acquire versatility of multiple differentiation, exhibit chemoresistance, experience cancer recurrence, and metastasize to distant organs when pluripotent genes are activated¹⁷. GQs are endowed with essential genomic structural characteristics that function as an inbound

link between multipotency and cellular differentiation¹. The operative mode of oncogenes and pluripotent stemness genes may be switched on by a defective GQ with faulty functioning at the regulatory region of the gene. GQs may serve as indicators of transcriptionally active zones by altering histone marks and the compaction of nucleosome, according to the relationship between GQs and chromatin structure^{12,13}. It is anticipated that the deregulation of crucial GQs driven by defective cellular functions could promote cellular plasticity in cancer cells. REX1 is a zinc finger protein (also known as ZFP42), expressed predominantly in multipotent cells. There are reports on overexpression of REX1 in ES cells, adult stem cells, and in several types of cancer^{18,19}. The proficiency in metastasis by a few cancer cells may be aided by the deregulated expression of REX1 gene. Literature reviews have reports of the presence of one or more putative GQ forming motifs in genes mostly deregulated in cancer^{8,20}, indicating the significance of the G rich segments identified by us in the REX1 gene promoter. One of the G-rich stretch located distal to the TSS in the REX1 gene was observed to form a highly stable GQ structure, suggesting its putative regulatory role.


In this chapter, we have reported a GQ motif (R1WT) in the DPR of the stemness marker gene REX1. We have found the R1WT GQ to have the predilection for folding into two parallel GQ topologies, R1mut1 and R1mut2. We have employed a series of biophysical and biochemical approach, and molecular dynamics to characterize the wildtype GQ (R1WT) and its monomeric conformers R1mut1 and R1mut2. For selective GQ structure-based drug designing, *in-depth* atomic-level knowledge of the tetrad plates, its loops, and interactivity with drug molecules are the prerequisites^{12,21,22}. TmPyP4 and BRACO-19 are established GQ-binder molecules with preferential binding interaction with parallel GQs^{23–25,26,27}. TmPyP4 is a cyclic ligand and BRACO-19 is a linear ligand²³. We have selected these two small molecules as models in our study to characterize the newly identified GQ. In our study, we have performed a comparative study of TmPyP4-GQ and BRACO-19-GQ interaction. The extent of the structural flexibility

of R1mut2 as obtained from the ligand-GQ interaction studies strongly suggest its functional role as a gene regulatory switch. We also report the physical interaction of selected transcription factors (CNBP, Sp1, hnRNPK, and NM23-H2) with the proximity of R1WT in the chromatin context of metastatic cell. We observed the occupancy level of GQ resolving TFs, NM23-H2, and CNBP to outweigh the binding interaction of R1WT with the GQ stabilizer TF NCL, indicating the maintenance of R1WT in an unfolded state in metastatic cancer cell model. From our results, we speculated a correlation between the relaxed promoter region and the unfolded configuration of R1WT, prime the cancer cells with pluripotent characteristics. R1mut2 was observed to stably interact with BRACO-19, which suggests the possibility of designing a GQ-targeted molecule specific for turning off REX1 gene expression. Successful disruption of stemness traits in the metastatic cells shall drastically reduce the mortality rate in patients diagnosed with advanced-stage cancer. Our investigation presented in this chapter shall accelerate the designing of GQ-conformation-based targeted molecule. Selectively impeding the promoter DNA of REX1 from procuring a relaxed chromatin state may be a smart approach to suppress metastasis and development of secondary malignant tumors. Our research adds to the growing repertoire of the GQ topology toolbox, assist in targeted drug invention programs²², and the construction of nanodevices for medical diagnostic purposes²⁸.

2.2 Results and Discussion

2.2.1 Mining of G rich stretches in the gene promoter of REX1 (ZFP42)

```
>NC_000004.12:187994737-187995871 Homo sapiens chromosome 4,
GRCh38.p12 Primary Assembly
ATCGATCTGGGTTTCGTTCTAATGGTCTGGGTGTACTTATACGACCTATTACATTTCACTAGACT
GTAGCCAATAGTGAGCGTTGACTGACCATTGCTTTCAAAGTGAGTTGTGTTGCCTTTAGCAATAC
AGTCACATTAATGGCCAGAAAAAGAAAAAGAAAAACAAACACTGGGGGTGTTTGAAACAACTGGG
GGTAGGGTGGGGTGGGAAAGGTGTGGGGGTTTGCCTTCGAGTCAACCTCTATTTTCGGGGGGTAAA
AACTGCCTCCTCCGCTTGATTTTTTTTCTCGATCTGCAAAAGCGTAAATGTTTTGTTTTGTTCTA
TTGAGATGGAGCCTCGCTCTATCACCAGGCTGGAGCTTAGTGGCGGATCTCGGCTCACTGCAA
CCTCCGCATCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCG
AGCACCACCATGCCTGGCTAATTTTTGTATTTTTTCGGTAGAGACGGGGTTTCGCCATGTTGGCCA
GGCTGGTCTCGAACGCTTGACCTCATGTGATCCCCCGTCTCGGCCTCCCAAAGTGCTGGGACGA
CAGGCGTGAGCCACCGCGTCCGGCCTAAAAGGGTAAATGTGATTACACCCACGCGTATTTGTTCA
ACAGACATTTATTGAGCGCTCACCACGTGCCAACGCCGGGCGTCTGGGCTCTGGAGGCGCTGCCA
CGTGCGGATGCGCAGTGCCCGGCGGCGGGCTGAGGGGTGAACGCGGGGGTCAGGGGGCCCCGG
GCTGAGGGTTTTCGTGCGGGCCGGGTGCCTGGGGGCCCGGGCTGAGGGTGAGCGCGCGCCAG
GTGCCAGGCGGCTCCGGGCAGAGAGTGAACGCGCGGGCCAGGTGGCTCCAGGGCGGCGCCCCAGG
GCGGGGCGGCGCCACGCCCTCCCTAACCTGGCGGAGCTGATGGGTGGCTGTAGCCTGATTAGACC
GCGTCAGTCCGGAGGGTGGGTCTTGGGAGGGGGCGCAGGGCAGTCCACGTTTCCACTGCAGTTTC
TCCTTTGTTTTACGTTTGGGAGGAGGTGGCATTGGAAATAGCAGAGTGCTTCGCGGTAACAGGGG
TGAGTCTTGTTTCATGGAACTTTTTTCAA
```

Figure 1: Promoter region of REX1 gene downloaded from NCBI PubMed database. The TSS is marked with a symbol () and with highlighted putative GQ forming sequences. R1WT highlighted in yellow and R2, R3, and R4 are highlighted in sky blue colour.

We identified four G-tracts in the REX1 gene's promoter with the propensity to form G-quadruplexes [Figure 1]. The promoter sequence of REX1 (also known as ZFP42) was retrieved from NCBI PubMed gene database and verified using the MatInspector Genomatix software suite. Next, we used QGRS mapper software to obtain the GQ scores of the selected G rich sequences [Table 1]. We acquired the 1D-1H-NMR scan and CD spectra of the selected oligos in the GQ signature range. The R1WT sequence highlighted in yellow in Figure 1 is located at the DPR of REX1, while R2, R3 and R4 are situated in the proximal zone of the

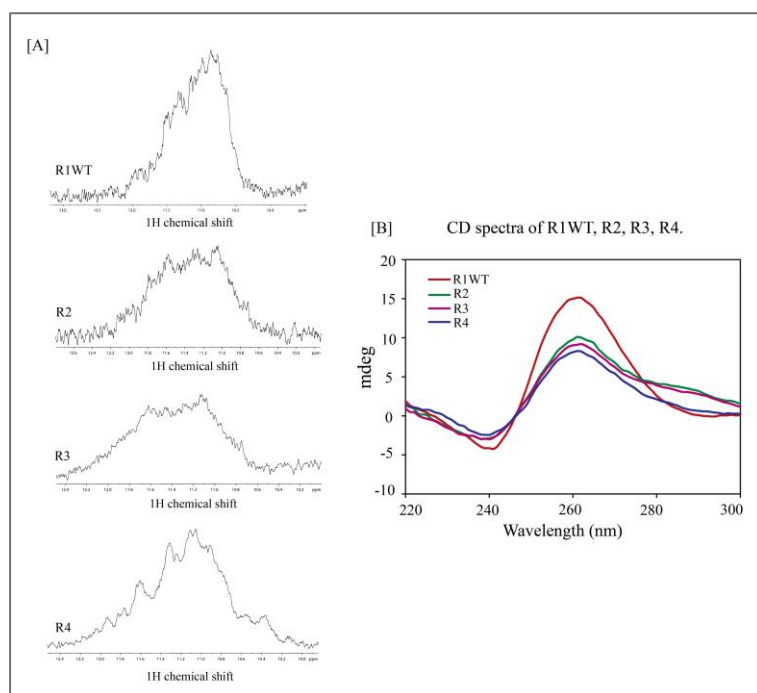


Figure 2: [A] 1D-¹H NMR spectra of the G-rich sequences: R1WT, R2, R3, and R4 in 10 to 12 ppm range. [B] CD spectra of R1WT, R2, R3, and R4 putative GQ forming sequences.

TSS. R1WT displayed much clear imino proton signal (in 10-12 ppm range) featuring the property of G-quadruplex compared to R2, R3, and R4 [Figure 2A and 2B]. R1WT also displayed sharper signature spectra of parallel GQ in buffer supplemented with K⁺ ions compared to the other sequences

under study. We detected an intense positive peak at 262 nm

and a negative peak at 240 nm by R1WT in the CD spectroscopic scan. [Figure 3C-(i)]. R1WT do not form GQ in Tris-Cl buffer. Next, we conducted stability assay in CD spectrophotometer and determined the temperature dependent stability of the quadruplex [Figure 3C-(ii)]. A melting temperature (T_m) of 73.28°C±2.12°C was obtained for R1WT GQ. According to reports K⁺ ions favor shorter loops in GQs^{4,29}. The stability of the GQ hinge upon the nucleotide base composition of the loop as well as its configuration (T/C>A)^{4,30}. We speculated on the presence of short loops with thymine (T) residues that possibly account for the high stability of the said G-quadruplex.

2.2.2 Analysis and structural clarification of the promoter REX1 GQ (R1WT)

The analysis of the participating guanines (Gs) in the tetrad plate and the nucleotide bases comprising the loops of R1WT structure was studied by conducting DMS footprinting assay. We performed the experiment under two reaction conditions: (i) KP-KCl buffer (pH 7), (ii) Tris-Cl buffer at pH 7. The CD and footprint data confirmed the formation of R1WT in KP-

KCl buffer media and not in Tris-Cl environment [*Figure 3C-(i) and 3D*]. The guanines that do not participate in Hoogsteen base pairing are methylated at the N7 position by dimethyl sulphate (DMS), which are subsequently cleaved by piperidine³¹. Since we speculated that R1WT would conform into a GQ in the presence of potassium ions, the Gs participating in Hoogsteen H-bonds are protected from DMS mediated methylation and thus protected from piperidine cleavage which otherwise cleaves the unprotected Gs in the loops methylated by DMS. We unravelled each of the Gs participating in the tetrad plate and loop formation visualizing the cleavage pattern of the DMS reaction. On comparing the intensities of the resolved nucleotide bands in KP-KCl versus Tris-Cl buffer, it determined G4-6, G9-11, G14, G15, and G18-20 to display the protection pattern as evidenced by the lighter intensities of the bands. Therefore, we confirmed the above-mentioned guanines to participate in quartet plate construction *via* Hoogsteen H bond network. We observed darker bands of the nucleotides (NTs): T1, G2, and G3. Since these 3 NTs are situated at the 5' end of R1WT, these are not involved in the plate or loop of the GQ structure, while T7, A8, T12, T17, and A21 were speculated to build the loops [*Figure 3D*]. We observed equal intensity of G13 and G16 bands that led us to speculate their co-occurrence in tetrad plate or loop with a certain level of dynamicity that allows the formation of two GQ conformation simultaneously *in solution*. Our previously obtained 1D NMR spectra of R1WT with a positive hump instead of resolved peaks and the DMS footprint result are congruent, suggesting the dynamic nature of the wildtype GQ R1WT. We further mutated G13 and G16 and formed 2 sequences [R1mut1 and R1mut2] and determined the structure of the wildtype GQ R1WT. The sequences are referred in *Table 1*. Both the sequences displayed similar characteristics of parallel GQ in the CD spectral scan, while their 1D NMR spectral signals were highly resolved with sharp imino

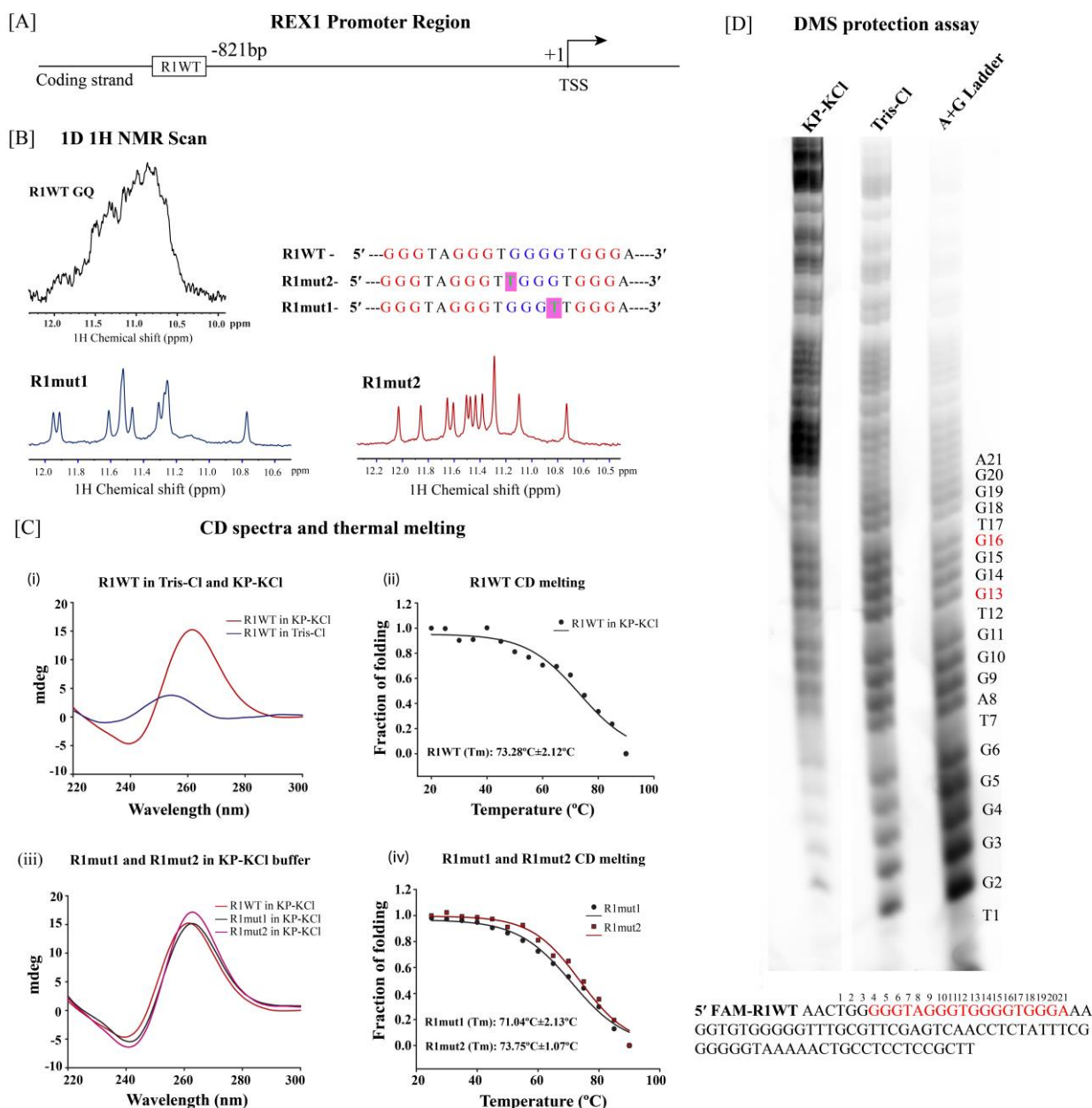


Figure 3 [A] **Representative diagram** of the REX1 gene promoter with the marked TSS and the location of R1WT GQ sequence in the sense strand. [B] **1D NMR spectra** of R1WT, R1mut1 and R1mut2 sequences in 10 to 12 ppm range. The G rich sequences are mentioned with the Gs engaged in the quadruplex are highlighted in red. [C] **CD data** (i) **CD spectral scan** of R1WT in KP-KCl buffer and Tris-Cl buffer. (ii) **CD melting** plot of R1WT in KP-KCl. Fraction of folding is plotted against incremental rise in temperature. (iii) **CD spectral scans** of R1mut1 and R1mut2 overlaid on the CD spectral scan of R1WT. (iv) **Overlaid CD melting profile** of the two mutated sequences. [D] **DMS footprint results** of REX-1 promoter R1WT encompassing region under different conditions. From right to left: A+G ladder, Tris-Cl, KP-KCl. The R1WT comprising sequence tagged at its 5' terminal is represented under the gel image. The NTs engaged in the quartet plate construction are highlighted in red. The lighter guanine bands signify their protection from methylation by DMS thereby, confirming their participation in Hoogsteen hydrogen bonding.

proton peaks [*Figure 3B and 3C-(iii)*]. The scan results clearly demonstrate that each of the mutated sequences adopts a stable monomeric GQ form with minimal dynamicity.

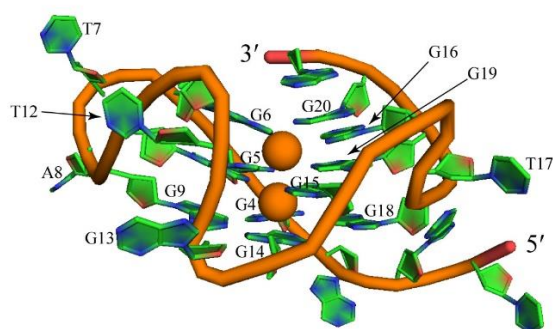
2.2.3 Molecular modelled structures of R1WT and its mutated conformers

We applied computational biology to identify the guanines and categorize each GQ tetrad plate of the expected G-quadruplex structure. We utilized the analyzed DMS footprint data to construct the molecular models of the GQ structures using the Schrödinger Modelling Software Suite and energy minimized in AMBER14. Residue numbers G13 and G14 of R1WT were observed to co-occur in the tetrad as well as the loop, allowing the co-existence of two putative GQ conformers with different loop members. We predicted two structures (R1mut1 and R1mut2), each with a different loop composition based on the information acquired from the analyzed DMS data.

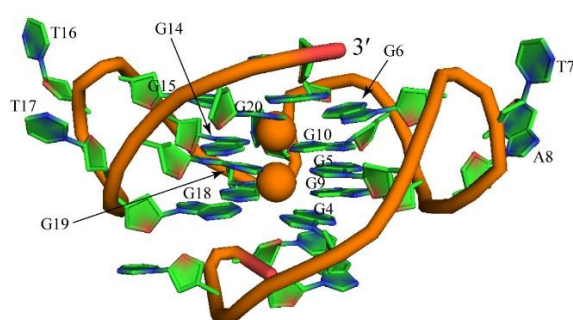
The R1WT was found to have a ‘2-2-1’ loop configuration (loop 1: T7 and A8, loop 2: T12 and G13, loop 3: T17). G4-G9-G14-G18 formed the quartet-1, G5-G10-G15-G19 formed quartet 2, and G6-G11-G16-G20 formed the third quartet [*Figure 4A-(i) and 4B-(i)*]. The model of R1mut2 was built based on the framework of R1WT wildtype topology, by substituting 13th guanine (G13) of the loop with thymine (T13) [*Figure 4A-(iii) and 4B-(i)*]. Molecular modelling results after energy minimization revealed R1mut2 to possess the same configuration of R1WT, both having the same guanines in each tetrad plate with a ‘2-2-1’ loop topology. R1mut2 differed from R1WT only at nucleotide 13th position, T13 substituted for G13 in loop 2. R1mut1 sequence with mutated nucleotide base at the 16th position (G16 to T16) was observed to conform into a GQ with ‘2-1-2’ loop topology (loop 1: T 7 and A 8, loop 2: T12, loop 3: T16 and T17). The sequence of guanines forming the tetrad plates also differed from that of R1WT and R1mut2. The first quartet of R1mut1 was formed by G4-G9-G13-G18,

[A] Molecular modelling

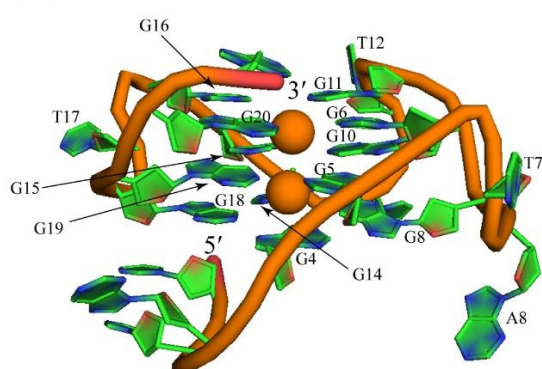
(i) R1WT



(ii) R1mut1

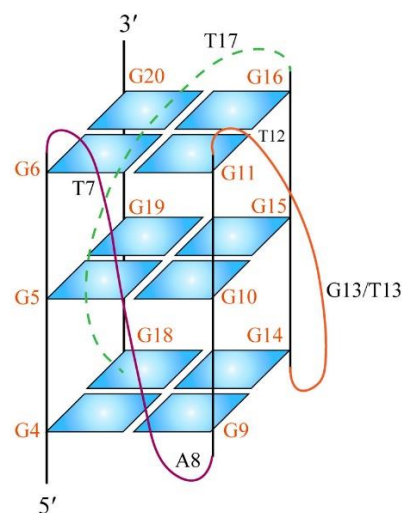


(iii) R1mut2



[B] Cartoon representations

(i) R1WT/R1mut2



(ii) R1mut1

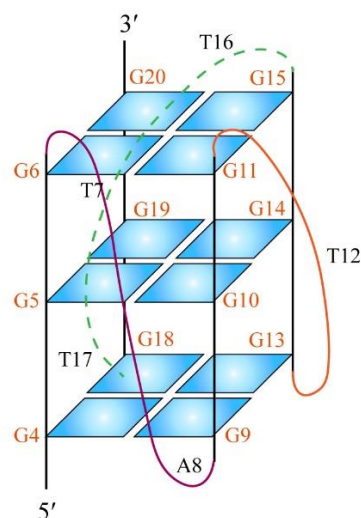


Figure 4: [A] **Molecular modelling results** (i) R1WT, (ii) R1mut1, and (iii) R1mut2 molecular models build using Schrodinger software suite. The nucleotide bases are labelled in the models. [B] **Representative cartoon models** (i) R1WT or R1mut2, and (ii) R1mut1 G-quadruplex structures. Each nucleotide participating in the tetrad plates and connecting loops as determined from the modelling data were used to draw the cartoons.

the 2nd quartet plate by G5-G10-G14-G19, and third by G6-G11-G15-G20 [Figure 4A-(ii) and 4B-(ii)]. R1WT, R1mut1 and R1mut2 were found to conform into 3D monomolecular parallel G-quadruplexes with propeller type loop configuration. The highly planar arrangement of the

quartets limits multiple conformations, while providing optimal geometrical stability to the structure^{4,32}. Therefore, the two conformers of the R1WT: R1mut1 and R1mut2, were anticipated from molecular modeling. Our computational research suggested that R1mut2 and R1WT have a structural similarity. Both conformations utilized the '2-2-1' loop configuration and shared identical Gs that were sequentially engaged in the construction of quartet plates and loops. The close structural proximity between R1mut2 and R1WT points towards the conformational dominance and the putative significance of this structure in the cellular context. R1WT and R1mut2 with '2-2-1' loop topology and the shorter loop comprised of 1 nucleotide at the 3' end of the GQs is presumed to assist in enhancing the stability of the structure. R1mut1 having '2-1-2' loop configuration was mildly less stable than the R1mut2 structure.

2.2.4 Comparative structural stability of the REX1 GQ sequences

From 1D NMR and molecular modelling data, we speculated on two conformations of the wildtype GQ-R1WT that result in the dynamic nature of the structure *in-solution*. For maintaining precision of our analyzed data, we chose to advance with the characterization of R1mut2, the mutated sequence of R1WT, in which the structure is restricted to a single conformer. The melting temperature (T_m) of R1WT was analyzed to be 73.28 ± 2.12 °C, while the T_m of R1mut2 was 73.75 ± 1.07 °C and R1mut1 was 71.045 ± 2.13 °C. The CD melting assays, revealed the thermal stability of R1mut2 to overlay on that of R1WT [*Figure 3C-(ii) and 3-(iv)*], while R1mut2 also structurally resembles R1WT. Based on our biophysical and modelling results, we selected R1mut2 for further characterization of the GQ structure. We performed a series of biophysical and biochemical experiments, and molecular dynamics simulation to investigate the details of the binding interaction of R1mut2 with the small molecules selected in our study TMPyP4 and BRACO-19.

2.2.5 UV-Visible spectroscopic investigation suggested the interactional behaviour of TMPyP4 or BRACO-19 with REX1 GQ to be remarkably different

UV-Visible absorption spectroscopy was used to examine the interactional behaviour of TMPyP4 (TMP) or BRACO-19 (BRC) with R1mut2. Keeping the ligand concentration constant, it was titrated with R1mut2 oligo. The UV spectra was measured in 350-550 nm wavelength range, where the solet band was measured as a function of oligo concentration. The addition of R1mut2 resulted in significant hypochromic effect on the absorption of TMPyP4, along with a 21nm of large bathochromic shift (422 to 443nm) with an intense isosbestic point at 430nm. UV titration results indicated a binding interaction between TMP and R1mut2 [Figure 5A]. The observed bathometric shift and absorption hypochromicity of the solet-band was at a wavelength of 422nm for TMPyP4, which may have occurred because of the binding interaction involving N7 guanine (N7G)¹. $\Delta A/\Delta A_{\max}$ versus R1mut2 concentration was plotted to derive the K_d value (dissociation constant). K_d of $1.29 \pm 0.62 \mu\text{M}$ was derived [Figure 5A in-set]. Upon further titration of TMPyP4 with R1mut2, we observed an increase in absorption [Figure 6B] which led us to speculate instability of R1mut2 GQ upon interaction with TMPyP4. We presumed the local disruption of Hoogsteen H-bonds and the G-tetrad plates upon interaction with TMPyP4 as the plausible explanation of the observed latter increase in absorption. The final shifted solet band remained stable upon further titration. We presumed the binding of TMPyP4 with R1mut2 unfolds the GQ structure, which is unable to reform in the presence of the TMP ligand molecule.

The plausible binding interaction between BRACO-19 and R1mut2 was also investigated by performing UV titration in a UV-visible absorption spectrophotometer. BRACO-19 displayed two absorption maximum (A_{\max}) at 294 nm and 361nm. We observed hypochromism in the absorption of BRACO-19 at both these wavelengths upon titration with R1mut2 [Figure 5B].

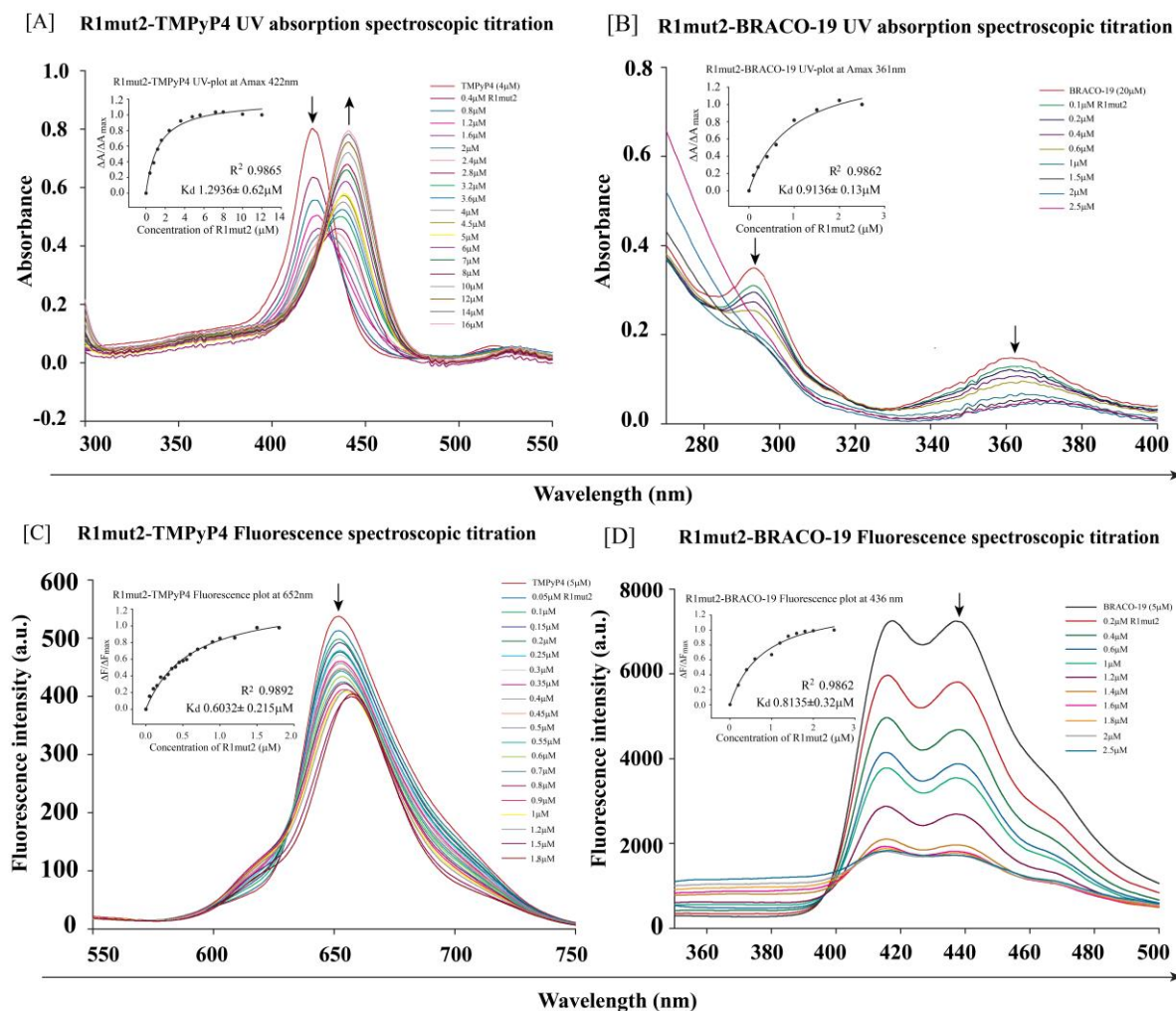


Figure 5. The figure displays the titration data of R1mut2 with small molecules. The graph A shows the UV absorption titration spectra of R1mut2 with an increment of TMP, where the in-set graph displays the fractional absorption at 422nm on the y-axis plotted against the oligonucleotide (R1mut2) concentration on the x-axis. Graph B shows the UV absorption titration spectra of R1mut2 with an increment of BRC, and the in-set graph represents the fractional absorption versus R1mut2 oligonucleotide concentration. Graph C displays the fluorescence titration spectra of R1mut2 with an increment of TMP, and the in-set graph shows the fractional fluorescence versus oligo concentration. Finally, graph D shows the fluorescence titration spectra of R1mut2 with incremental BRC, and the in-set graph shows the fractional fluorescence versus R1mut2 oligonucleotide concentration.

We speculated that π - π stacking interaction of BRACO-19 with R1mut2 may have resulted in the significant hypochromism of BRACO-19. BRACO-19-R1mut2 complex formation lowered the π - π^* transition energy gap that stabilized π^* -orbital of BRC. The fractional absorbance of BRC at 361nm was plotted against oligo concentration of R1mut2. The

calculated K_d value of BRACO-19-R1mut2 interaction is $0.914 \pm 0.13 \mu\text{M}$ [Figure 5B inset graph].

2.2.6 Fluorescence spectroscopic study revealed substantially distinct mode of interaction between R1mut2 and TMPyP4 or BRACO-19

The binding interaction of TMP and BRC with R1mut2 was analyzed by performing fluorescence spectroscopic titration. The ligand (TMPyP4 or BRACO-19) was held constant and titrated with oligonucleotide. In the titration of TMP with R1mut2, the fluorescence of TMPyP4 initially decreased, then gradually deviated towards the right (red-shift) followed by an increase in fluorescence upon further titration with R1mut2 [Figure 5C]. The initial decrease in fluorescence of TMPyP4 suggested end-stacking mode of binding-interaction, while the observed red shift (from 652 nm to 661 nm) indicated significant ligand-receptor (TMPyP4-R1mut2) binding interaction. The later increase in fluorescence is suggestive of unfolding of the GQ structure that is no more able to quench the fluorescence of TMPyP4 [Figure 6B]. The decrease in fluorescence intensity of TMPyP4 at 652nm with the gradual addition of R1mut2 oligo was used to derive the K_d value and the binding stoichiometry of TMPyP4 with R1mut2 GQ. We derived a K_d of $0.6032 \pm 0.215 \mu\text{M}$ for TMPyP4-R1mut2 interaction [Figure 5C in-set]. In the F_0/F versus oligo concentration graphplot, we observed a linear trend which suggest static quenching of TMP by R1mut2 [Figure 6D]. The Fluorescence spectrophotometer detected significant fluorescence hypochromicity of BRACO-19 upon titration with R1mut2 [Figure 5D]. The observed decrease in fluorescence is suggestive of external mode of binding of BRC to R1mut2 GQ. We considered the fluorescence emission max values of BRC at 436nm to determine the K_d of the binding interaction ($K_d = 0.8135 \pm 0.32 \mu\text{M}$) [Figure 5D in-set]. F_0/F versus oligo concentration graphplot of BRC-R1mut2 interaction displayed an initial linear trend showed which indicate static quenching of BRC, while the later portion of the graph is non-linear suggestive of dynamic quenching [Figure 6E]. The results indicate the possibility

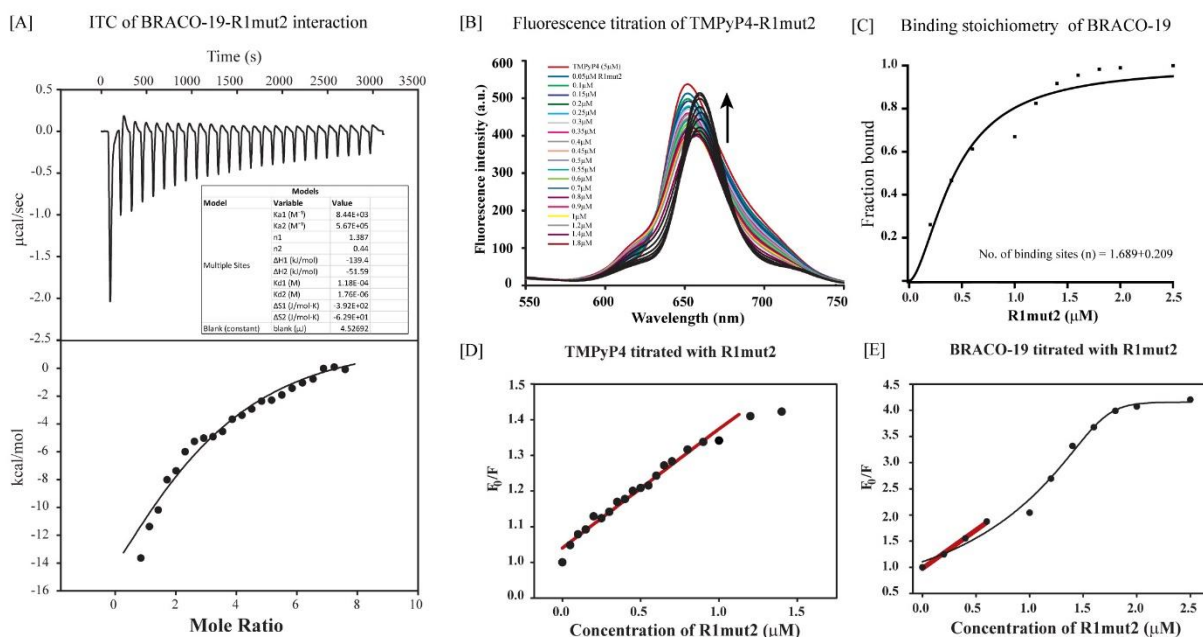


Figure 6. [A] Isothermal titration calorimetry (ITC) of BRACO-19-R1mut2 disclosing 2-site-binding of BRC to R1mut2, [B] Fluorescence titration spectra of TMPyP4 with increment of R1mut2. The spectra of increase in fluorescence of TMPyP4 at the later stage of titration are highlighted in black with an upward arrow, [C] Fluorescence data used to plot the fraction of ligand bound (BRACO-19) against concentration of R1mut2 and analyze the binding constant. No. of binding sites(n) was calculated out to be 1.7 which we presumed as 2. [D] The linear characteristic of F_0/F graph plot of TMPyP4-R1mut2 fluorescence titration indicating static quenching, [E] The F_0/F plot of BRACO-19-R1mut2 fluorescence titration data showing linear+non-linear graphical characteristic indicating both static and dynamic quenching of the fluorophore BRACO-19.

of multiple binding mode of interaction of BRC with R1mut2. Subsequently we performed an isothermal titration calorimetric (ITC) experiment to obtain the number of binding sites (ns) of BRACO-19. We confirmed two-site binding of BRC with R1mut2 GQ structure from the fluorescence and ITC data [Figure 6A and 6C].

2.2.7 Structural stabilization or destabilization of R1mut2 upon interaction with small molecules (TMP and BRC)

We investigated ligand-induced structural change by performing CD titration and 1D-1H NMR titration assay. Upon titration with TMP, R1mut2's spectral intensity at 260nm decreased gradually. The CD melting assay revealed the thermal stability of R1mut2 to decrease upon complexation with TMP [T_m : $67.49 \pm 2.4^\circ\text{C}$]. The drop in the CD spectral intensity and the

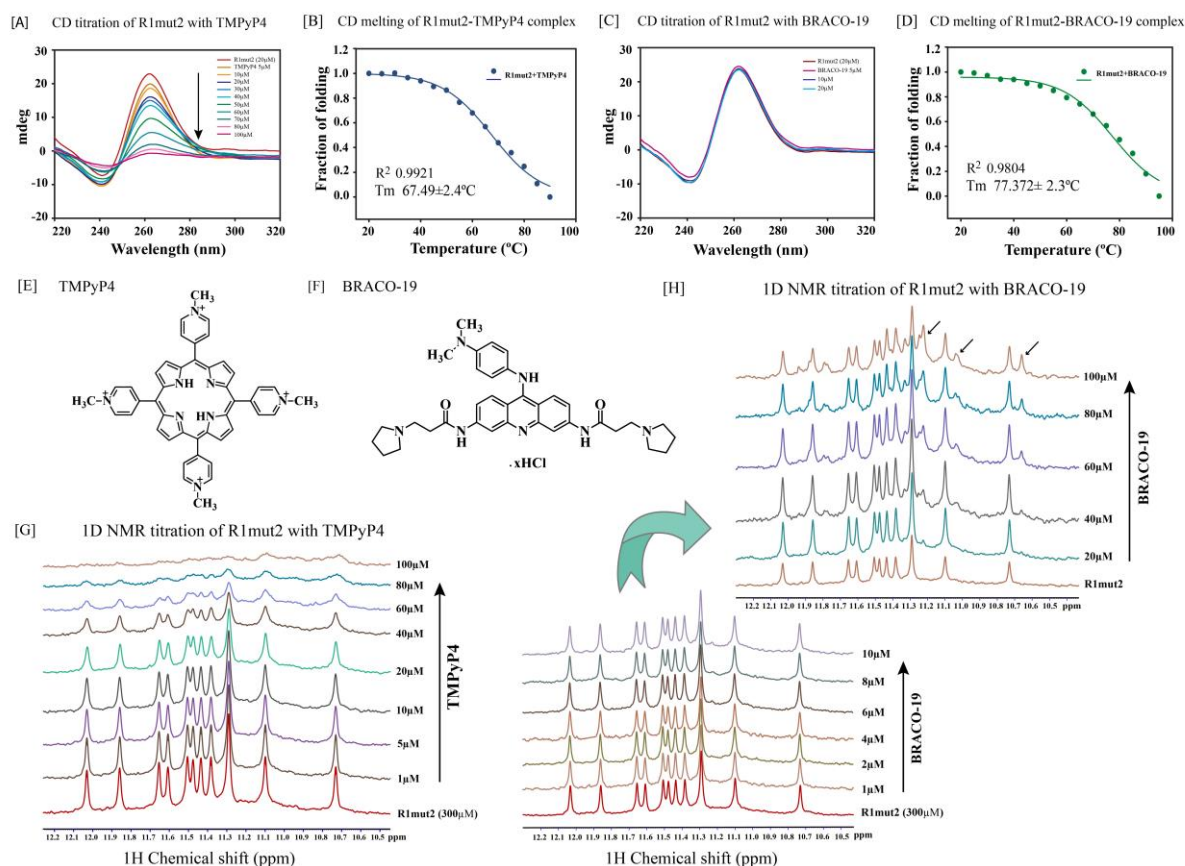


Figure 7. The impact of small molecule binding on R1mut2 is demonstrated. [A] The **CD titration** spectra of R1mut2 with an increment of TMP show a decrease in the CD spectral features of the R1mut2 GQ. [B] The **CD melting** profile. The CD melting data was used to calculate the melting temperature of the R1mut2-TMP complex. A decrease in the T_m was observed confirming R1mut2 quadruplex destabilization via interaction with TMP. [C] In contrast, the **CD titration** data of R1mut2 with increasing concentration of BRC shows no observable alteration in the intensities or spectral properties. This suggests that the parallel GQ structure is retained by R1 mut2 in the presence of BRC. [D] The **CD melting** assay was used to calculate the melting temperature of the R1mut2-BRC complex. T_m enhancement was observed, suggesting stabilization of the R1mut2-GQ by BRC. The **Molecular structures** of TMP and BRC are shown in [E] and [F], respectively. [G] Additionally, **1D-NMR titration** spectra of R1mut2 with increasing concentrations of TMP in the 10-12 ppm region were obtained. The GQ-specific imino proton signals decrease gradually with an increment of TMP, suggesting the unfolding of the GQ structure. [H] **1D-NMR titration** spectra of R1mut2 with increasing concentration of BRC in the 10-12ppm region. The imino proton signals, in this case, are seen to intensify upon the addition of BRC, confirming the stabilization effect of BRC.

thermal stability of R1mut2-GQ indicate TMPyP4-induced structural destabilization [Figure 7A and 7B]. However, we did not observe any conformational transition. In the 1D-1H-NMR, the incremental addition of TMP to R1mut2 led to gradual fading of the characteristic imino peaks in 10-12 ppm range [Figure 7G]. The loss of imino proton signals upon incremental

addition of the ligand suggest structural destabilization of R1mut2 upon interaction with TMPyP4 [Figure 7E].

In the titration of R1mut2 with incremental BRC, no significant change in the spectral characteristic of R1mut2 was observed. CD titration data suggest the maintenance of the parallel conformation of R1mut2 GQ in the presence of BRC [Figure 7C]. The melting temperature of R1mut2-BRACO-19 complex increased (T_m : $77.372 \pm 2.3^\circ\text{C}$), suggesting enhanced stability of R1mut2 upon complexation with BRACO-19 [Figure 7D]. In the 1D NMR titration of R1mut2 with BRC, initially BRC was observed to stabilize the GQ as indicated by the higher resolution of the imino peaks of R1mut2, further titration led to evolution of additional imino peaks within 10-12 ppm, suggesting elevated dynamicity of the proton signals due to better engagement of guanines in GQ formation [Figure 7H]. Thus, the titration results indicate BRC mediated stabilization of R1mut2-GQ.

2.2.8 Corroboration of the interactive behavior of TMP or BRC with REX1 GQ

We performed DMS footprint assay to verify the biophysical findings with the wildtype conformation. 5' FAM tagged R1WT was further investigated under different circumstances including a variable concentration of TMP and BRC. DMS reaction of R1WT sequence was conducted under the mentioned conditions:

- (i) R1WT annealed in water,
- (ii) R1WT annealed in KP-KCl buffer,
- (iii) R1WT+ TMPyP4 in 1:1 ratio
- (iv) R1WT+ TMPyP4 in 1:3 ratio
- (v) R1WT+ BRACO-19 in 1:1 ratio
- (vi) R1WT+ BRACO-19 in 1:3 ratio

DMS Protection assay under different conditions

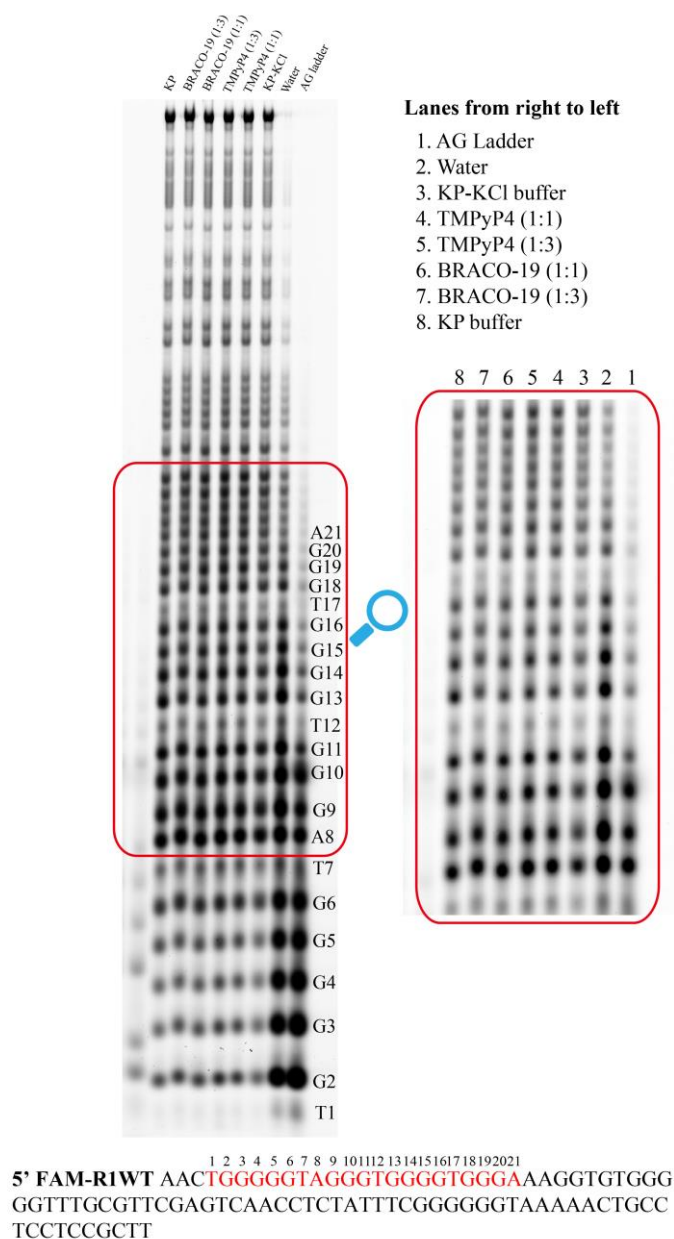


Figure 8. DMS protection assay: We can see the results of a DMS protection/ footprinting assay on the DNA sequence containing the R1WT-GQ region. The lanes on the gel from right to left represent the following conditions: A/G ladder, R1WT annealed in water, R1WT annealed in KP-KCl, R1WT annealed in KP-KCl in the presence of TMP at 1:1 and 1:3 ratio, R1WT annealed in KP-KCl in the presence of BRC at 1:1 and 1:3 ratio, and R1WT annealed in KP buffer. The results show that R1WT is unable to adopt a GQ structure when annealed in water, as observed by the lack of guanines participating in Hoogsteen base pairing. The GQ structure is formed in the presence of KCl and is further stabilized by BRC, while it is destabilized by TMP. This is inferred by a lightening of G-bands in the presence of BRC and a darkening of the same in the presence of TMP. Finally, a magnified image of the middle nucleotides is provided for better viewing.

Upon analyzing the DMS data, we observed darker intensity of the guanine bands in Tris-Cl or water compared to KP-KCl. Thus, we confirmed that the R1WT sequence does not conform to a G-quadruplex when annealed in water or Tris-Cl buffer. Also, we observed darker intensity of the bands (Gs) in TMPyP4 compared to KP-KCl condition. This data confirms the destabilizing effect of TMPyP4 on R1WT G4 structure. TMP unfolds the GQ, making the N7 position of guanines susceptible to DMS methylation and subsequent cleavage by piperidine [Figure 8]. In case of the BRACO-19-R1WT complexed condition, the DMS reaction yielded lighter intensity of the G bands leading us to presume the stabilizing impact of BRC upon R1WT. The N7 atom of the guanines in the quadruplex are shielded from DMS-methylation and hence not cleaved by piperidine, resulting in the lighter band intensity of the specific guanines [Figure 8]. We observed a good correlation of our biophysical data with the DMS data, thereby validating the destabilizing effect of TMP and the stabilizing impact of BRC upon interaction with the monomeric conformer R1mut2.

2.2.9 Computational scrutiny of the interaction between REX1 GQ with TMP or BRC

We analyzed the interaction between TMP and BRC with R1WT and R1mut2 by *in silico* methods. Molecular docking and molecular dynamics (MD) simulation were employed in the investigation of receptor-ligand interaction. Docking results of the R1WT/R1mut2- TMP complex revealed TMP to bind terminally to R1mut2 and R1WT GQ structures, interacting primarily *via* hydrophobic interactions. In the molecular simulation of R1mut2-TMP, we observed distortion in the alignment of the GQ structure, thereby strengthening our biophysical and biochemical results of the destabilizing effect of TMPyP4 on the GQ [Figure 9A]. At 20 nanoseconds of the simulation run, the first potassium ion was displaced out of the complex, and at 180 nanoseconds, the second potassium ion moved away from the center. The simulation result of R1WT-TMPyP4 complex buttressed the destabilizing effect of TMPyP4 on R1WT-GQ. Guanine numbers:18, 19, and 20 of R1WT flipped back from the tetrad plane, thereby

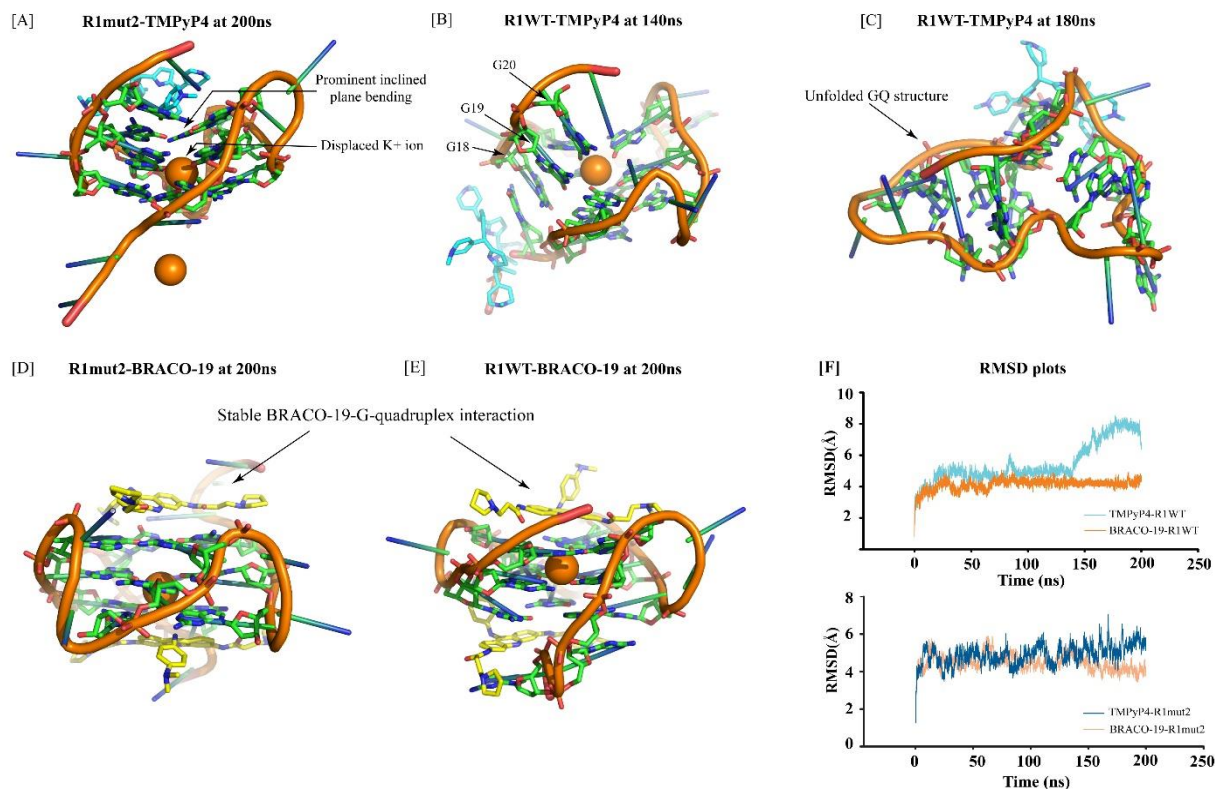


Figure 9. Molecular dynamics (MD) simulation. [A] Simulation end result of R1mut2-TMPyP4 complex after 200 nanoseconds of simulation run showing distorted orientation of R1mut2 GQ upon interaction with TMPyP4. The prominent inclined plane bending of the guanines and displaced potassium ions from the centre position indicates TMPyP4 induced instability of R1mut2 GQ structure. [B] Simulation result of R1WT-TMPyP4 at 140 nanoseconds showing the flipping away of the guanines numbered G18, G19, and G20 from the centre tetrad plate orientation, thereby initiating destabilization of R1WT GQ. [C] Simulation result of R1WT-TMPyP4 at 180 nanoseconds showing completely unfolded or destabilized R1WT GQ upon interaction with TMPyP4. [D] Simulation end result of R1mut2-BRACO-19 at 200 nanoseconds showing stacking of two BRACO-19 molecules on the terminal ends of R1mut2 interacting stably via π - π interaction. [E] Simulation end result of R1WT-BRACO-19 at 200 nanoseconds showing two molecules of BRACO-19 stacked on the two terminals of R1WT GQ structure, thereby enhancing the stability of the quadruplex structure. [F] RMSD plots. Upper panel graph showing overlayed RMSD plot of the simulation trajectory of R1WT simulated with TMPyP4 (sky blue) and with BRACO-19 (orange); lower panel graph displaying comparative RMSD plot of the simulation trajectory of R1mut2 simulated with TMPyP4 and BRACO-19. Higher RMSD value denotes destabilizing interaction. The colour code of BRACO-19 molecule is yellow and the colour code of the TMPyP4 molecule is cyan.

distorting the ordered structure of the GQ. At 25 nanoseconds, the first potassium ion displaced out of the complex, at 140 nanoseconds the GQ structure partially broke, and at 180 nanoseconds, the entire structure disintegrated with the release of both the K^+ ions [Figure 9B and 9C]. TmPyP4 mediated destabilization of R1WT and R1mut2 is reflected in the RMSD plot [Figure 9F].

ITC and fluorescence spectroscopic data disclosed two-site binding of BRACO-19 with R1mut2. So, we docked two molecules of BRACO-19 onto each of R1WT and R1mut2. The molecular double docking results revealed stacking mode of interaction between BRACO-19 and R1WT/R1mut2. The three interconnected cyclic rings of BRC stack upon the end-quartet plate interacting through π - π interaction. Terminal stacking of BRC at both ends of the R1WT and R1mut2 is presumed to enhance the structural stability. The two potassium ions at the center stayed confined at their places for 40 nanoseconds of R1mut2-BRC complex simulation and 115ns for R1WT-BRC complex simulation. The structure and stability of R1WT and R1mut2 GQs remained unimpeded, supporting our presumption of BRACO-19-mediated stabilization of R1WT and R1mut2 quadruplexes [Figure 9D and 9E]. The RMSD plot and the simulation data indicated that the GQ structure was stable in the complex form with BRACO-19, thereby supporting BRACO-19 mediated R1WT and R1mut2 GQ stabilization [Figure 9F].

2.2.10 *In vitro* existence of REX1 G-quadruplex

We have employed the GQ-specific antibody BG4^{33–35} to investigate the *in vitro* existence of R1WT GQ. By conducting chromatin immunoprecipitation assay (ChIP), we have pulled down the DPR of REX1, embracing the R1WT sequence by antibody treatment. Positive pull-down of the R1WT region with the BG4 antibody confirmed the existence of this newly identified GQ inside the cell [Figure 10A].

2.2.11 Endogenous R1WT sequence serves as a docking platform for multiple gene regulatory proteins

To maintain the cellular identity, a very intricate genetic programme encompassing the chromatin architecture, regulatory transcription factors, and epigenetic markers is tightly enforced¹². According to the available data, GQs control transcription by engaging multiple transcription factors (TFs) and chromatin remodeling proteins. Additional reports on the physical interaction of distally located GQs in the genome affirm the spatio-temporal existence and regulatory activity of GQs³⁵. GQ formation is linked to transcriptional enhancement and gene silencing³⁶. To explore the biological importance of R1WT GQ, we investigated R1WT-TF interaction at the chromatin level in the cancer cell model A549. Initially, we confirmed the expression of REX1 in A549 lung cancer cell-line. We speculated, that if R1WT acts as a silencer element, GQ formation will be less favored in the promoter coding strand to maintain active state of REX1 gene. According to our hypothesis, the R1WT element acts as the recognition site for various TFs that bind to the same genomic loci and execute a molecular program. We have performed ChIP of the R1WT region with antibodies against selected transcription factors documented to have binding preference for GC-rich segments. The master transcription factor Sp1 (specificity protein) is a ubiquitous protein that is reported to control multiple molecular programs at the chromatin level³⁷. We have reports from the literature survey that Sp1 binds to GC-boxes and G-rich quadruplex forming sequences with a stronger binding affinity for GQ sequences over the Sp1 specific canonical consensus motif³⁸. The binding of Sp1 to GQ forming sequences in the genome ensues active transcription. Earlier reports correlate the association of Sp1 with demethylated CpG islands that are the pre-requisites for active gene expression³⁵. Our ChIP assay revealed a positive interaction of Sp1 at the R1WT region pulled down by anti-Sp1 antibody [*Figure 10B*].

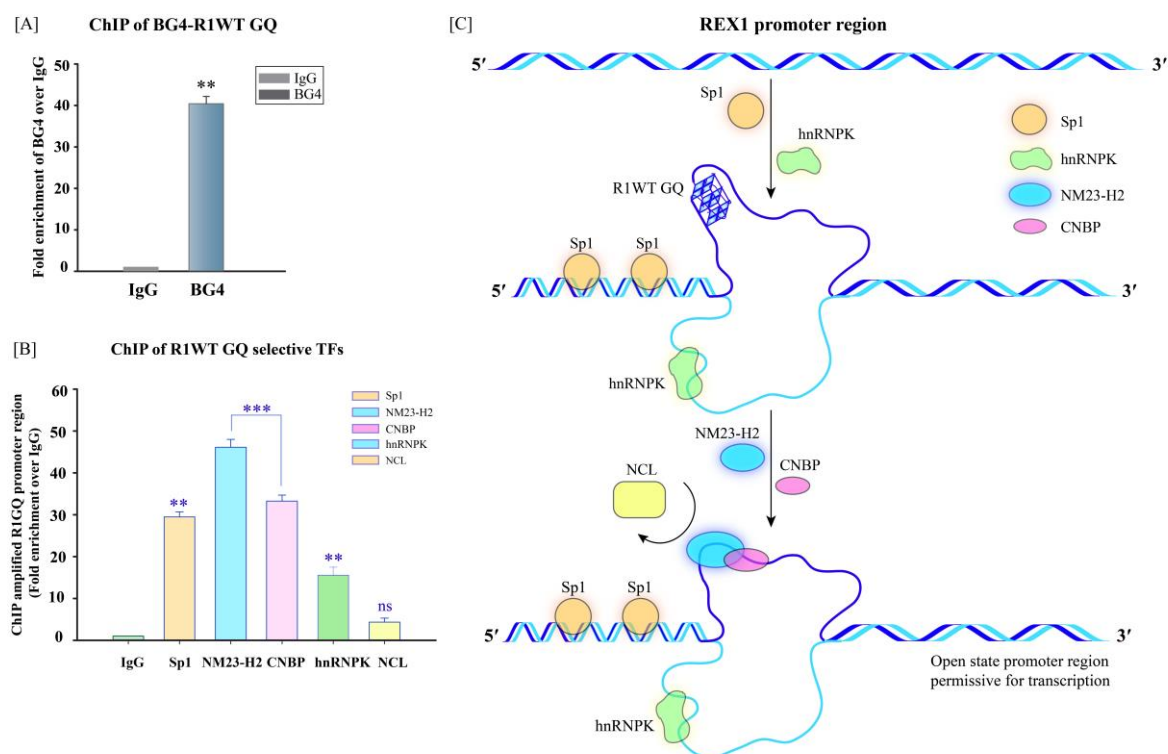


Figure 10. The *in vitro* interaction module of R1WT-GQ with the transcription factors (TFs). **[A]** Bar plot showing the relative precipitation of R1WT GQ region with BG4 antibody. Positive pulldown of BG4 antibody with R1WT segment confirms the *in vitro* evolution of R1WT quadruplex. The mean of normalized values (\pm SE) from the set of triplicate data is represented by error bars. Asterisks (*) indicate statistical significance, determined by paired t-test (** indicates $p < 0.01$), showing significant binding of BG4 with R1WT region compared to mock IgG in the chromatin context. **[B]** ChIP shows the differential occupancy level of selected TFs (CNBP, NM23-H2, Sp1, hnRNPK, and NCL) at R1WT distal promoter region relative to IgG. The quantification of TF-binding by fold enrichment method based on RT-qPCR data is shown on the histogram bar plot. IgG acted as the negative control. Asterisks (*) indicate statistical significance determined by a paired t-test. (** indicates $p < 0.01$, *** indicates $p < 0.001$, and ns indicates non-significant difference). These symbols denote the significance of the association of different TFs with the R1WT region compared to the mock IgG. **[C]** The R1WT segment of the DPR of the REX-1 gene is shown in the schematic representation. Sp1 induces negative superhelicity at the 5' end of R1WT. This makes it easier for the hnRNPK protein to bind to the opposite strand's C-rich element in R1WT. The unwinding of this region then attracts the NM23-H2, followed by CNBP forming a temporary complex to maintain the REX-1 promoter region in a state that allows for easier transcription.

Our study's excavation suggests that R1WT is a crucial epigenetic marker for TF-Sp1 recruitment. We hypothesized that Sp1 either directly identifies and binds to the R1WT motif or that other TFs assist in its recruitment. We presumed the recruitment of Sp1 to the chromatin region featuring co-binding proteins as it offers better access to efficient DNA-TF interaction³⁹. The member of NM23 metastatic suppressor protein family⁴⁰, the protein factor NM23-H2 is reported as a G-quadruplex destabilizing transcription factor⁴¹. There are reports on the correlation of NM-23-H2 binding sites with depleted nucleosomal regions⁴². NM23-H2-GQ interaction is reportedly associated with upregulated *c-MYC* in multiple cancers⁴². The transcription factor Cellular Nucleic Acid Binding Protein (CNBP) is well-reported to be more oriented toward G-rich sequences with a higher propensity of GQ formation⁴¹. A recent report characterized CNBP as a GQ-unfolder that enhances the expression of oncogenes⁴¹. CNBP reportedly unfolded the GQs in the promoter of genes: VEGF, c-MYC, PDGF, and c-KIT genes that elevated their expression level and promoted cancer malignancy⁴¹. There are reports on the association of NM-23-H2 and CNBP at the NHE_{III1}-element of the *c-MYC* promoter governing the upregulation of *c-MYC* oncogene in cancer⁴³.

In our precipitation assay, we observed significant high occupancy of CNBP and NM-23-H2 at the R1WT promoter segment of REX1 gene, indicating nucleosome depletion. Thus, we anticipate the R1WT sequence as the nuclear hypersensitive element in the REX1 promoter [Figure 10B]. From our ChIP results, we contemplated that the positive binding of CNBP and NM-23-H2 transcription factors to R1WT region triggers quadruplex unfolding and maintains the underneath chromatin (promoter region) in an active transcriptional state [Figure 10C]. hnRNPK (heterogenous nuclear ribonucleoprotein K); a well reported TF that interacts with single stranded DNA (ssDNA). According to studies, hnRNPK and GQs are tightly associated with the regulation of transcription⁴⁴. Our immunoprecipitation assay divulged the positive binding interaction of hnRNPK with the R1WT GQ region, suggesting an unwinded state of

the promoter in A549 metastatic cell line [Figure 10B]. In the chromatin state, G-quadruplex structures evolve, relaxing the complementary C-rich stretch with the propensity to conform to an i-Motif (iM) structure⁴⁵. We speculated that the Cytosine rich sequence complementary to the R1WT is the plausible binding site of hnRNPK. The multifunctional phosphoprotein nucleolin (NCL) is a documented GQ binding transcription factor. Repression of *c-MYC* transcription is reported upon NCL binding to the regulatory GQ^{3,38}. In the chromatin immunoprecipitation assay, we found an insignificant precipitation of NCL with the R1WT chromatin DNA fragment. ChIP results suggest that the open state of the R1WT sequence gets favored over the folded-GQ conformation. We hypothesised that low NCL enrichment at R1WT was the result of Sp1-induced torsional stress and hnRNPK binding at the complementary C-rich region, which relieved the transcriptional blockade.

The results of our investigation provide credence to the hypothesis that R1WT in the REX1 DPR is a potential docking platform for NM-23-H2 and hnRNPK in the chromatin state. NM-23-H2 plausibly unfolds R1WT into single stranded relaxed DNA. CNBP transiently couples with NM-23-H2 and maintains the relaxed state of the promoter DNA. The binding of NM-23-H2 assist in the recruitment of hnRNPK to the Cytosine-rich strand, which subsequently impedes the binding of nucleolin (NCL). The recruitment of Sp1 to the R1WT region may be aided by co-binding proteins. Sp1 subsequently adds negative supercoiling stress that exposes the ssDNA, facilitating hnRNPK recruitment to the C-rich sequence [Figure 10C]. Therefore, in collaboration with Sp1 and hnRNPK, NM-23-H2 modulates the distal R1WT region of the REX1 promoter, which assists CNBP to warrant the expression of REX1. The dominance of the unfolded state of R1WT suggests its plausible link with the transcriptionally permissive state of chromatin that primes A549 cancer cells with pluripotent propensity.

2.3 Conclusion

Regulatory G-quadruplex sequences are promising drug targets for next-generation cancer therapeutics. An extensive investigation of the GQ structural configuration is essential for the successful designing of an efficacious cancer targeted drug³². We identified and characterized a stable GQ in the DPR of REX1 gene. We observed the G-rich sequence R1WT conform into a parallel tetrad structure consisting of three stacked quartet plates connected by short loops. Structural anatomy revealed the architecture of the quadruplex to render high stability and flexibility to R1WT. We determined the structural flexibility of the newly identified quadruplex by conducting a series of biochemical, biophysical assays and *in silico* analysis. Small molecule-GQ interaction studies revealed the destabilizing effect of TMP and the stabilizing impact of BRC upon the monomeric conformer of R1WT (R1mut2). *In vitro* targeting of R1WT is a promising avenue in the development of anti-metastatic drug. Drug mediated stabilization of R1WT into its folded GQ conformation may suppress the stemness trait of metastatic cells. Within the genome, R1WT GQ most likely serves as a gene-regulatory-switch controlling transcription. It orchestrates the TFs and holds the trigger of REX1 gene activation. At the chromatin level, deregulated transcription in cancer disrupts the balance between the formation and dissolution of G-quadruplex structures, which directly affects gene regulation by altering chromosomal structure⁴⁶. Our research aligns with the function of the R1WT motif as a gene regulatory element serving as a docking site for several transcription factors. We reported the binding interaction of NM-23-H2 in association with CNBP, Sp1, and hnRNPK transcription factors at the R1WT region that confers regulatory control on the promoter activation-deactivation state of REX1.

The R1WT sequence in the REX1 promoter is a gene regulatory element that provides a docking platform for important regulatory transcription factors. This DNA-TF system is promising for developing stemness-targeting anticancer drugs. The research has the potential

to result in a significant advancement in the treatment of malignant cancer, possibly leading to a breakthrough. We detected CCCTC sequence (the binding site for CTCF transcription factor) within the REX1 promoter span. The presence of CCCTC recognition element suggests the implication of long range control (LRC) that mediates distal DNA interaction⁴⁷ by the promoter of REX1 in regulating transcription. The location of R1WT in the DPR of REX1 either directly assists in the activity of the Polymerase or functions as an accessory regulator that assist in REX1 gene regulation *via* DNA looping¹². Thus, the regulation of REX1 transcription is anticipated to be much more intricate than the mere interplay of NM-23-H2, CNBP, Sp1 and hnRNPK at the R1WT site. In the GQ research field, the innovative use of duplex/quadruplex motifs in the design and discovery of highly selective dual-motif targeting molecules is in its infancy with a promising perspective on drug development²⁷. A combinatorial strategy is currently under investigation to circumvent the specificity problems of small molecules. It combines synthesized small molecules with established tag proteins²⁷. Extensive research into the structural and functional aspects of GQs and their interactive communication with other factors in the cellular environment will lead to successful and innovative cancer drug development.

2.4 Material and methods

Bioinformatic analysis

The Genomatix software suite's MatInspector was used to determine the promoter region of the REX1 (ZFP42) gene. The gene sequence for REX1 was retrieved from the NCBI gene database via Gene ID 132625⁴⁸. Analysis of the G rich sequences were done using the QGRS mapper tool⁴⁹. The Eukaryotic Promoter Database (EPD) was hunted to retrieve the TSS of the REX1 gene.

Reagents and Oligonucleotides

The REX1 promoter region comprises the G-rich oligonucleotides termed R1 oligos (R1WT, R1mut1, and R1mut2), R2, R3, and R4. These oligonucleotides were obtained from Eurofins India Pvt. Ltd and stored in commercially available, ultra-pure, DNase and RNase-free distilled water. R1mut1, R1mut2 refer to the mutated forms of R1WT. We also obtained TMPyP4 (TMP), a cationic porphyrin derivative, 5,10,15,20-Tetrakis(1-methyl-4-pyridinio) porphyrin tetra(p-toluenesulfonate), and synthetic acridine BRACO-19 (BRC) from Sigma-Aldrich (catalog no. 323497 and SML0560, respectively). We prepared the main stocks of TMP and BRC in ultra-pure water (Invitrogen) at concentrations of 10 mM and 5 mM, respectively. We stored TMP in the dark at -20°C and BRC in dark at 4°C.

Serial No.	Sequence name	Sequences	Length	G-Score
1.	R1WT	TGGGGGTAGGGTGGGGTGGGA	21	41
2.	R2	GGGTCAGGGGGCCCCGGGCTGAGGG	25	41
3.	R3	GGGTGCCTGGGGGGCCCCGGGCTGAGGG	27	40
4.	R4	GGGTCTTGGGAGGGGGCGCAGGG	23	41
5.	R1mut1	TTAGGGTAGGGTGGGTGGGA	21	41
6.	R1mut2	TTAGGGTAGGGTGGGTGGGA	21	41

Table 1: A list of the sequences of oligonucleotides that have been studied. QGRS mapper tool predicted the G scores for the sequences.

Sample preparation

The working concentrations of the ligand molecules were freshly prepared and equilibrated at RT for 1 hour. Two buffer conditions KP-KCl and Tris-Cl at neutral pH were kept for the analysis of the oligonucleotides. KP-KCl is 10mM of potassium phosphate (KP) buffer supplemented with 100mM potassium chloride (KCl). The prepared oligonucleotide samples were annealed at 95°C for 5mins and gradually cooled to RT and further used in experiments. The leftover samples were divided into smaller portions and then stored at a temperature of -20°C for future use in experiments.

1D-¹H NMR spectroscopy

We used Bruker 700 MHz spectrometer to perform our NMR experiments. The machine is equipped with 5mm SMART probe. NMR samples of oligonucleotides were prepared in KP-KCl buffer with 10% D₂O. We conducted the 1D NMR scans in shigemi tube with a 300 μM of oligonucleotide concentration and sample volume of 350 μl. TSP was used as an internal standard reagent at 0.0 ppm. It served as a reference for each recorded spectral scan. The imino proton signals of the putative GQ forming sequences under study were monitored using the pulse program Bruker Pulseprog 'zgesgp' with a 20ppm spectral width, number of scans(ns): 64, and calibrated pulse length (p1): 12.48 μs. The 1D proton spectra were monitored for imino proton signals in the 10-12 ppm range. In the 1D NMR titrations, each of TMPyP4 and BRACO-19 were incrementally added to a constant oligonucleotide R1mut2 concentration of 300 μM. The ratio of oligonucleotide to ligand (TMPyP4/BRACO-19) at the final titration point was 1: 0.3. Further incremental rise in ligand concentration was not done to avoid aggregation and precipitation. The samples were homogenized after each addition of the ligand and allowed to attain thermal equilibrium at 25°C before acquisition of the proton spectra. For precise signal intensity comparisons during overlaid spectral representation, each spectrum was fixed at the same scale⁵⁰.

Circular dichroism (CD) spectroscopy

The signal spectra were measured in the spectrophotometer (Jasco J-815) connected to a Peltier system. The cuvette used had a path length of 1mm. Spectra were collected in the wavelength range (220-320 nm) with band width: 1nm, step size: 1nm and spectral scanning speed: 100nm/min. We performed the CD titration experiments at 298K. The working concentration of the oligonucleotides (20 μM) were freshly prepared in different buffer conditions (KP-KCl or Tris-Cl) for monitoring the CD spectral characteristics. In the CD titration assay the oligonucleotide concentration was maintained at 20 μM with an incremental addition of ligand

molecule (TMPyP4 or BRACO-19) upto a ratio of 1:5 (oligo: ligand). Each measurement was collected after 5mins incubation to ascertain proper complex formation and equilibration. Each spectrum was an average of three measurements collected after baseline correction. 1:3 ratio of oligo: ligand (R1mut2: TMPyP4/BRACO-19) was used in the CD melting experiments. A temperature of 5°C incremental increase from 20°C to 90°C was set for the CD melting assay. The calculated values of the fraction of folding were plotted against the temperature, followed by curvefit using sigmoidal 3-parameter binding equation to obtain the T_m value of the oligos in free state versus in [oligo+ligand] complexed state.

UV-visible spectroscopy

Cary60 UV-Vis spectrophotometer was used to perform the UV titration experiments. Quartz cuvette with 1cm path-length was used for sample holding. The concentration of the ligand TMPyP4 (4 μM) was kept constant and titrated with R1mut2. The absorption spectra were collected between 350-500 nm wavelength range. The values of maximum absorption of TMPyP4 (A_{max}) were observed at 422nm which was used for deriving data. Fractional absorbance was plotted against oligonucleotide concentration. In the UV titration experiment of R1mut2-BRC, the small molecule ligand was held constant and titrated with R1mut2, while the spectra were monitored between 250-400 nm wavelength range. The absorption A_{max} of BRACO-19 was used to plot the fractional absorption versus R1mut2 concentration. The data were fitted using ligand binding equation and the dissociation-constant(K_d) was derived for each of the ligand TMP and BRC interaction with R1mut2.

$$\text{Fractional absorbance} = (A_{\text{max}} - A) / (A_{\text{max}} - A_{\text{min}})$$

A_{max} represents the maximum absorbance, while A_{min} represents the minimum absorbance. A refers to the absorbance at each point of titration. In SigmaPlot 12.0 software, the following inbuilt equation is used for the interaction between TMPyP4-R1mut2: Ligand Binding with one site saturation. The equation is as follows: $f = B_{\text{max}} * \text{abs}(x) / (K_d + \text{abs}(x))$. In SigmaPlot 12.0 software, the following

inbuilt equation is used for the interaction between BRACO-19-R1mut2: Ligand Binding with two site saturation. The equation is as follows: $f = B_{\max 1} * \text{abs}(x) / (K_{d1} + \text{abs}(x)) + B_{\max 2} * \text{abs}(x) / (K_{d2} + \text{abs}(x))$.

Fluorescence spectroscopy

The fluorescence titration experiments were done using the Hitachi F-7000 FL spectrophotometer. Each of TMP and BRC ligand concentration was kept constant at 5 μM and titrated with R1mut2 GQ. The buffer environment was KP-KCl at pH 7. Excitation of TMP was at Ex 435nm and the emission was monitored in the range (λ_{Em} 550-750nm) upon titration with increasing concentration of R1mut2. In the fluorescence titration of BRC with R1mut2, the sample excitation was 300nm (λ_{Ex} of BRACO-19), and the emission spectra was measured between 350-550 nm (the emission range of BRACO-19). The fluorescence intensity [F_0/F] versus oligo concentration plot was done that represented the type of quenching of each of the two ligands upon interaction with GQ.

The fluorescence intensity of the free ligand is denoted by F_0 , while F represents the fluorescence intensity of the ligand in the presence of oligo. Additionally, we also calculated the fractional fluorescence using the following formula:

$$\text{Fractional fluorescence} = (F_{\max} - F) / (F_{\max} - F_{\min})$$

Here, F_{\max} represents the maximum fluorescence of the ligand, F_{\min} represents the minimum fluorescence of the ligand, and F denotes the fluorescence at each point of titration. Curvefit and K_d calculation was done using the inbuilt ligand binding equation of Sigma Plot12.0 software. The curve fit equations mentioned under the methodical section ‘UV-visible spectroscopy’ was used to fit the curves of the fluorescence titration data.

DMS (dimethyl sulfate) footprint assay

The REX1 promoter segment of 89-nucleotides possessing the R1WT sequence was commercially synthesized and purchased from IDT. The segment was tagged with fluorescein amidite (FAM) at its 5' end. A stock solution of FAM-tagged R1WT in Tris-Cl buffer was prepared for the DMS reaction. 10 μ M tagged oligo sub-stocks were prepared in KP-KCl or Tris-Cl followed by heating at 95°C and gradual cooling to room temperature. Different reaction conditions were kept according to the aim of the experiment. Various reaction conditions for the DMS reaction were set up with FAM-tagged-R1WT sequence (10 μ l) and a reaction volume of 100 μ l prepared. Each DMS reaction was performed with FAM tagged R1WT DNA sequence (2 μ M) reacted with 1% DMS solution for 2 minutes at RT. To stop the methylation process, 25 μ l of a stop solution was added to each reaction. [Stop solution with a total volume of 100 μ l was prepared. To prepare it, 43 μ l of sodium acetate (pH 7) was added, 7 μ l of Beta-mercaptoethanol (β -ME), and 50 μ l of double distilled water (ddH₂O) to make up the remaining volume]. The reacted samples were further treated with chilled absolute ethanol (250 μ l) and Sodium Acetate of pH 5.2 (37.5 μ l). After 30mins of centrifuging the samples at 13,000xg, the formed pellet was washed with 70% chilled ethanol. Samples were lyophilized for further piperidine treatment. For preparation of the A/G ladder, the FAM tagged R1WT containing DNA sequence was treated with 4% formic acid (10 μ l) for 30 minutes at 37°C. The A/G ladder samples were also lyophilized. The lyophilized samples were treated with piperidine (1M) by incubating them at 95°C for 30 minutes. After that, the treated samples were washed with water and lyophilized again. Next, the samples were dissolved in formamide dye and loaded onto a 12% denaturing acrylamide-bisacrylamide gel and allowed to separate at 1000 V for 4 hours. The resulting gel image was captured using a Typhoon phosphorimager.

Molecular modelling

The Molecular Modeling suite of the Schrödinger software was utilized for building the GQ models under study. As a starting point, we utilized the c-MYC G-quadruplex NMR structure with PDB ID 1XAV and constructed molecular models of the R1 oligonucleotides (R1WT, R1mut1, and R1mut2). Accuracy was maintained while editing and renaming the nucleotide bases and atoms. AMBER14 with ff14SB and parmBSC1 force- fields were employed to execute energy minimization of the modelled GQ structures.

Molecular docking and dynamic simulations

The Schrodinger software suite was used in preparing the structures of TMP and BRC for further docking and simulation. The structures were further sculpted in Ligprep. Molecular docking was done using HDOCK web server to dock the ligands with R1WT or R1mut2. Using a hybrid algorithm of template-based modeling and ab initio free docking, this bioinformatic server was utilized in predicting ligand receptor interactions. R1WT- TMP/BRC and R1mut2- TMP/BRC docking was done. Molecular simulation was performed for each of the four GQ-small molecule docked complexes. The molecular models were processed utilizing the Graphical User Interface of Chemistry at Harvard Macromolecular Mechanics (CHARMM-GUI)⁵¹. Ligands were parameterized using CGENFF (ParamChem server)⁵². The CHARMM36 force field was utilized for system runs⁵³ while a TIP3P cubic water box was utilized to solvate the system. This ensured a minimum of ~9 Å uniform thickness of the water layer. The overall charge neutralization of the system was done by addition of K⁺ and Cl⁻ ions into the system at a concentration of 0.15M. The GQ structures incorporated two potassium ions at their center. About 17,200 atoms in total, comprising a ligand and 21 nucleotide residues, were present in each system. NAMD 2.12 package⁵⁵ was used to perform the molecular dynamics (MD) simulation at 310K. The non-bonded interactions with a short range were truncated at 12Å, and a switching function was utilized between the 10Å and 12Å range.

The particle mesh Ewald (PME) method was used to compute long-range electrostatic interactions⁵⁶.

After performing energy minimization⁵⁷, heating, followed by equilibration steps, a 200ns of simulation was conducted. To control pressure and temperature, Langevin thermostat and Nosé-Hoover piston pressure control algorithm was used^{58,59}. Under consideration were the equilibrated structures as the initial conformation for molecular dynamics simulations that ran for 200 ns using NAMD 2.12 package. To maintain the hydrogen bonds, the SHAKE algorithm was employed with an integration time step of 2fs. At regular intervals of 4ps the frame of the trajectories was saved. VMD⁶⁰ and PyMol⁶¹ was used to visualize the data.

Cell culture

DMEM Media Gibco Dulbecco's Modified Eagle Medium (Catalog no. 11995-065) was used for maintaining A549 lung cancer cell-line. 10%(v/v)-FBS, 1%Penicillin-Streptomycin, 50µg/ml Gentamicin, and 2.5µg/ml Amphotericin-B were added to the final working media. CO₂ Incubator from ESCO, [Model no. CCL-1708-8- UV] was used to maintain the mammalian cells. A stable temperature of 37°C with 5% CO₂ supply and humid environment was maintained in the incubator for proper cell culture. Cells were maintained and passaged every 3 to 4 days with media replenishment every 24 hours to maintain healthy growth of cells.

Chromatin immunoprecipitation (ChIP) assay and RT-qPCR

The binding interaction of TFs (CNBP, NM-23-H2, Sp1, hnRNPK, and NCL) with the R1WT region of REX-1 DPR was monitored by performing chromatin immunoprecipitation assay (ChIP) experiment. PierceTM Magnetic ChIP kit (Cat. # 26157) by Thermo ScientificTM was used to carry out the ChIP experiment. A549 lung cancer cells were seeded at a density of 1×10^6 in 100 mm culture dish and cultured in the CO₂ incubator under ambient environment for mammalian cell culture. The cells were harvested on reaching proper confluency. The cells

were crosslinked with molecular biology grade formaldehyde (1%) for 10mins at RT. Crosslinking was halted by adding Glycine solution (1X). The kit protocol was followed to process the crosslinked product. The cellular product was sonicated and the supernatant was collected. For the Input, 10% supernatant was aspirated and stored at -20°C. The digested chromatin present in the supernatant was treated with different antibodies (each antibody of 1µg per reaction) and incubated at 4°C for 5 hours under rotating condition. DNA-protein-antibody complex was pulled down using the magnetic beads solution available in the kit.

ChIP class antibodies were used:

Anti-BG4 (MABE917),
 Anti-NM23-H2 (sc-100400),
 Anti-CNBP (ab83038),
 Anti-Sp1 (ab13370),
 Anti-hnRNPK (ab39975),
 Anti-nucleolin (ab13541),

Anti-IgG was provided with the kit. The purified genomic DNA fragments were used as template for the qPCR amplification of the R1WT GQ region. PowerUp™ SYBR™ Green Master Mix (Cat. # A25742) and the specific designed primers were used to prepare the samples for RT-PCR. QuantStudio 5 thermal cycler was used to perform the PCR cycles.

The PCR run set up:

Initial denaturation step: Heating at 95°C for 10 min

40 cycles:

Initial denaturation at 95°C for 30 secs,
 In appropriate melting temperature for 30 secs
 Extension at 72°C for 30 secs.

Fold enrichment method used to quantify the binding of TFs,

$$\Delta\Delta Ct = Ct(\text{target}) - Ct(\text{IgG})$$

$$\text{Fold enrichment: } 2^{(-\Delta\Delta Ct)}$$

Specific primers to amplify the R1WT region of REX1 gene promoter		
R1 ChIP primers	Forward Primer-TGAGCGTTGACTGACCATTG Reverse Primer-AGCGGAGGAGGCAGTTT	200 bp amplicon length

Data analytics tools.

The G-scores of the selected guanine rich oligonucleotides were predicted by the QGRS predictor software. The molecular skeleton of TMP and BRC were drawn using ChemSketch software. Microsoft Excel Office 2019, SigmaPlot version 12.0, GraphPad Prism version 8.4.2 employed for the statistical analysis, data representation. Adobe Illustrator CS6 was used to prepare the schematic diagrams.

2.5 References

1. Zyner, K. G. *et al.* G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun* **13**, 142 (2022).
2. Alessandrini, I., Recagni, M., Zaffaroni, N. & Folini, M. On the Road to Fight Cancer: The Potential of G-Quadruplex Ligands as Novel Therapeutic Agents. *Int J Mol Sci* **22**, 5947 (2021).
3. Roy, A. *et al.* Curcumin arrests G-quadruplex in the nuclear hyper-sensitive III1 element of c-MYC oncogene leading to apoptosis in metastatic breast cancer cells. *J Biomol Struct Dyn* 1–17 (2021) doi:10.1080/07391102.2021.1940284.
4. Jana, J. & Weisz, K. Thermodynamic Stability of G-Quadruplexes: Impact of Sequence and Environment. *ChemBioChem* **22**, 2848–2856 (2021).
5. Jana, J. *et al.* Chelerythrine down regulates expression of VEGFA, BCL2 and KRAS by arresting G-Quadruplex structures at their promoter regions. *Sci Rep* **7**, 40706 (2017).
6. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research* **36**, 5482–5515 (2008).
7. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90**, 1149–1171 (2008).
8. G-quadruplexes: a promising target for cancer therapy | Molecular Cancer | Full Text. <https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-021-01328-4>.

9. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol* **19**, 414–422 (2009).
10. Prioleau, M.-N. G-Quadruplexes and DNA Replication Origins. *Adv Exp Med Biol* **1042**, 273–286 (2017).
11. Linke, R., Limmer, M., Juranek, S. A., Heine, A. & Paeschke, K. The Relevance of G-Quadruplexes for DNA Repair. *Int J Mol Sci* **22**, 12599 (2021).
12. Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Di Antonio, M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Research* **49**, 8419–8431 (2021).
13. Hou, Y., Guo, Y., Dong, S. & Yang, T. Novel Roles of G-quadruplexes on Enhancers in human chromatin. 2021.07.12.451993 Preprint at <https://doi.org/10.1101/2021.07.12.451993> (2021).
14. Orkin, S. H. & Hochedlinger, K. Chromatin Connections to Pluripotency and Cellular Reprogramming. *Cell* **145**, 835–850 (2011).
15. Kashyap, V. *et al.* Regulation of Stem Cell Pluripotency and Differentiation Involves a Mutual Regulatory Circuit of the Nanog, OCT4, and SOX2 Pluripotency Transcription Factors With Polycomb Repressive Complexes and Stem Cell microRNAs. *Stem Cells Dev* **18**, 1093–1108 (2009).
16. Son, M.-Y., Choi, H., Han, Y.-M. & Cho, Y. S. Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. *Stem Cells* **31**, 2374–2387 (2013).
17. Renčiuk, D. *et al.* G-quadruplex formation in the Oct4 promoter positively regulates Oct4 expression. *Biochim Biophys Acta Gene Regul Mech* **1860**, 175–183 (2017).
18. Scotland, K. B., Chen, S., Sylvester, R. & Gudas, L. J. ANALYSIS OF REX1 (ZFP42) FUNCTION IN EMBRYONIC STEM CELL DIFFERENTIATION. *Dev Dyn* **238**, 1863–1877 (2009).
19. Zeng, Y.-T., Liu, X.-F., Yang, W.-T. & Zheng, P.-S. REX1 promotes EMT-induced cell metastasis by activating the JAK2/STAT3-signaling pathway by targeting SOCS1 in cervical cancer. *Oncogene* **38**, 6940–6957 (2019).
20. Nakanishi, C. & Seimiya, H. G-quadruplex in cancer biology and drug discovery. *Biochemical and Biophysical Research Communications* **531**, 45–50 (2020).
21. Awadasseid, A., Ma, X., Wu, Y. & Zhang, W. G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. *Biomedicine & Pharmacotherapy* **139**, 111550 (2021).
22. Lv, L. & Zhang, L. Characterization of G-Quadruplexes in Enterovirus A71 Genome and Their Interaction with G-Quadruplex Ligands. *Microbiol Spectr* **10**, e00460-22.
23. Sullivan, H.-J. *et al.* Binding of Telomestatin, TMPyP4, BSU6037, and BRACO19 to a Telomeric G-Quadruplex–Duplex Hybrid Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *ACS Omega* **3**, 14788–14806 (2018).

24. Yang, D. & Okamoto, K. Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med Chem* **2**, 619–646 (2010).
25. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Molecular Cancer* **20**, 40 (2021).
26. Brooks, T. A. & Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **1**, 641–649 (2010).
27. Asamitsu, S., Bando, T. & Sugiyama, H. Ligand Design to Acquire Specificity to Intended G-Quadruplex Structures. *Chemistry* **25**, 417–430 (2019).
28. Mendes, E., Aljnadi, I. M., Bahls, B., Victor, B. L. & Paulo, A. Major Achievements in the Design of Quadruplex-Interactive Small Molecules. *Pharmaceuticals (Basel)* **15**, 300 (2022).
29. How long is too long? Effects of loop size on G-quadruplex stability | Nucleic Acids Research | Oxford Academic. <https://academic.oup.com/nar/article/38/21/7858/2411469>.
30. Puig Lombardi, E. *et al.* Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Research* **47**, 6098–6113 (2019).
31. Sun, D. & Hurley, L. H. Biochemical Techniques for the Characterization of G-Quadruplex Structures: EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. *Methods Mol Biol* **608**, 65–79 (2010).
32. Reshetnikov, R. V., Kopylov, A. M. & Golovin, A. V. Classification of G-Quadruplex DNA on the Basis of the Quadruplex Twist Angle and Planarity of G-Quartets. *Acta Naturae* **2**, 72–81 (2010).
33. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat Chem* **5**, 182–186 (2013).
34. Fig. 5 Identification of TFs binding to BG4-IP regions in 93T449 cells.... *ResearchGate* https://www.researchgate.net/figure/identification-of-TFs-binding-to-BG4-IP-regions-in-93T449-cells-A-Consensus-sequences-of_fig4_352683025.
35. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome | Nature Communications. <https://www.nature.com/articles/s41467-021-24198-2>.
36. Da Ros, S. *et al.* G-Quadruplex Modulation of SP1 Functional Binding Sites at the KIT Proximal Promoter. *Int J Mol Sci* **22**, E329 (2020).
37. Fulciniti, M. *et al.* Significant biological role of Sp1 transactivation in multiple myeloma. *Clin Cancer Res* **17**, 6500–6509 (2011).
38. Meier-Stephenson, V. G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys Rev* **14**, 635–654 (2022).
39. Different SP1 binding dynamics at individual genomic loci in human cells | PNAS. <https://www.pnas.org/doi/10.1073/pnas.2113579118>.

40. Lee, J.-J. *et al.* Small molecule activator of Nm23/NDPK as an inhibitor of metastasis. *Sci Rep* **8**, 10909 (2018).
41. David, A. P. *et al.* CNBP controls transcription by unfolding DNA G-quadruplex structures. *Nucleic Acids Research* **47**, 7901–7913 (2019).
42. NM23/NDPK proteins in transcription regulatory functions and chromatin modulation: emerging trends | Laboratory Investigation. <https://www.nature.com/articles/labinvest201798>.
43. Chen, S. *et al.* Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. *Biochim Biophys Acta* **1830**, 4769–4777 (2013).
44. G-quadruplexes are transcription factor binding hubs in human chromatin | Genome Biology | Full Text. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02324-z>.
45. Sengupta, P., Bose, D. & Chatterjee, S. The Molecular Tête-à-Tête between G-Quadruplexes and the i-motif in the Human Genome. *ChemBiochem* **22**, 1517–1537 (2021).
46. Thakur, R. K. *et al.* Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res* **37**, 172–183 (2009).
47. Rezzoug, F., Thomas, S. D., Rouchka, E. C. & Miller, D. M. Discovery of a Family of Genomic Sequences Which Interact Specifically with the c-MYC Promoter to Regulate c-MYC Expression. *PLOS ONE* **11**, e0161588 (2016).
48. MatInspector: Search for transcription factor binding sites. https://www.genomatix.de/online_help/help_matinspector/matinspector_help.html.
49. QGRS Mapper | Analyze and Search. <https://bioinformatics.ramapo.edu/QGRS/analyze.php>.
50. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C. & Basu, S. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Research* **40**, 4137–4145 (2012).
51. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545–1614 (2009).
52. Vanommeslaeghe, K. & MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).
53. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135–2145 (2013).
54. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
55. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802 (2005).

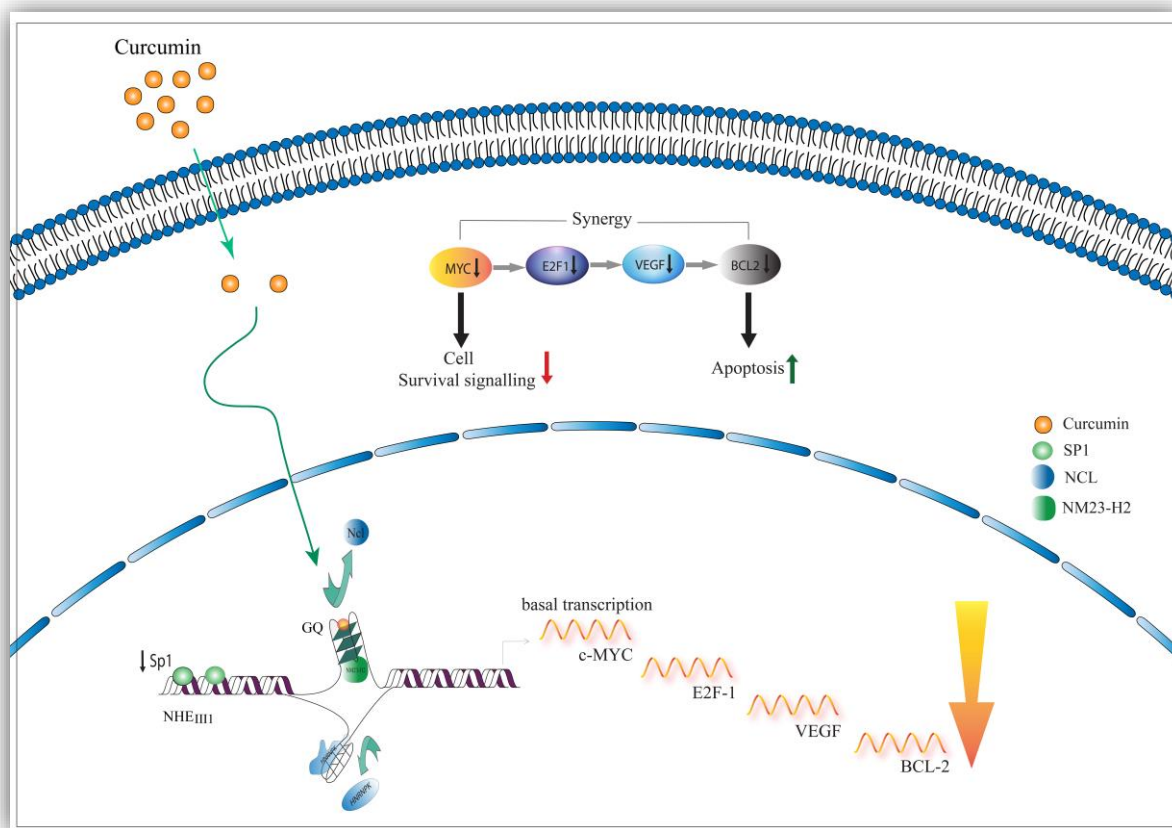
56. Darden, T., Perera, L., Li, L. & Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **7**, R55-60 (1999).
57. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187–217 (1983).
58. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
59. Hoover, null. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys* **31**, 1695–1697 (1985).
60. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).
61. Lavrischeva, E. M. Scientific Basis of System Programming. *JSEA* **11**, 408–434 (2018).



Chapter 3

Curcumin induces apoptotic death in TNBC cells by downregulating c-MYC oncogene through targeting G-quadruplex in the Nuclear Hypersensitive Element III1.

This chapter is based on the article 'Curcumin arrests G-quadruplex in the nuclear hyper-sensitive III1 element of c-MYC oncogene leading to apoptosis in metastatic breast cancer cells' *J Biomol Struct Dyn.* 2021 Jun 30;1-17. doi: 10.1080/07391102.2021.1940284.



3.1 Introduction

For a long time, Myc has been infamous in the scientific community as an oncogene underpinning most human cancers. MYC gene is known to be amplified and overexpressed in malignant cells. Several studies have revealed that MYC amplification is linked to the

development of breast tumor and metastatic progression, highlighting MYC as an enticing target in cancer therapeutics¹⁻³. Since Myc protein is essential in embryonic development and for other normal functioning of proliferative tissues, its complete inhibition results in fatal side effects⁴. A promising strategy for combating cancer is to delineate *c-MYC* gene-targeted drug molecule to rehabilitate MYC expression back to its basal-level^{5,6}. MYC gene expression is stringently regulated by a composite mechanism that involves the cooperative participation of enhancers, several promoters, and multiple TFs⁷. The main TFs participating in the tight regulation of MYC transcription are specificity protein (Sp1), cellular nucleic acid binding protein (CNBP), heterogeneous nuclear ribonucleoprotein K (hnRNPK), FBP, CTCF, nucleolin (NCL), and NM23-H2⁸. Two DNA elements [CT element and Far upstream element (FUSE)] present in the PPR of *c-MYC* are critical players in gene regulation⁸. The NHE_{III1} region (also called CT element) located between -142 to -115 basepair upstream of the P1 gene promoter of *c-MYC*^{9,10} is a crucial DNA element that work in conjunction with DNA binding TFs to regulate *c-MYC* transcription⁸. This regulatory region contains a 27-nucleotide long guanine rich sequence (*Pu-27*) in the non-coding strand that folds into a GQ¹¹. While the complementary cytosine rich strand is reported to conform into higher order i-motif (iM) structure¹¹. *Pu-27* acts as a regulatory switch that is proclaimed to control more than 80% of *MYC* transcription^{8,11}. In the previous chapter, I discussed about the importance of GQ as a regulatory element in controlling the expression of REX1 gene, upregulation of which induces pluripotent characteristic into cancer cells. In this chapter, I shall be discussing about the critical biological regulatory role of G quadruplex in controlling the expression of the proliferative marker gene *c-MYC*.

GQs are non-canonical secondary structures composed of G-quartet plates stacked *via* π - π interaction. Each plate consists of 4 Gs stabilized by eight Hoogsteen H-bonds (N2 bonds with

N7 and N1 bonds with O6) organized in a coplanar fashion^{12,13}. The ring of hydrogen bonds coordinates with a central cation ($K^+/Na^+/NH_4^+/Pb^{2+}$) to render stability to the quadruplex structure¹³. Evidence of the existence of GQs inside cell and the critical role in controlling gene expression have intensified research on GQs worldwide^{14–16}. In healthy proliferating cells, *Pu-27* in its folded conformation in the NHE_{III} downregulates *c-MYC* expression¹⁷. X-ray structure (PDB ID: 6AU4) and resolved NMR structure of MYC G-rich promoter sequence (PDB ID: 1XAV) have validated a parallel GQ topology conformed by MYC G-quadruplex⁸. Researchers have designed several small-molecule compounds specifically to target the G-quadruplex. Reported small molecules like TmPyP4¹⁸, Telomestatin¹⁹, Chelerythrine²⁰, Berberine²¹, Sanguinarine²², Pyridostatin (PDS)²³, BRACO-19²⁴, PhenDC3²⁵ have planar π -aromatic aspect that interacts with GQ *via* stacking interaction^{13,17,26,27}. Although these compounds reportedly stabilized GQs and downregulated genes overexpressed in cancer, these molecules are found to be hazardous to healthy cells as well. These molecules act by entirely shutting down the transcription of targeted genes resulting in adverse consequences.

In our investigation, we have employed the use of the natural compound curcumin (turmeric) to investigate its binding interaction with *Pu-27* GQ. Curcumin (CUR) is observed to share structural features with the established GQ-interacting molecules. Literature survey revealed evidence on telomeric GQ-CUR interaction under molecular crowded condition³¹. The chemical formula of CUR: $C_{21}H_{20}O_6$ ^{13,28}. It is also called diferuloylmethane, a hydrophobic polyphenol integrant extracted from the rhizomes of *Curcuma longa* with potential GQ binding ability. Curcumin is a compound that consists of two aromatic rings. These rings hold *o*-methoxy phenolic groups and are connected by a seven-carbon linker that consists of an α , β -unsaturated β -diketone moiety²⁸. Because of its unique structure and functional groups, curcumin has efficient electron transfer capability and several π electrons that can conjugate between the two phenyl rings^{8,29}. The alkoxy-side chains of curcumin provide electron-rich

aromatic surface, while the flat aromatic rings connected by internal H-bond network are suitable for interaction with GQ³⁰.

3.1.1 The mechanosensor mechanism controlling GQ/iM dynamicity at the NHE_{III} element in the gene promoter of *c-MYC*

Pu-27 is a negative regulator of *c-MYC* expression, ensuring optimal level of transcription when normal cells are not in a proliferative condition¹⁷. In the cellular or chromatin state, the NHE_{III} maintains the *c-MYC* promoter in a dynamic equilibrium condition i.e., intermittent between transcriptionally active and silenced forms. This dynamic equilibrium depends on the evolution and resolution of GQ and iM at the NHE_{III} element³². During transcription, local unwinding of the dsDNA is induced by added negative superhelicity which in turn facilitates iM and GQ formation on the C-rich and G-rich strands respectively⁶. Sutherland et.al. have shown the mutual exclusive evolution of GQ and iM, i.e., when GQ forms, the opposite Cytosine-rich strand remains in an unfolded state, and while iM forms the opposite G-rich stretch is maintained in an unfolded state³⁴. GQ in its folded conformation acts as the negative regulator of transcription, while iM formation is correlated with an active transcriptional state³³ [Figure 1]. Thus, the evolution and resolution of these higher-order conformations in the NHE_{III} element of *c-MYC* gene promoter is responsible for the active or silenced state of transcription. The formation and resolution of these non-canonical DNA elements are tightly controlled, not just by the transcription-induced alteration in the negative superhelicity but also by the active participation of regulatory TFs³³. Sp1 is a well-known TF that binds to G-rich DNA regions and induces negative supercoiling that leads to the evolution of ssDNA, allowing the folding of a GQ or an iM according to the cell's requirement³³. The transcription of genes with TATA-less promoters is dominantly controlled by Sp1³³. hnRNPK is a reported TF that interacts with C-rich DNA elements having a higher propensity to conform into an i-motif, hence proving its vitality in transcriptional regulation³⁵. The two KH domains of hnRNPK

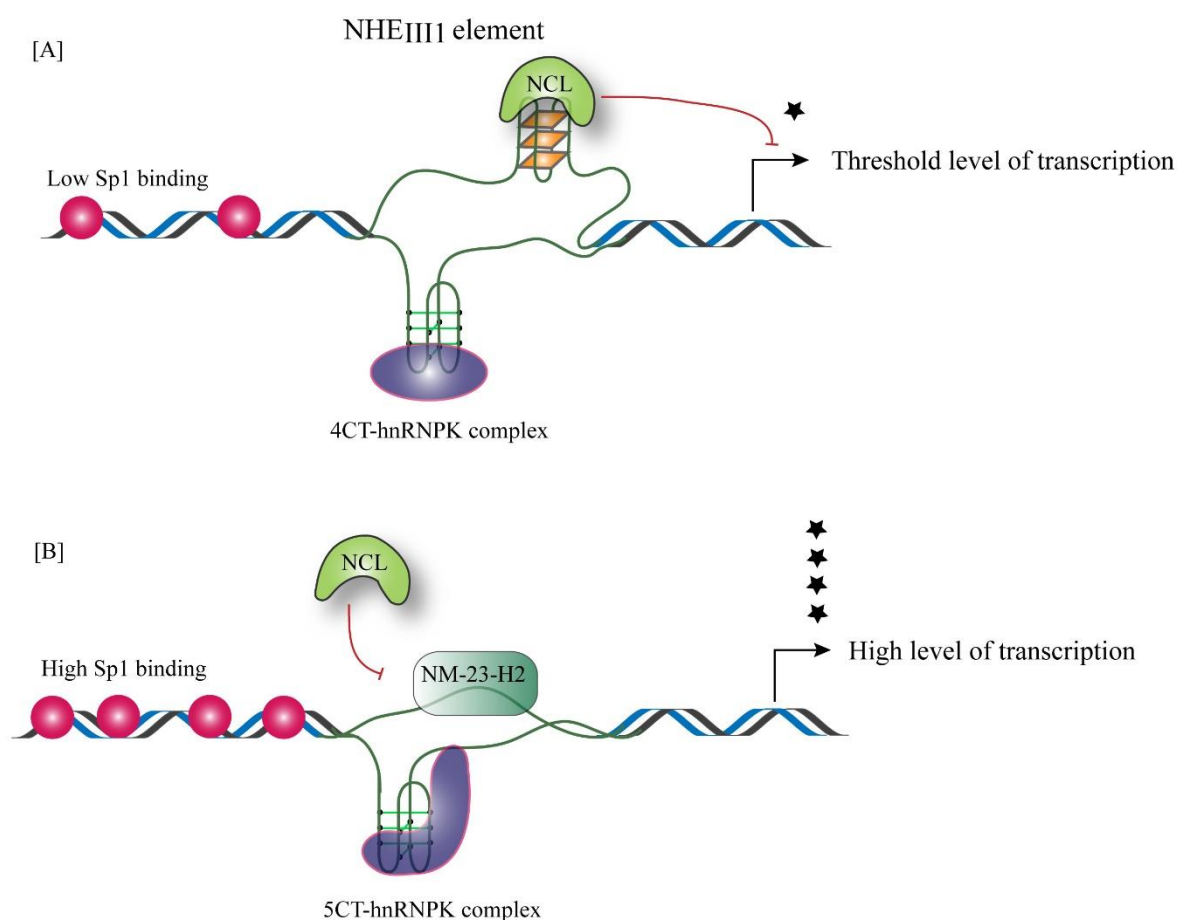


Figure 1. [A] Schematic diagram representing the NHE_{III1} element. Low Sp1 binding at the NHE_{III1} element results in weak 4CT-hnRNP complex formation and allows NCL-GQ stable structure formation which in turn limits the transcription of c-MYC gene. [B] High level of Sp1 binding causes negative superhelicity that facilitates strong 5CT-hnRNP complex formation, while inhibiting the stable formation of GQ. NM-23-H2, hnRNP and Sp1 coordinates to actively transcribe c-MYC gene at a higher rate.

initially recognize and bind to the 4CT sequence of the NHE_{III1} element. The evolution of more ssDNA exposes the 5CT element, which the hnRNP recognizes and associates tightly to form a thermodynamically stable binary complex that activates transcription^{33,35}. Reduction in Sp1 occupancy level lowers the negative superhelicity stress, low torsional stress is insufficient to expose the 5CT element for binding by hnRNP. hnRNP-4CT complex is a weaker interaction compared to hnRNP-5CT complex [Figure 1A]. Thus, the hnRNP is displaced

from its complexation with 4CT element by nucleolin (NCL) TF. Recruitment and binding of NCL to the GQ, stabilizes the structure and results in *c-MYC* transcriptional repression³⁴. Thus, high Sp1 occupancy at the promoter region induces greater negative superhelicity that exposes the 5CT element allowing stable hnRNP-K-5CT complex formation that overcomes the repressive action of the GQ and results in full-scale transcriptional activation [Figure 1B]. The mutually exclusive formation of GQ or iM *via* the interplay of TFs especially Sp1 and hnRNP-K dictates *c-MYC* transcription³³. There are reports of NCL supporting the progression of cancer metastasis by promoting angiogenesis *via* the upregulation of VEGF and also help in evasion of apoptotic cell death *via* the stabilization of BCL2 transcript^{36,37}. *In-vivo* application of anti-NCL treatment led to reduction in tumor vasculature in neoplastic cells *via* the suppression of angiogenesis^{36,37}. On the contrary, NCL is reported to stabilize the silencer element GQ in the NHE_{III1} element that results in the downregulation of *c-MYC* oncogene³³. The NM-23 family proteins assist in several molecular signalling pathways. NM-23 H1/H2 majorly play metastasis suppressor role³⁸. There are evidences of the enhancement effect of CUR on NM-23 expression³⁹. RNAi mediated knockdown experiments confirmed the effect of NM-23-H2 in activating *c-MYC* transcription. NM-23-H2 is reported to destabilize or unfold the GQ structure³⁸. *Pu-27* GQ and iM are the docking platforms for NM-23-H2, NCL, and hnRNP-K. These TFs share overlapping DNA-binding sequences at the NHE_{III1} region. There is a constant dynamicity between the recruitment and detachment of these TFs across the *Pu-27* element. For successful design and development of an anti-*MYC*-anti-cancer drug, it is imperative to understand how these TFs coordinate at NHE_{III1} to regulate *c-MYC* expression in presence of ligand.

Curcumin exhibits antiproliferative⁴⁰, pro-apoptotic⁴¹, antiangiogenic⁴², antitumorigenic⁴³, antioxidant⁴⁴, and anti-inflammatory⁴⁵ properties. Curcumin (CUR) is reported to negatively regulate a number of TFs, growth factors, inflammatory cytokines, protein kinases, and various

[A]

Pu-27 TGGGGAGGGTGGGGAGGGTGGGGAAGG

[B]

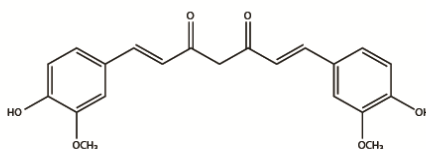


Figure 2. [A] *Pu-27* G quadruplex sequence of *c-MYC* [B] Curcumin (*CUR*) molecule.

other oncogenic factors^[16-19]. However, the bioavailability of *CUR* in the human system is inadequate to exert its anti-cancerous effect in metastatic patients. *CUR*

failed to qualify as an anticancer drug molecule. In this chapter, we have shown the utility of *CUR* as a model drug skeleton. Our study examined the binding interaction between *CUR* and the G-quadruplex of the wildtype *c-MYC* promoter. Through various biophysical experiments, we have evaluated the interaction between *CUR* and *Pu-27* at the atomic-level. Next, we studied the mode of action of *CUR* in downregulating the *c-MYC* oncogene. Through several *in-vitro* assays, we studied the binding property of *CUR* with *Pu-27* in cancer cells and tried to unravel the mechanism by which *CUR* drives *c-MYC* silencing in aggressive cancers with an upregulated *c-MYC* oncogene. We have decrypted the difference between *CUR*-mediated *c-MYC* suppression and other small molecule-mediated downregulation of *c-MYC* that makes *CUR* non-toxic to normal healthy cells. Upregulated *c-MYC* and *BCL2* synergize to aggressively accelerate breast cancer metastasis. In our study, we have also reported an alternative apoptotic signalling pathway [*c-MYC-E2F-1-BCL-2* axis] that gets activated in cancer cells upon *CUR* treatment. Our research could help design selective anti-metastatic drug targeting GQ, based on a natural compound that shall minimize the risks of fatal off-target effects.

3.2 Results and Discussions

In this chapter, we report the binding interaction of CUR with *Pu-27*, the wildtype GQ positioned in the proximal promoter region (PPR) of the *c-MYC* gene. The *Pu-27* GQ sequence and the molecular structure of curcumin are provided in [Figure 2](#). We have utilized CUR as a model-drug ligand and studied the interaction efficiency of this small molecule and the mechanism behind CUR-mediated *c-MYC* gene downregulation.

3.2.1 Significant binding interaction between curcumin (CUR) and *c-MYC* GQ

We have employed fluorescence spectrophotometry to study the interaction between curcumin (CUR) and the wildtype G4sequence (*Pu-27*) of the NHE_{III1} element of *c-MYC*. CUR itself is a non-fluorescent molecule (CUR exhibits minimal fluorescence in an aqueous solvent). However, it fluoresces upon binding to a macromolecule. When CUR is excited at 427 nm, it displays fluorescence at around 540 nm. Titration of CUR with incremental oligo addition resulted in a 3nm blue shift [540 to 537nm] along with a consistent enhancement of fluorescence intensity (FI) [[Figure 3A](#)]. The observed intensification of fluorescence intensity is indicative of the enhanced quantum efficiency of CUR. The fluorescent spectral characteristic suggests the physical interaction of CUR with a non-polar segment of the quadruplex³¹. The binding constant of CUR with the GQs selected in this study was calculated using the fractional change in the FI. We have used the modified version of the Scatchard model (multiple binding sites) to calculate the K_d of GQ-CUR complex⁴⁶. The K_d value of *c-MYC*-CUR complex was calculated to be 5.84 μ M [[Figure 3B](#)]. A comparative fluorescent spectroscopic study was parallelly conducted with CUR and three different reported GQs (VEGF, KRAS, and BCL2 promoter regions). We observed a low binding affinity of curcumin with these reported GQs compared to that of *Pu-27* GQ [[Figure 4C](#)] indicating the specificity of the interaction.

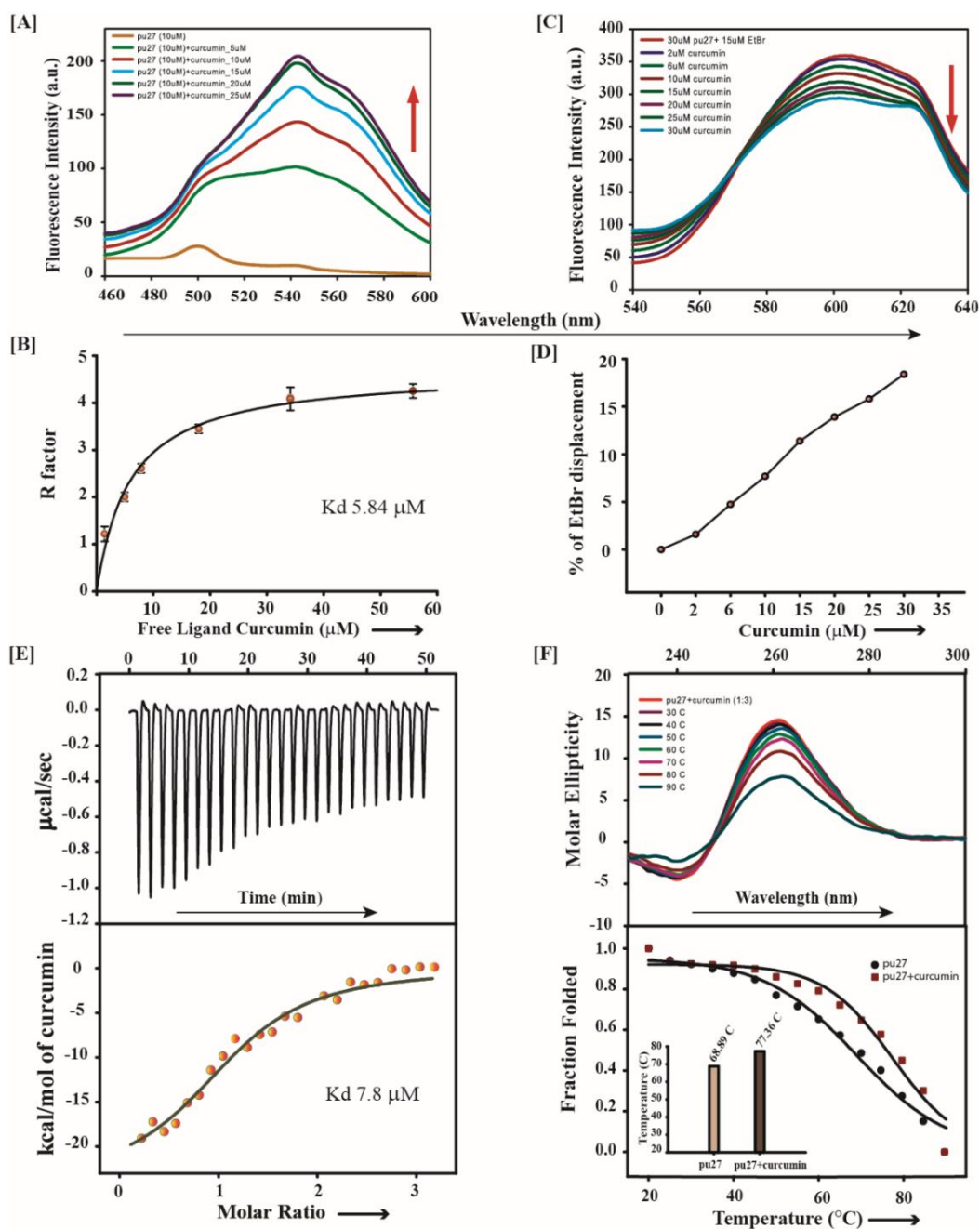


Figure 3. [A] Upon titration with incremental increases of Pu-27, the observed fluorescence emission spectra of Curcumin (20 μM) [B] Graph plot of r factor versus free CUR. [C] EtBr displacement assay of Pu-27-EtBr complex titrated with incremental concentrations of CUR and the observed lowering in EtBr fluorescence. [D] Percentage of EtBr displacement plotted as a function of CUR concentration. [E] The ITC profile of CUR binding with Pu-27 is displayed in two panels. The above panel displays the thermogram of the Pu-27-CUR complex formation in $\mu\text{cal/sec}$ versus time in minutes. The panel below shows the binding isotherm resulting from the integrated peak area plotted against the molar ratio. [F] CD titration and melting assays of the Pu-27-CUR complex. The CD spectra of the Pu-27-CUR complex at increasing temperature intervals are shown in the upper panel. The melting profile of the free Pu-27 and the GQ-CUR complex are shown in the lower panel (Pu-27: CUR = 1:3). The inset bar plot represents the stabilization of Pu-27 by curcumin addition as observed by an increase in the T_m values.

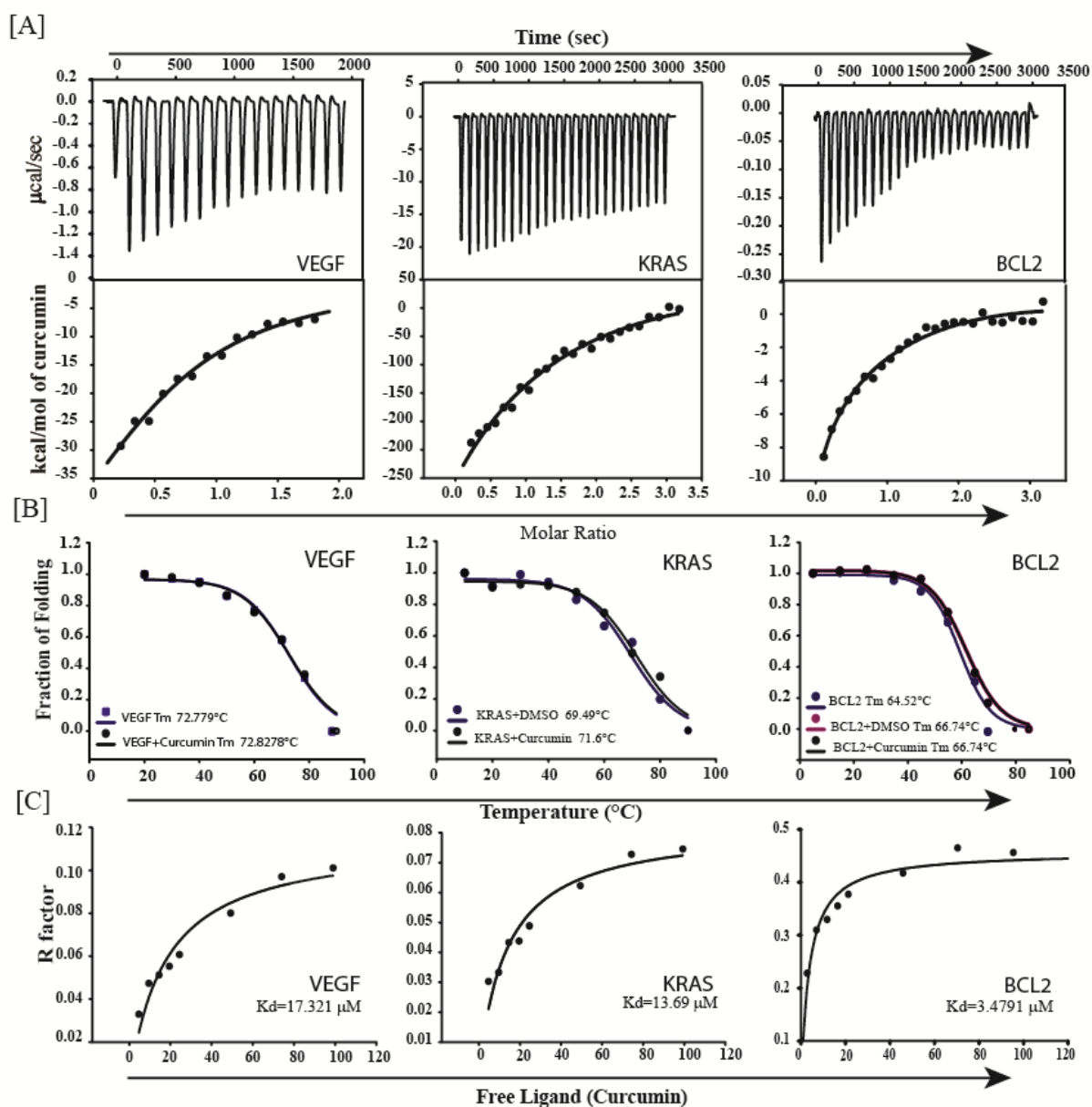


Figure 4 [A]. ITC profile of CUR with VEGF, KRAS, BCL-2 G quadruplex sequences. The upper panel on the display shows the isothermal plot of the CUR-GQ complex formation, with the y-axis showing $\mu\text{cal/sec}$ and the x-axis showing time in minutes. On the other hand, the lower panel shows the integrated binding isotherm, which is generated from the integrated peak area as a function of molar ratio. **[B].** CD melting profile of free quadruplexes and CUR-GQ complex of VEGF, KRAS and BCL-2 respectively with mentioned change in melting temperatures. **[C].** The graph plot of r factor versus free curcumin to derive the binding constant of interaction of CUR with VEGF, KRAS, and BCL-2 respectively.

3.2.2 Curcumin and *Pu-27* interacts via stacking mode of interaction

We have used EtBr displacement assay to study the binding mechanism of CUR. According to reports, ethidium bromide interacts with GQ-DNA through end-stacking mode of interaction⁴⁷. We gradually titrated the EtBr-GQ complex with CUR and examined the change in fluorescence intensity [FI] within the range of 540 nm to 640 nm. Initially, we observed a gradual enhancement of fluorescence of EtBr upon complex formation with an increment in GQ concentration. This indicates intercalative mode of interaction of EtBr with GQ. Further, upon titrating the EtBr-GQ complex with CUR, we observed a drop in fluorescence intensity of EtBr. This indicates plausible stacking of the CUR molecule on *Pu-27*-GQ structure by displacing EtBr from the complex [*Figure 3C and 3D*]. Thus, we evidenced that curcumin undergoes end stacking mode of interaction with *Pu-27* GQ with a higher propensity than EtBr. The obtained results have further been correlated with molecular docking results.

3.2.3 Enhancement of thermal stability of the *c-MYC* quadruplex conformation upon complexation with CUR

The CD spectrophotometer is a precise instrument to examine ligand-induced conformational changes in nucleic acids. Earlier reports confirmed the acquisition of a parallel GQ conformation by *Pu-27* in an aqueous environment supplemented with potassium (K⁺) ions. The characteristic CD spectra of *Pu-27* has a positive hump with maxima at 260nm along with a trough with negative maxima at around 240nm^{48,49}. To investigate any structural change in *Pu-27* upon interaction with CUR, we have evaluated the CD spectral information of the GQ in presence of CUR and without CUR. Upon titrating *Pu-27* (20μM) with incremental addition of CUR, we observed insignificant variation in the characteristic GQ spectra up to an oligo-to-curcumin ratio of 1:3. The spectral intensification at 260nm in CD with a stable parallel GQ spectrum indicates unflurried or unaltered GQ structure after interaction with CUR. The ellipticity enhancement at 260nm observed indicated $\pi-\pi^*$ transition to be the more preferred

electronic transition that lowered the conformational-dynamics of the GQ when complexed with CUR. For comparison of Pu-27-GQ stability with and without CUR, we carried out thermal denaturation assay. The melting curves are obtained by plotting the change in molar ellipticity at 260nm against temperature. We detected $\sim 9^{\circ}\text{C}$ enhancement in melting temperature of *Pu-27* when complexed with CUR compared to the T_m value of the GQ *in solution* without the presence of the ligand [Figure 3F]. We also performed CD melting experiments of CUR complexed with each of the reported KRAS, BCL2, and VEGF G-quadruplexes for a comparative analysis, where we did not observe significant stability enhancement of these GQs in presence of CUR [Figure 4B]. Thus, the CD-melting data, clearly evidenced the enhanced thermal stability of *Pu-27* GQ by curcumin. We obtained the melting temperature of *c-MYC*-CUR complex: $T_m = 77.35 \pm 3.194^{\circ}\text{C}$ which is significantly higher than that of the free *c-MYC* quadruplex ($T_m = 68.92 \pm 2.7984^{\circ}\text{C}$) [Figure 3F]. We therefore speculate that curcumin preferentially binds to *Pu-27* GQ.

3.2.4 Thermodynamic feasibility of Curcumin-*c-MYC* GQ interaction

To examine the thermodynamics of the *Pu-27*-CUR interaction, we conducted an isothermal titration calorimetric study (ITC). The equilibrium dissociation constant (K_d), enthalpic change (ΔH), entropic change (ΔS), and change in Gibbs free energy (ΔG) were measured. The change in the thermodynamic parameters as determined from the isothermal calorimetric analysis are provided in Table 1. For a comparative study, we have conducted ITC profiling of CUR with VEGF, KRAS, and BCL2 GQs, where we observed a low binding affinity of curcumin for the mentioned GQs [Figure 4A]. The ITC profile with the thermogram of *Pu-27*-CUR interaction is provided in [Figure 3E]. The thermogram clearly indicates that *Pu-27*-CUR binding interaction results in a substantial quantity of heat change. Curcumin binds strongly to *Pu-27* (K_d value for the dissociation constant, $7.8 \mu\text{M}$), indicating the participation of hydrogen bonds

<i>Pu-27-CUR interaction</i>		
Model	Variables	Values
Independent	K_d (M)	7.82E-06
	ΔS (J/mol.K)	16.57
	ΔH (kJ/mol)	-23.91
	n	1.17
	ΔG (kJ)	-24.324

Table 1. The thermodynamic properties of the interaction between CUR and Pu-27.

and Van der Waals force of attraction. Both these forces are involved in stabilization of the complex that results in a decreased RMS motion. From the ITC result, we obtained a positive value for entropy change (ΔS) and a negative value for Gibbs free energy change (ΔG) [Table 1]. This indicates that water molecules are pulled out of the Pu-27 GQ receptor upon interacting with curcumin. The negative enthalpy of *c-MYC* GQ-CUR interaction suggests stacking mode of interaction. Negative free energy change (ΔG) exemplifies the spontaneity of the interaction. Compared to curcumin's reported binding affinity for KRAS G-quadruplex⁵⁰, the strength of its interaction with *c-MYC* Pu-27 GQ is notably greater. The efficient interaction between CUR and Pu-27 agrees well with the proven binding of CUR to telomeric GQ. and other established GQ interacting small molecules³¹. The high binding affinity of curcumin for Pu-27 provides a potential lead for the designing of *c-MYC*-targeted anticancer ligands. Our biophysical studies provide evidence that CUR interacts firmly with the native Pu-27 *c-MYC* G-quadruplex. Using all-atom molecular modelling, the precise mechanism through which curcumin interacts with MYC GQ has been further investigated.

3.2.5 Atomic level investigation of curcumin-*c-MYC* GQ interaction by Molecular Dynamics (MD) simulation

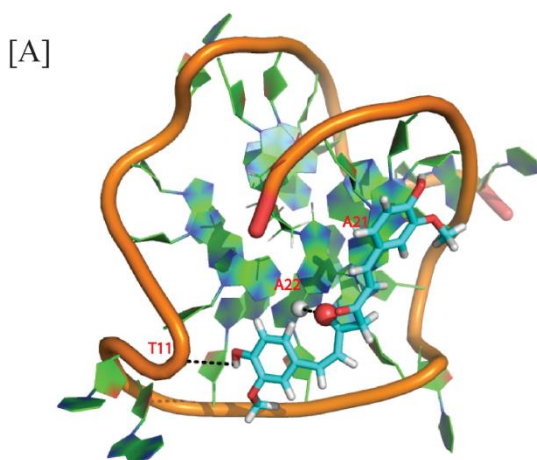


Figure 5. [A] Docking and simulation of *c-MYC*-GQ-curcumin complex

Fluorescence spectroscopic results and the EtBr displacement analysis suggested CUR interacts with MYC-GQ via stacking mode. Molecular docking was done in Schrodinger to find the details of the atomic interaction. The docking results revealed CUR to bind to the 3'-end of *MYC* GQ through the stacking mode of interaction [Figure 5A]. CUR is a beta-diketone methane molecule with two hydrogens replaced by feruloyl groups⁵¹.

With the aid of the hydroxyl and methoxy groups, one of the feruloyl groups attaches to Adenine 21 (A21). The other terminal of the CUR molecule interacts with Guanine14 (G14) with the assistance of hydroxyl in the 2nd feruloyl group. G14 is a participant in the middle tetrad plate of the quadruplex. This feruloyl group also forms a strong contact by tethering to the Guanine 10 (G10). The observed binding results demonstrate that CUR stacks on top of the G-quartet and stabilizes the tetrad structure. Our previous spectroscopic and calorimetric results are supported by docking investigations that have a good correlation showing the stabilizing effect of CUR on *c-MYC* GQ.

3.2.6 Transcriptional downregulation of *c-MYC* by curcumin in metastatic cancer cell line.

Curcumin treatment downregulates abnormally expressed genes in MDA-MB-231 cells^{52,53}. Since we observed CUR to efficiently interact with the GQ in the NHE_{III1} element of *c-MYC* gene promoter *in solution*, we speculated that CUR treatment in metastatic cell line might negatively affect the expression of *c-MYC*. Cell line MDA-MB-231 is reported with

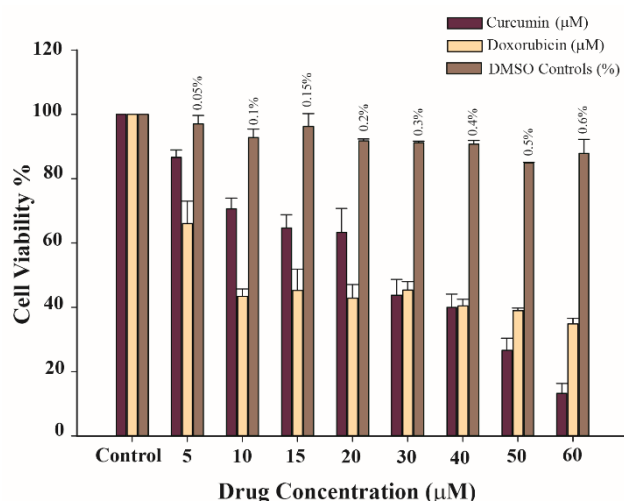


Figure 6. MTT cell viability assay performed on Curcumin treated MDA-MB-231 cell line. Curcumin concentration (5µM to 60 µM) was selected for treatment, Doxorubicin chosen as positive control with DMSO as solvent in the negative-controls with a range of DMSO percentage ranging from 0.05 to 0.6%. The treated values normalized with the control group.

upregulated *c-MYC* gene⁵⁴. We treated metastatic breast cancer cells (MDA-MB-231) with a gradient of CUR dosage and performed MTT assay to select the efficient dose of CUR for further analysis-based experiments. Almost 50% of proliferation was inhibited at around 30 µM treatment

concentration of CUR [Figure 6]. We have performed semiquantitative RT-PCR to analyze the change in expression of *c-MYC* at the mRNA level. At 20 µM of CUR treatment, the mRNA expression level of *c-*

MYC decreased by 1.4-fold (30%). At 30 µM of CUR dose, the expression of *c-MYC* decreased by 1.95-fold, which is equivalent to 48.75%. [Figure 7A]. Our results suggest, the detrimental effect of CUR on cancer cell proliferation may be due to the repression of *c-MYC*. We speculated that CUR binds to the *Pu-27* GQ and stabilizes the silencer switch that results in *c-MYC* downregulation. At higher doses of CUR treatment (40µM and 60µM), *c-MYC* expression does not completely shut down, instead the basal level of expression got restored [Figure 7A]. CUR treatment did not have any effect of the expression of the positive control gene GAPDH.

3.2.7 Curcumin-*Pu-27* GQ interaction mediates *c-MYC* transcriptional repression

To confirm the mechanism by which curcumin mediates *c-MYC* downregulation *in vitro*, we conducted luciferase assay. Our speculation about CUR's selective binding to *Pu-27* in *c-MYC* promoter *in cellulo* shall result in the downregulation of hRlucCP expression, since MYC GQ is reported as a silencer regulatory switch^{8,11}. The deletion clone with the GQ sequence deleted

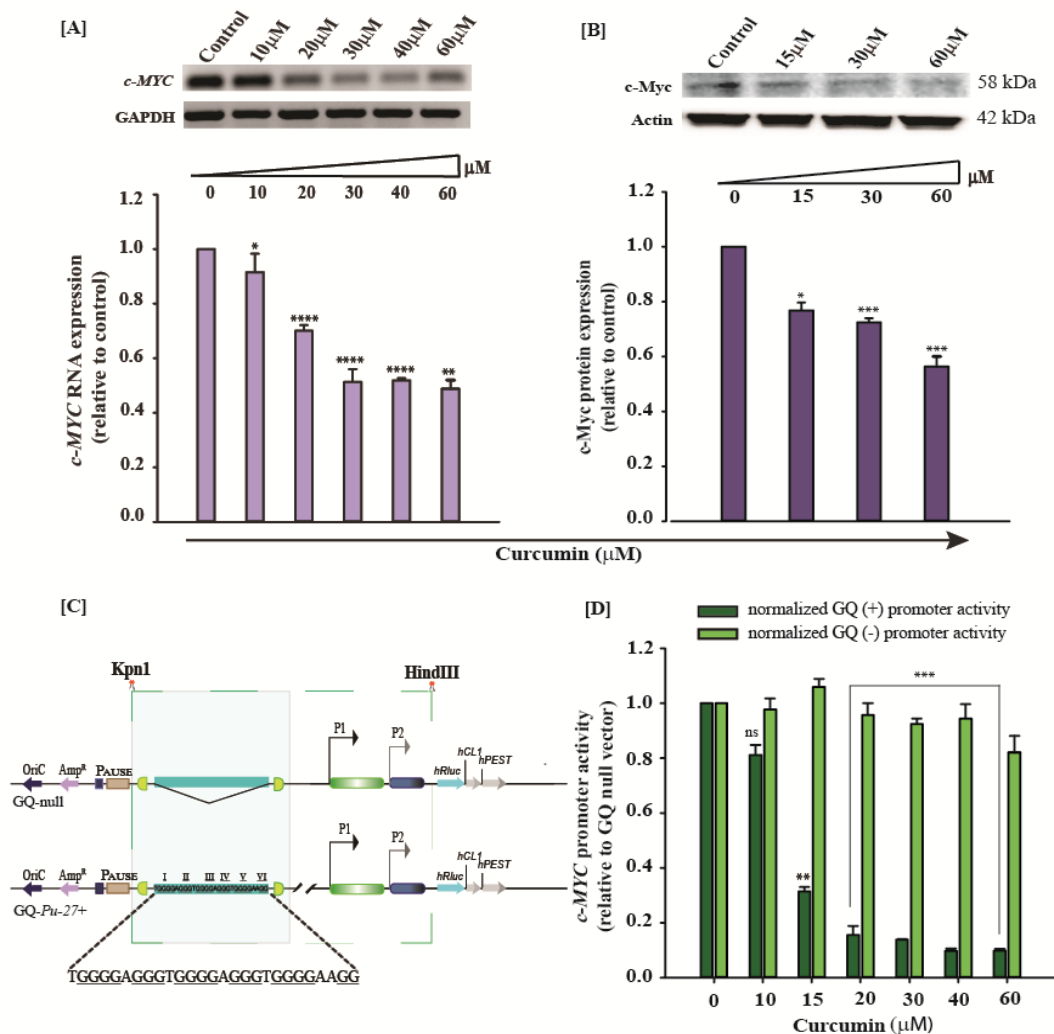


Figure 7. **[A]** Semiquantitative RT-PCR data showing CUR mediated transcriptional repression of c-MYC in MDA-MB-231 cell line after a gradient of CUR treatment for 24 hours. Bar plot data of the agarose gel bands' densitometric analysis showing the relative RNA expression of c-MYC upon CUR treatment (\pm SE) normalized to the control in the panel below. **[B]** After treating MDA-MB-231 cells with CUR (15, 30, 60 μ M), Western blot analysis was done to check Myc protein levels. The lower panel histogram plot shows Myc protein level quantified by densitometric analysis and normalized to the Control set. **[C]** This diagram illustrates a reporter luciferase construct, using the pGL4.72 (hRlucCP) vector with a c-MYC promoter sequence (P1 and P2) inserted, both with and without an upstream Pu-27 GQ sequence containing six G-tracts (I-VI) ahead of the hRluc coding region. The promoter sequences are cloned into restriction sites Kpn-I and Hind-III, with or without the wild-type GQ scaffold. The diagram includes abbreviations for the important components: hRluc (Renilla luciferase gene), oriC (origin of replication), hCL1 and hPEST (protein-destabilizing sequences), SV40 (simian virus 40 polyadenylation signal cassette), and AmpR (ampicillin resistance gene). **[D]** In the study, the promoter activity of c-MYC oncogene was evaluated using the Dual luciferase assay. The relative promoter activity was determined by comparing Fluc/Rluc values to those of cells that were transfected with the GQ-null vector after treatment with CUR. The results showed a significant reduction in luciferase activity for the constructs that contained the quadruplex motif upstream of the promoter compared to the Pu-27 deleted construct after CUR treatment. Error bars used in the histograms (Figure 3 A, B and D) represent means of normalized values \pm SE from triplicate set of data. Asterisks (*) indicate statistical significance as determined from paired t-test (* indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$, **** indicates $P < 0.0001$), denoting significant difference in RNA and Protein level as well as a significant difference in promoter activity upon CUR treatment.

from the inserted promoter region would exhibit higher promoter activity independent of CUR treatment [Figure 7C]. We analyzed the extent to which c-MYC promoter activity is inhibited by treatment with CUR. This analysis takes into account that CUR stabilizes Pu-27 GQ in the c-MYC promoter, which represses the expression of the renilla luciferase gene in GQ+ plasmid transfected cells compared to GQ-/ GQ null plasmid transfected cells. Our observations show significant inhibition of luciferase activity in MDA-MB-231 cells transfected with GQ+ luciferase clones [Figure 7D]. Our results confirm the efficient targeting of MYC promoter GQ by curcumin *in cellulo* that leads to downregulation of the gene. We observed *c-MYC* promoter activity to reduce by 3.5-fold (71.7%) at 20 μ M of CUR treatment, while the promoter activity curtailed by 3.65-fold (72.62%) at 30 μ M CUR treatment compared to the treated GQ-promoter activity. Thus, curcumin binds and stabilizes the silencer regulatory switch (*Pu-27* GQ) in the cellular environment.

3.2.8 Curcumin represses c-Myc protein expression in metastatic cancer

At 15 μ M, 30 μ M and 60 μ M of CUR treatment dosages, the translation of c-Myc was reduced by 1.3, 1.4, and 1.8-fold respectively, when compared to the expression change of Actin (Positive Control) [Figure 7B]. We measured the expression of Actin for each treatment as an internal control. We observed Actin to remain constant for the treatment course, suggesting that curcumin selectively suppresses c-Myc at the protein level.

3.2.9 Alteration of TF recruitment and stabilization of endogenous *Pu-27* by curcumin triggers *c-MYC* downregulation in metastatic cancer

An *in-depth* analysis of the mechanism behind curcumin-mediated *c-MYC* transcriptional regulation was carried out. We monitored the binding of selected TFs: NM-23-H2, Sp1, hnRNPK, and NCL in the NHE_{III1} region by conducting chromatin immunoprecipitation (ChIP) experiment [Figure 8A]. We observed substantial reduction in the occupancy level of hnRNPK and Sp1 TFs at the NHE_{III1} region on CUR treatment [Figure 8B and 8C]. 2.3-fold

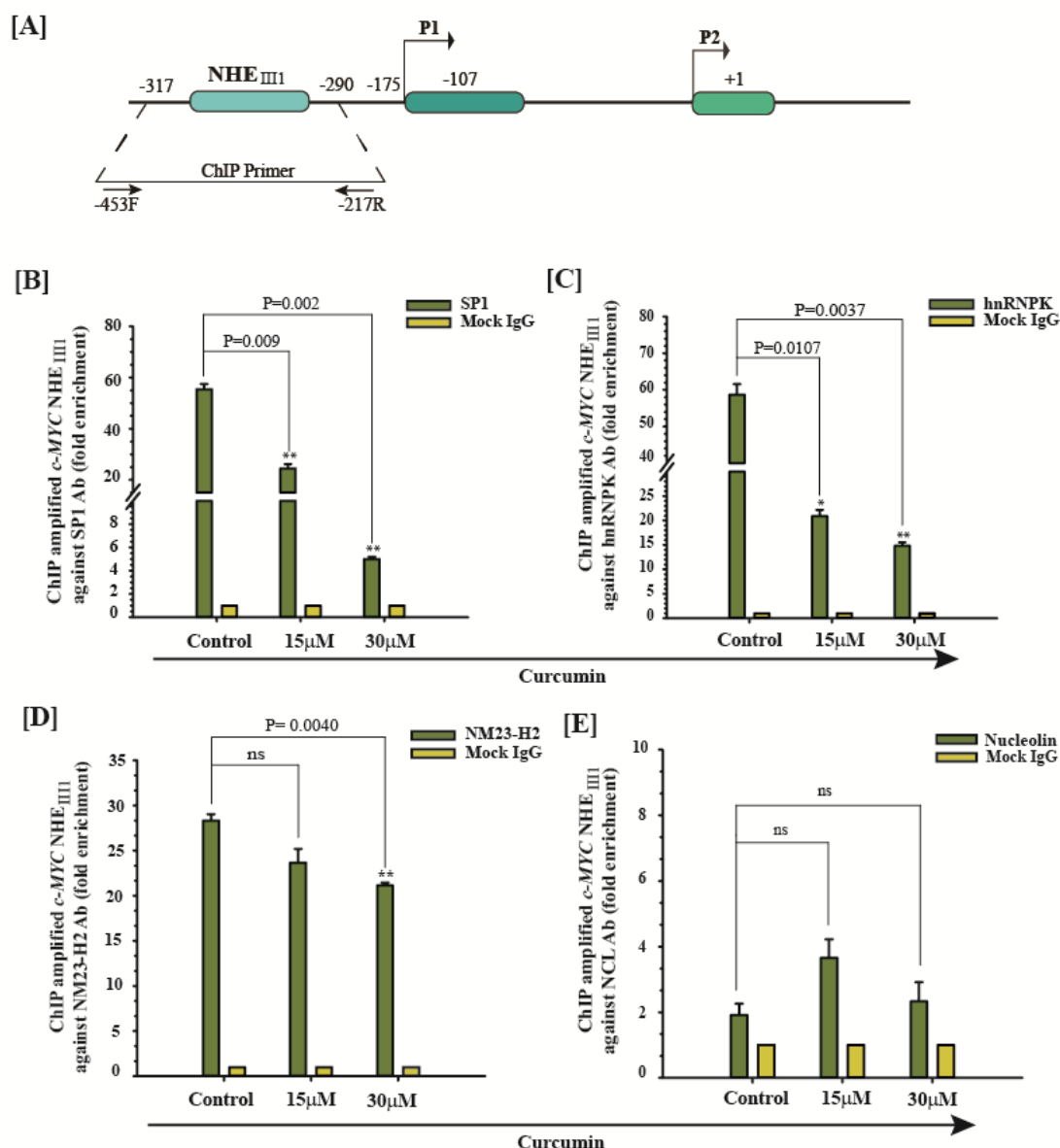


Figure 8. Differential occupancy level of transcription factors (SP1, hnRNP K, NM-23-H2 and Nucleolin (NCL)) at c-MYC NHE_{III} element. **[A]** Schematic representative diagram of the c-MYC promoter region. Nuclease hypersensitive element (NHE_{III}) containing the Pu-27 G-quadruplex motif is present upstream of P1 and P2 transcription initiation sites. The position of the ChIP primers used in real time PCRs following immunoprecipitation is marked. ChIP results show occupancy level of the transcription factors using specific antibodies against **[B]** SP1, **[C]** hnRNP K, **[D]** NM23-H2, **[E]** NCL transcription factors. The respective bar plots show the quantification of the binding of transcription factors at NHE_{III} by fold enrichment method compared to the IgG control based on quantitative real time PCR data. Error bars represent means \pm SE of triplicate set of data. Asterisks (*) indicate statistical significance as determined from paired t-test (* indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$, **** indicates $P < 0.0001$), denoting significant difference in occupancy level of the transcription factors (TFs) in the promoter region of c-MYC gene upon CUR treatment.

reduction in binding of Sp1 on 15 μ M of CUR treatment was noted, while its occupancy scaled-down by 11-fold upon 30 μ M of CUR treatment. We detected a 2.8-fold reduction in hnRNPK occupancy at the NHE_{III1} element on 15 μ M of CUR treatment, while the hnRNPK occupancy curtailed down by 4-fold at 30 μ M of CUR treatment. In our conducted ChIP assay, we did not notice significant enrichment of NM-23-H2 at the NHE_{III1} element. The recruitment of NCL at the NHE_{III1} region was trivial [Figure 8D and 8E]. The substantial reduction in the occupancy level of Sp1 at the NHE_{III1} region upon curcumin treatment relaxes the torsional stress created due to induced negative supercoiling. This weakens the recruitment and formation of hnRNPK-iM complex especially hnRNPK-5CT complex formation is hindered, while CUR stabilizes the GQ. The hindrance to hnRNPK-5CT complex formation upon CUR treatment is a plausible reason for low occupancy level of hnRNPK observed at the NHE_{III1} element. NCL TF was observed to be sparsely recruited at the NHE_{III1} element [Figure 8E]. We speculated that in CUR mediated GQ stabilization, the binding position of CUR on the GQ geometry may hinder NCL binding that results in the weak interaction of NCL and *c-MYC* GQ. The ChIP data of NCL occupancy is strengthened by MD simulation results later in this chapter. There are reports on high expression of NCL in proliferating cells³⁷. NCL acts as both a *c-MYC* GQ stabilizer and an inducer of proangiogenic and antiapoptotic factors³⁶. The alternative plausible explanation to our findings of insignificant recruitment of NCL is since NCL is highly over-expressed in actively dividing cells, whereas CUR treatment drastically inhibits cancer cell proliferation, curcumin may regulate *c-MYC* downregulation by stabilizing *Pu-27*-GQ without the assistance of NCL. Thus, curcumin downregulated *c-MYC* expression by stabilizing *Pu-27* GQ silencer switch in the NHE_{III1} promoter element without the stability dependency of the GQ on NCL transcription factor. Additionally, we evidenced undetectable alteration in the occupancy level of NM23-H2 at the NHE_{III1} element following CUR treatment [Figure 8D]. A plausible explanation to this result is that CUR binds to *Pu-27* GQ at a different site that do not

mask the topological site where NM-23-H2 interacts. There are reports that NM-23-H2 is unable to induce transcription on its own, since NM-23-H2 transcription factor lacks the characteristic activation domain for inducing transcription⁵⁵. Although, NM-23-H2 binding to the NHE_{III1} element remained unaltered even after CUR treatment, NM-23-H2 is unable to resolve the GQ structure stabilized by CUR and fails to induce *c-MYC* transcription.

Thus, low torsional stress at the upstream of NHE_{III1} element due to decreased binding of Sp1 synergizes with CUR mediated *Pu-27* GQ stabilization and hindered hnRNP-K-5CT complex formation result in downregulation of *c-MYC* transcription. CUR binding to *Pu-27* GQ drives the equilibrium in favor of stable GQ formation curtailing down *c-MYC* transcription to a threshold level. We have selected these TFs as a prototype to investigate *c-MYC* gene activity regulation by CUR. Our findings shall accelerate the development of novel ligands based on natural compounds like curcumin mimicking the mechanism of *c-MYC* gene downregulation by curcumin. Development of such drugs can specifically target and stabilize the *c-MYC* quadruplex in the NHE_{III1} element irrespective of NCL mediated stabilization and NM23-H2 induced destabilization.

3.2.10 *In silico* investigation of the interaction of NCL with CUR-MYC-GQ complex

We noticed an unstable interaction and low occupancy of NCL at the NHE_{III1} element, yet we obtained stable GQ formation assisted by CUR binding and subsequent downregulation of *c-MYC*. CUR mediated stabilization of *c-MYC*-GQ independent of the role of NCL is supported by our docking and molecular dynamic (MD) simulation results. We docked NCL onto CUR-*MYC*-GQ complex to identify the interacting areas of NCL with *c-MYC*-GQ. Our docking results revealed NCL to bind at the 3' end of *c-MYC* quadruplex. NCL binding is observed in the loop region of *MYC*. Adenine12 (A12) of *MYC*-GQ interacts *via* polar contact with Asp72

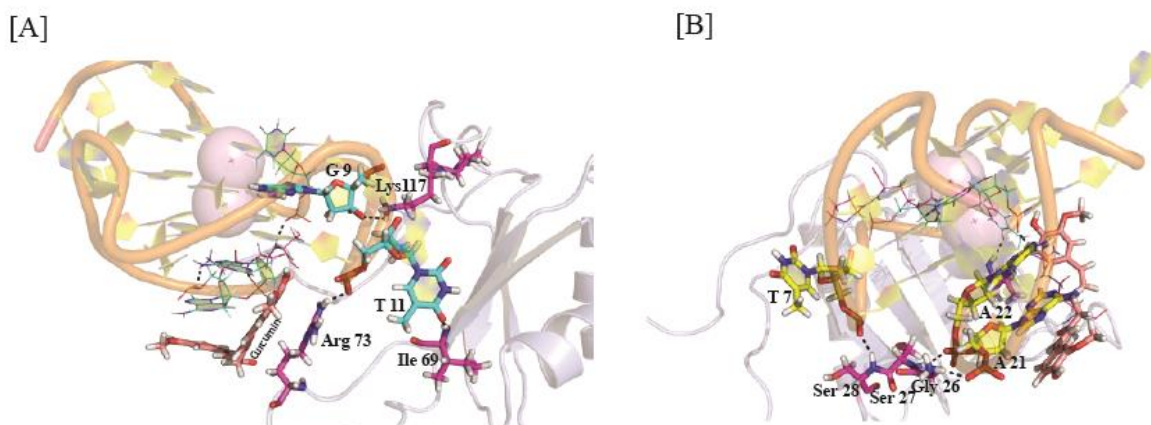


Figure 9. Molecular simulation. [A] and [B] showing the hydrogen bonding interaction between *c*-MYC-GQ and NCL in curcumin bonded state.

and Lys79. Thymine11 (T11) also interacts with Lys79. We observed CUR in complexed state with MYC-GQ to block the proper stable interaction of NCL with the GQ, since the ketone group of CUR also interacts with NCL by forming hydrogen bond with Lys79. CUR is also observed to interact with Gly47 and Asn107 with the methoxy group [Figure 9]. Upon simulation of CUR-MYC-GQ-NCL complex, we observed NCL to detach away from the complex at 34 nanoseconds. The validity of the unstable interaction of NCL in presence of curcumin at the NHE_{III1} element as obtained from our ChIP results is supported by our simulation data. We speculated curcumin to engross the central binding zone of MYC GQ that hinders NCL binding. We observed NCL to move from its binding position towards the loop of the GQ specifically with nucleotides: T7, T11, A21, and A22, and that reduces the stability of NCL interaction with CUR-MYC-GQ complex. Further analysis of the simulation data revealed CUR-MYC-GQ complex to retain stable interaction regardless of the presence or detachment of NCL. CUR stays bonded to the GQ at its position interacting *via* its central carboxyl group with A22, while one terminal of the ferulyl group of CUR bonds with A21 *via* π - π interaction and the other terminal bonds with T11 [Figure 9]. We have also deduced the RMSD of CUR-MYC interaction from both CUR-MYC and CUR-MYC-NCL complex till NCL

detaches from *MYC* GQ. RMSD revealed that CUR-*MYC* complex is highly stable ($<5 \text{ \AA}$) and is comparable both in presence and absence of NCL.

3.2.11 Activation of an alternative cell death pathway by CUR upon *c-MYC* downregulation

MYC alters the balance between anti-apoptotic and pro-apoptotic factors thereby, sensitizing the cells to apoptosize. There are reports on a positive correlation between upregulated *MYC* and activated p53-dependent and independent pathways mediated apoptosis⁵⁶. Thus, tumors with upregulated *MYC* are susceptible to self-destruction *via* activation of apoptosis². However, tumors cells with upregulated *MYC* evades apoptotic cell death by over-activating proto-oncogene *BCL2*⁵⁷. *MYC* assists in accelerating proliferation while *BCL2* maintains the cell survival signalling pathway in an active state. Constitutively activated *BCL2* synergizes with altered *MYC* to drive the tumor cells toward malignancy^{58,59}.

Depending on the cellular circumstances, altered *MYC* expression either accelerates cell-proliferation or triggers apoptosis. The bipartite function of *MYC* complicates the strategical cum conventional gene silencing method to inhibit cancer metastasis^{60,61}. CUR triggers the intrinsic pathway that leads to caspase activation and apoptosis. This process is achieved by increasing the expression of Apaf-1⁶². It is reported to activate Apaf-1-mediated caspase 9 that in turn activates mitochondria-dependent apoptotic cell death⁶². We investigated the downstream pathway of CUR mediated *MYC* downregulation that triggers apoptotic death in MDA-MB-231 cell line. We examined the expression-level of BCL-2 (anti-apoptotic marker) and *E2F -1* (upstream oncogene of *BCL-2*)⁶¹ in MDA-MB-231 upon CUR treatment. At 30 μM CUR treatment we detected, 1.8-fold(44.8%) transcriptional inhibition of *BCL-2* and 1.8-fold(44.2%) transcriptional inhibition of *E2F-1* gene [Figure 10A and 10C]. Delineating previous reported data of curcumin mediated apoptosis by upregulation of Apaf-1 and downregulation of VEGF in MDA-MB-231 cells^{62,63},

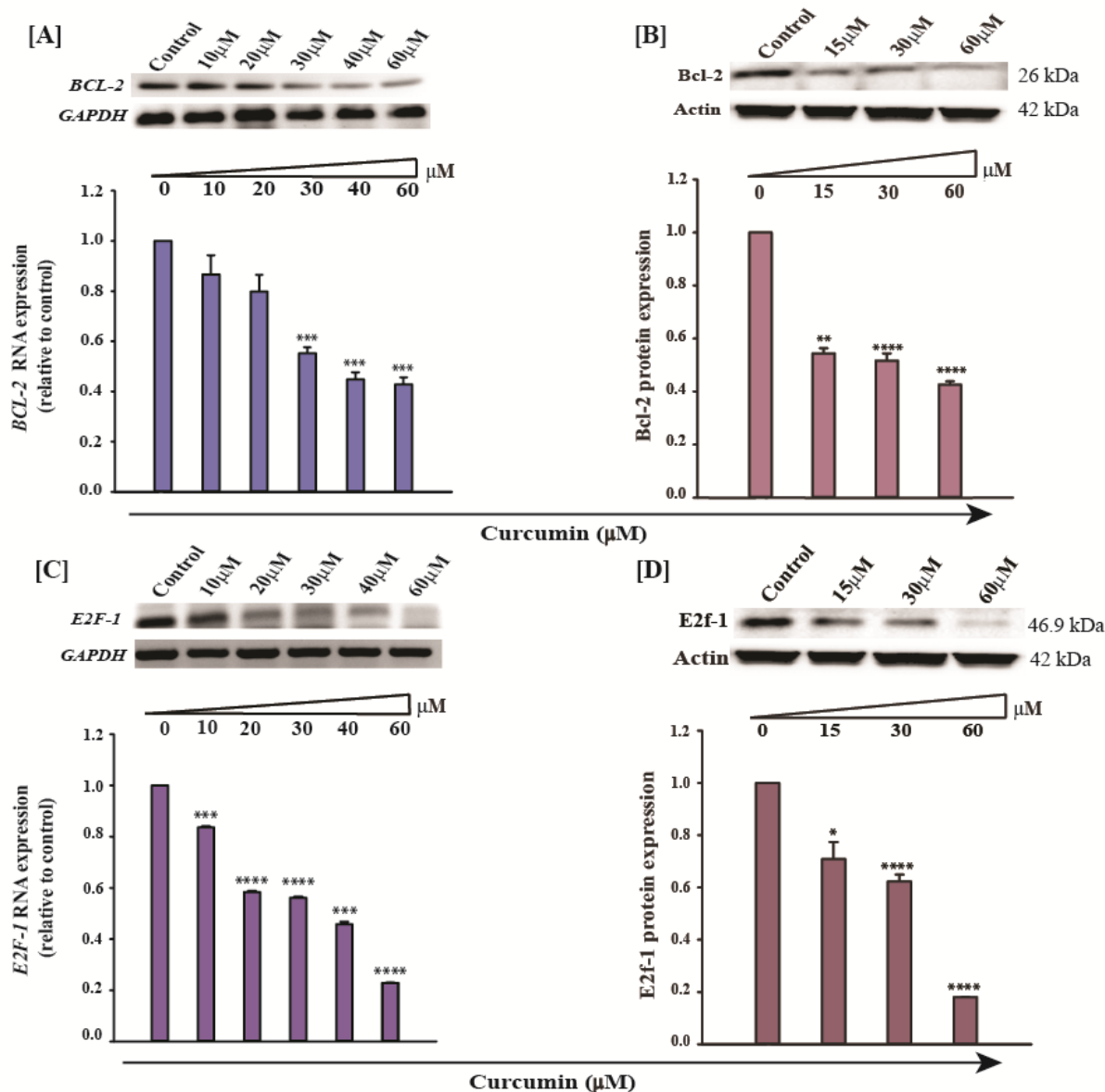


Figure 10. Analysis using *sq*-PCR demonstrates CUR-mediated repression of anti-apoptotic markers. The above panel represents the studied gene product, while the below panel represents the positive controls GAPDH and Actin. [A] BCL-2 transcription level and [C] E2F-1 transcription level in MDA-MB-231 cell line upon CUR treatment (0, 10, 20, 30, 40, 60 μM) along with the bar diagram plots of densitometric analysis of the gel bands normalized to control. [B] and [D] The levels of Bcl-2 and E2f-1 proteins were analyzed using Western blotting, and their corresponding bar graphs were generated after densitometric analysis. The values were normalized to the Control, and the histogram error bars represent the mean of normalized values \pm SE from a triplicate data set. Asterisks (*) indicate statistical significance as determined from the paired t-test, with * indicating $P < 0.05$, ** indicating $P < 0.01$, *** indicating $P < 0.001$, and **** indicating $P < 0.0001$. These statistical symbols denote a significant difference in RNA and protein levels and a significant difference in promoter activity upon CUR treatment.

Apoptosis detection assay values from Flow Cytometry (FACs)						
Events (%)	Control	DMSO (0.3%) Control	Curcumin 10μM	Curcumin 20μM	Curcumin 40μM	Curcumin 60μM
Live Cells	97.82	96.79	86.06	88.09	71.57	3.09
Early Apoptosis	1.92	1.52	13.37	9.92	22.30	63.98
Late Apoptosis	0.21	0.31	0.49	0.81	3.94	32.71
Necrosis	0.05	1.38	0.06	1.18	2.19	0.22

Table 2: Apoptosis detection assay in MDA-MB-231 cells upon CUR treatment.

we observed the downregulation of *BCL-2* via *E2F-1* pathway culminated in apoptosis. Previously, MTT result showed how MDA-MB-231 proliferative rate was dose-dependently suppressed by CUR. Our FACs data [Table 2] explained the translational repression of Bcl2 [1.9-fold (48%)] at 30 μM of CUR treatment thereby, validating the activation of apoptosis by CUR in metastatic cell line [Figure 10B].

We envisaged from our analyzed data that curcumin stabilizes the *MYC*-GQ regulatory switch and downregulates its activity. Activation of a specific downstream signalling cascade upon lowering *MYC* levels leads to significant repression of *BCL-2* and induction of apoptosis. From our data, we observed 1.6-fold (38%) translational downregulation of E2f-1 at 30μM of CUR treatment that correlates with the downregulation of c-Myc translated product [Figure 10D]. Analysis of our data and the reports that are currently available revealed that curcumin inhibits *c-MYC* transcription by stabilizing *Pu-27* GQ silencer element upstream of the *c-MYC* promoter, that further disrupts the *VEGF-E2F1-BCL2* axis, thereby inducing apoptotic cell death in metastatic cancer cells [Figure 11].

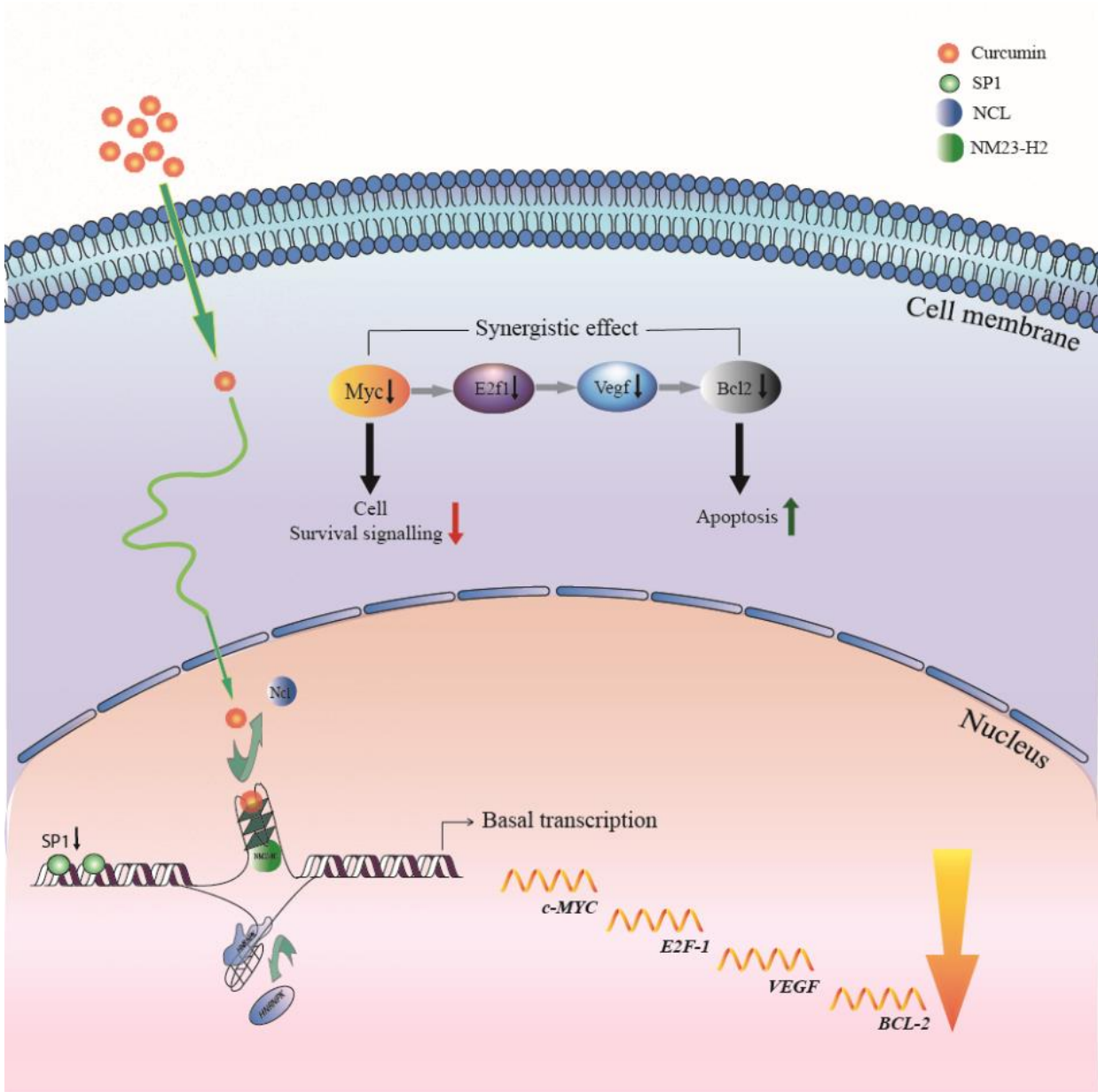


Figure 11. Schematic representation illustrates how c-MYC GQ scaffold at NHEIII1 element predominates in the presence of curcumin (CUR) without the assistance of NCL TF. As a result of CUR binding, hnRNPK and SP1 occupancy at NHEIII1 decreases. Due to the insufficient cooperative action of hnRNPK and SP1 upon CUR treatment, the transcription of MYC is kept low. CUR-mediated downregulation of c-MYC downregulates anti-apoptotic BCL-2 via downregulation of E2F-1 TF, present upstream of BCL-2, thereby triggering apoptosis in metastatic breast cancer cells.

3.3 Conclusion

In brief, our study focused on examining how curcumin interacts with c-MYC G-quadruplex in buffer environment (*in solution*). Various biophysical and *in silico* experiments collectively determined the interaction of c-MYC GQ with CUR. Our findings through CD spectroscopic analysis showed that CUR enhanced the stability of the GQ and also maintained the parallel-looped tetrad scaffold of the c-MYC-GQ. ITC experimental study revealed efficient binding potential of CUR with Pu-27 GQ. The ITC data also suggested that Pu-27-CUR complex formation is thermodynamically favorable. Results from molecular docking and MD simulations show that CUR forms a stable complex with the c-MYC GQ at its 3' end. Extracting medicinal compounds from natural products is a significant process. These compounds offer a primary source for the development of therapeutic agents. The selectivity and binding affinity of CUR for the c-MYC GQ make it a potential model for developing GQ-specific ligands. We investigated curcumin's interaction with c-MYC-GQ *in vitro* in MDA-MB-231 metastatic breast cancer cell. We have shown how curcumin orchestrates the transcription factors at the NHE_{III1} element and control c-MYC activity. The drastic depletion of SP1 at NHE_{III1} element is identified as the crucial determinant in the downregulation of MYC upon CUR treatment. Our further investigation on the downstream signaling pathway of CUR mediated Pu-27 GQ conformational arrest revealed the deregulation of an alternate pathway [MYC-VEGF-E2F1-BCL2] through which CUR triggers apoptotic death in cancer.

Rational drug design and GQ ligand optimization require a clear comprehension of the interaction between small-molecule ligands and their G4 targets. Based on our findings, we may be able to develop concepts for designing compounds that can drive c-MYC expression to its baseline, without posing a risk to healthy proliferating cells. The development of a diagnostic platform for future therapeutic intervention to treat metastatic cancer can avail the use of the spatiotemporal formation and stabilization of G-quadruplex.

3.4 Experimental section

Oligonucleotide sequences used in our study:

Gene (GQ)	Sequences
<i>Pu-27-cMYC</i>	5'-TGGGGAGGGTGGGGAGGGTGGGGAAGG-3'
<i>BCL2</i>	5'-CGGGCGCGGGAGGAATTGGGCGGGAGC-3'
<i>VEGF</i>	5'-GGGGGCGGGCCGGGGGCGGGTCCCGGCGGGGCGGAG-3'
<i>KRAS</i>	5'-GGGAGGGAGGGAAGGAGGGAGGGAGGGA-3'

Table 3: G-Quadruplex Sequences used in our study

Sample preparation:

We procured the oligonucleotides from Eurofins Genomics India Pvt. Ltd. through commercial synthesis. In this study we selected *Pu-27* (wild-type) 5'- TGGGGAGGGTGGGGAGGGTGGGGAAGG -3' sequence. Potassium phosphate dibasic(KH_2PO_4) and potassium phosphate monobasic(K_2HPO_4), potassium chloride (KCl) and curcumin (CUR) were purchased from Sigma Aldrich. Oligonucleotides were dissolved (1mM concentration) in a buffer solution of 10 mM potassium phosphate buffer (Neutral pH), supplemented with 100 mM KCl. For proper annealing, the oligonucleotide (oligo) samples annealed at 95 °C for 5 minutes in a heat-block, then slowly cooled to RT overnight. CUR stock solution of 20mM was prepared in DMSO solvent of 99.9% purity and filtered for further use. The working concentration range of CUR was 5 μM to 60 μM , with DMSO concentration ranging from 0.05% to 1%. A 5mg/ml stock solution of Doxorubicin (DOX) was prepared in ultrapure water.

Fluorescence spectroscopy

Fluorescence titration was carried out in Hitachi spectrophotometer(F-7000 FL spectrophotometer). A 1 cm of pathlength cuvette was used to record the fluorescence spectra. The Ex was fixed at 427nm and the Em was scanned from 475 to 600nm. Modified Scatchard model for multiple binding sites was employed to analyze the fluorescence titration data⁶⁴.

Each fluorescence emission scan was taken after 5 minutes of incubation with each increment of titrant. In the reverse titration, we kept 10 μM of CUR constant and titrated with an increment of oligo. $1/\text{fluorescence}$ was plotted against $1/[\text{quadruplex}]$ to obtain the standard curve. The fluorescence intensity of 10 μM quadruplex-CUR complex was measured. 10 μM oligonucleotide was kept constant in the forward titration which was gradually titrated with increasing concentration (5-100 μM) of CUR.

Concentration of bound drug was determined; the following equation was used:

$$\alpha = \frac{(F_f - F)}{(F_f - F_s)} \quad (1)$$

where, F_f and F_s , respectively are fluorescence for free and total bound CUR.

The free concentration of curcumin (C_f) was calculated using $C_f = (1 - \alpha)C$, and was subsequently used to determine the binding ratio.

$$r = \frac{(nK C_f)}{(1 + K C_f)} \quad (2)$$

r being the mole of bound ligand per mole of GQ, K : the association constant and n : ligand binding site on the GQ. r was plotted against c_f (free curcumin) to find out the binding constant. The plot of r verses C_f for BCL2, KRAS and VEGF G quadruplex structures were performed for a comparative study.

The fluorescence measurements were adjusted to account for the inner filter effect by applying the Lakowicz equation.[2].

$$F_{corrected} = F_{observed} [\text{anti log} \{ \frac{A_{ex} + A_{em}}{2} \}] \quad (3)$$

Ethidium Bromide (EtBr) displacement assay

EtBr displacement experiment was employed to ascertain the mechanism of curcumin's binding with G-quadruplex. At first, 15 μM EtBr was saturated with 30 μM of GQ. EtBr-GQ complex was titrated with increments of CUR with 3 minutes equilibrating time between each titration.

The E_x of EtBr was fixed at 510nm and E_m profile was monitored between 520 to 640 nm. Temperature was maintained at 25°C during the experiment. The % displacement is calculated from the fluorescence area (FA, 520–640 nm, λ_{ex} =510 nm) using: % displacement = $[100 - (FA/FA_0) * 100]$, FA_0 is the fluorescence of EtBr bound to DNA before titration with CUR. We obtained the FID curves by plotting EtBr displacement percentage against oligo concentration.

Circular Dichroism (CD) spectroscopic analysis

Jasco J815 spectrophotometer was used to perform circular dichroism (CD) spectroscopic analysis. The CD spectra were monitored from 220 to 320 nm with 100 nm/min scanning speed and 1 nm step size. The instrument was connected to a peltier system for maintaining the required temperature. Oligonucleotide at a concentration of 20 μ M were dissolved in 10 mM of potassium phosphate buffer (pH 7.0) supplemented with 100 mM KCl. Under this buffer condition the oligo was titrated with incremental addition (20 μ M, 40 μ M and 60 μ M) of CUR. The CD spectra were monitored for complete complex formation and equilibration after 5 minutes of each titration. After baseline correction, each spectra were calculated as the average of 3 observations taken at 25°C. We also performed control experiments in 0.3% DMSO-containing buffer without the ligand molecule. For the CD melting assays, we have used receptor to ligand ratio of 1:3 (concentration of quadruplex: CUR= 1:3). CD spectral scan of the melting experiments were taken at a 5°C temperature increment from 20 to 90°C. CD melting assays were repeated thrice. We presumed the quadruplex structure to unfold with gradual increase in temperature (denaturation) and it re-conforms back to its initial structure with gradual decrease in temperature (renaturation). We have analyzed the melting curves by fitting the data presuming a ‘two-state transition’ model.

Therefore, at any temperature (t), the fraction of folding is;

$$\alpha = ([F])/([F] + [U]) = ((\theta_t - \theta_U))/((\theta_F - \theta_U))$$

[F] is the concentration of folded quadruplex and [U] is the concentration of unfolded quadruplex. α denotes the fraction folded at temperature t , θ_t is the ellipticity of quadruplex in folded and unfolded state respectively and θ_t is the ellipticity of quadruplex at any (t)⁶⁵. We plotted the fraction of folding of the GQ with respect to temperature and fitted the curve applying sigmoidal 3-parameter binding in SigmaPlot software.

Isothermal titration calorimetry (ITC)

The thermodynamic properties of CUR-GQ interaction were studied using iTC200 Microcalorimeter with the temperature maintained at 25°C. The ITC cell was loaded with 400 μ M of GQ and the ITC syringe was loaded with 40 μ M of CUR. GQ was titrated with CUR. 40 μ M of CUR working concentration was mixed in 10mM KP buffer supplemented with 100mM KCl with DMSO 0.2%. The heat of dilution during the titration was reduced to insignificant level by maintaining identical buffer condition both in the cell and the syringe. Samples were degassed prior to ITC titration. 25 injections, each of 2 μ l oligonucleotide was injected into the ITC cell containing CUR. Blank titrations were also performed for baseline correction where oligonucleotide was injected in increments to the buffer containing identical DMSO (0.2%) percent. The heat of dilution in the CUR-GQ association reaction was normalized with that of the blank set. The corrected values were graphed in relation to the molar ratio. Independent fit model was used to fit the data and further analyzed in NanoAnalyzer in-built Origin 7.0 software. The best-fit values of the change in entropy (ΔS), enthalpy (ΔH), the binding stoichiometry(n), and K_d of CUR-GQ interactions were obtained. Gibb's free energy (ΔG) was calculated using the formula: $\Delta G = \Delta H - T\Delta S$

Mammalian cancer cell culture

Triple negative breast cancer cell line MDA-MB-231 was purchased from ATCC. Human cells were cultured in RPMI 1640 complete liquid media supplemented with FBS [10% (v/v)],

penicillin-streptomycin (1%), gentamycin (50 µg/ml), amphotericin B (2.5µg/ml). The cells were maintained in ESCO cell culture CO2 incubator under 95% humidity, 5% CO2 and 37°C.

Cell viability assay

The cell viability assay detects the reduction of MTT (3-(4,5-dimethylthiazolyl)-2,5-diphenyl-tetrazolium bromide) to formazan by mitochondrial dehydrogenase. This assay detects the percentage of cell viability upon a drug treatment and cytotoxicity level. MDA-MB-231 cells (3×10^3 cells/well) were grown in 96-well plate up to ~60% confluency followed by treatment with CUR (0 to 60 µM), doxorubicin (positive Control), and DMSO (negative control). Next, the treated cells were incubated for next 24 hours. After incubation, the treated cells were washed with 1X PBS. MTT (0.5mg/mL in PBS) was pipetted into each well followed by incubation for 3 hours at 37°C. After incubation 100µL DMSO was pipetted into each cell sample and then the plate was kept in rocking condition for 1 hour in dark place to dissolve the crystals. The ELISA reader (MultiskanEX) measured absorbance at 570nm with a reference wavelength set at 650nm. Drug toxicity was measured using the formula:

$$\text{Cytotoxicity (\%)} = 1 - \frac{(\text{mean absorbance of sample} - \text{blank})}{(\text{mean absorbance of negative control} - \text{blank})} * 100$$

$$\text{Viability \%} = 100 - \text{Cytotoxicity \%}$$

Plasmid amplification and transfection

For performing the luciferase assay, we have used the promoter-less luciferase vectors pGL4.72 [*hRlucCP*] with Kpn I and Hind III restriction sites. Luciferase reporter plasmid vectors were constructed by inserting the promoter sequence of *c-MYC* or without the *Pu-27-GQ* sequence utilizing the restriction sites for the insertion upstream of the luciferase reporter gene (*hRlucCP*) (*Renilla reniformis*). The wildtype promoter (GQ+) constructs and the deletion (GQ-) constructs were made to uptake by the competent *E. coli* DH5α cells for amplification via the transfection protocol. The transformed cells were selected by culturing the cells in Luria

Agar (LA) agar media with added ampicillin (100µg/ml) and kept at 37°C till prominent bacterial colonies developed. Single colonies were further cultured in Luria Broth (LB) and proceeded for plasmid isolation. We used the QIAGEN Plasmid Midi Kit (Catalog no. 12143) for plasmid isolation following protocol of the kit. Purified plasmid concentration was measured in Nanodrop instrument and further used in the luciferase assay.

Luciferase assay

A 24-well microtiter plate for adherent cell culture was seed with MDA-MB-231 cells at a cell density of 2×10^4 cells/well. After 24hrs, cells were transfected with 500ng of constructed GQ+ or GQ- clones (description provided under section: Plasmid amplification and transfection). pGL3 vector was as used as a control. The transfecting agent Lipofectamine 2000 (Thermo-Fisher Scientific Catalog no. 11668027) was used for efficient transfection as per the company's guidelines followed by incubation in a humidified environment at 37 °C with 5% CO₂. Transfection was carried out for 4 hours, following which the cells received CUR treatment at concentrations: 10, 15, 20, 30, 40 and 60µM for 24 hours. Following treatment incubation, the cells were washed with 1X PBS and lysed in passive lysis buffer included in the kit (1X PLB). LAR II was added and luminescence was measured and subsequently Stop and Glow solution was added and luminescence was measured. All was done according to the Promega Dual Luciferase Kit protocol. We used the GloMax® 20/20 Single-Tube Luminometer (Promega) for the measurement of luminescence.

Semiquantitative RT-PCR

Semiquantitative Reverse transcriptase-PCR was done for examining the change in RNA expression of *c-MYC*, *E2F1*, *BCL2* and *GAPDH* upon CUR treatment. Final PCR reactions were performed using Novagen KOD Hot Start PCR reaction mix following the company's guidelines in Prima 96plus TM thermal cycler.

Thermocycling setting:

Initial denaturation at 95°C (10mins)

40 cycles

Denaturation at 95°C (30 seconds)

Annealing temperature (30 seconds)

Extension at 72°C (45 seconds)

The amplified PCR products were separated in a 2% agarose EtBr gel to confirm presence of a single expected band. The data was normalized with the control to calculate relative gene expression. List of primers are provided in Table 4.

Primers	Oligo Sequences (5'-3')	Annealing Temperature (°C)
c-MYC RT Forward Primer	AATGAAAAGGCCCAAGGTAGTTATCC	55
c-MYC RT Reverse Primer	GTCGTTTCCGCAACAAGTCCTCTTC	
BCL2 RT Forward Primer	CTGCACCTGACGCCCTTCACC	61
BCL2 RT Reverse Primer	CACATGACCCCAACGAAGTCAAAG	
E2F1 RT Forward Primer	GTGTAGGACGGTGAGAGCAC	54.5
E2F1 RT Reverse Primer	TCAAGGGTAGAGGGAGTTGG	

Table 4: Primer pairs used in semi-quantitative PCR for RNA expression check.

Western blot (WB) assay

MDA-MB-231 cells were cultured in 60mm culture dishes. Upon attaining 70% confluency, the cells were treated with CUR (15, 30, and 60µM). Upon 24 hours of CUR treatment, the cells were lysed using 10X Cell Lysis Buffer (Cell Signaling Technologies) with added PMSF (Merck) and PIC cocktail (Merck). Cell lysis was performed according to the company's instructions. The concentration of isolated protein was determined using Bradford reagent as per Bradford's method of protein estimation^[37-39]. Sodium dodecyl sulfate polyacrylamide gel (4% stacking gel and 12% resolving gel) was used to resolve 40 µg of protein. Protein samples

were transferred by blotting onto to PVDF membrane. Membrane blocking was done with 5% skimmed milk in Tris buffered saline (TBS) followed by 1hr incubation at RT. The membranes were soaked in specific antibody dilutions overnight at a temperature of 4°C. Antibodies used: anti-c-Myc antibody was used at a dilution of 1:1000, similarly anti-Bcl2 at 1:500 dilution, anti-E2f1 at 1:1000 dilution, and anti-Actin at a dilution of 1:5000 was used. Following incubation with specific antibodies, the membranes were incubated with HRP-conjugated anti-rabbit IgG at a dilution of 1:8000 and HRP conjugated anti-mouse IgG at a dilution of 1:10,000 and incubated at RT for 2 hours. Using the Clarity™ Western ECL substrate (Bio-rad), target proteins were visualized on the membrane, and images captured in ChemiDoc MP system and Image Lab™ software for PC version 6.0 (Bio-rad).

Details of the antibodies used in the WB assay:

Anti-c-Myc (9402S, Rabbit polyclonal) was purchased from CST. Anti-Bcl-2 (sc-509, mouse monoclonal), and Anti-β-Actin (A5441, mouse monoclonal) were obtained from Santacruz Biotechnology Inc. Anti-E2f1 (8G9, mouse monoclonal) was purchased from Novus Biologicals.

Chromatin Immunoprecipitation Assay (ChIP)

The physical interaction of selected transcription factors (SP1, NM23-H2, hnRNPK, and NCL) at the NHE_{III} element of *c-MYC* promoter region was studied by performing chromatin immunoprecipitation assay. We used Pierce Magnetic ChIP Kit (Cat no. 26157) to conduct the experiment. 100 mm adherent cell culture plates were used to culture MDA-MB-231 at a cell density of 1×10^6 per plate. Following 24hrs, the cells were treated with CUR and incubated for another 24 hours. Cross-linking to identify protein-DNA interaction was done using 1% Formaldehyde and incubated for 10 minutes under shaking condition at RT. Further crosslinking was stopped by the adding 1X Glycine. The cross-linked products were washed in ice-cold PBS and dissolved in membrane extraction buffer (supplemented with inhibitors)

Primers	Oligonucleotide sequences (5'-3')	PCR annealing temperature (°C)	PCR amplicon length
c-MYC ChIP Forward primer	GCAGCAGAGAAAGGGAGAGG	56°C	238
c-MYC ChIP Reverse primer	GCGGAGATTAGCGAGAGAGG		

Table 5: Primer pairs used in Real-Time PCR for ChIP assay

present in the ChIP kit and placed on ice for 10 minutes. Then the samples were centrifuged at 9000xg for 3 minutes following which the nuclei content of each sample was collected and resuspended in MNase digestion buffer and the chromatin was digested with MNase following the kit's instructions. The partially digested product was resuspended in 1X IP dilution buffer with added inhibitors and proceeded to sonication for rupturing the nuclear membrane. The digested chromatin present in the supernatant was used for the immunoprecipitation according to the manufacturer's protocol. Prior to this 10% of each of the samples was stored at -20°C as Input. The main samples were treated with specific antibody dilutions prepared in 1X IP dilution buffer followed by incubation at 4°C for 6 h under rotating condition. The following antibodies were used (ChIP grade): Anti-hnRNPK (ab39975), anti-SP1 (ab13370), anti-NCL (ab13541), and anti-NM23-H2 (sc-14790).

The pull-down of protein-DNA complex was done with magnetic beads following the kit's protocol. Each pulled down product were washed in wash buffers provided in the kit. The bio-complex was eluted out from the bead's attachment by incubating the samples at 65°C for 40mins with recurrent resuspension of the products. The precipitated gDNA fragments and the Inputs were isolated from their protein attachment *via* treatment with 6 µl 5M NaCl and 2µl of 20mg/ml proteinase K at 65°C for 1.5 hours. Purified DNA was collected in 50µl of DNA column elution buffer present in the kit. We performed Realtime RT-PCR to detect the results of the chromatin immunoprecipitation assay. Maxima SYBR Green/ROX qPCR Master Mix

(2×) [Thermo-Scientific] was used to conduct RT-qPCR. We used the reported PCR primers for *c-MYC* NHE_{III} element [Table 5].

The thermo cycling conditions settings:

Initial denaturation at 95°C for 10 minutes

40 cycles

Initial denaturation at 95°C for 30 seconds,

Annealing at appropriate melting temperature based on the primer properties for 30 seconds

Extension at 72°C for 30 seconds.

Difference in occupancy was calculated by Fold enrichment method: $\Delta\Delta C_t = C_t (\text{Target}) - C_t (\text{IgG})$. Fold enrichment: $2^{(-\Delta\Delta C_t)}$.

Molecular Docking and Molecular Dynamic Simulations

From the RSCB PDB website, the coordinate file for the solution NMR structure of *c-MYC* (PDB ID: 1XAV), was obtained and further analysed in PyMOL. The Schrodinger software package was used to generate the curcumin structure. The curcumin molecule was made ready using Ligprep and further docked with MYC GQ using Schrodinger software suite (Glide module)^[40]. The protein preparation wizard module of the software was used to process the docked data. The selection and refinement of the best pose was done using AMBER^[41]. HDock is a web server that facilitates the docking of proteins with proteins, DNA, and RNA. It uses a hybrid algorithm of template-based modelling and ab initio free docking to predict the interaction between curcumin and MYC. The resulting CUR-MYC complex is refined through a process of minimization using equal steps of steepest descent and conjugate gradient method, and is then slowly heated to 300K. The equilibrated system was used for simulation. 40 nanosecond (ns) of simulation run was conducted on an NPT ensemble at a pressure of 1 atm, temperature of 300K, and step size of 2 fs. The MD trajectories were analyzed in CPPTRAJ

module of AMBER ^[43]. The final trajectories were visualized in VMD and PyMol software. The co-ordinate file of *c-MYC* (PDB ID: 1XAV) complexed with curcumin prepared in our study was utilized for further studies. Nucleolin protein was docked to *MYC-CUR* complex. We obtained the PDB file of NCL from the RCSB-PDB database and then conducted docking using HDOCK. ^[42].

During Molecular Simulation every atom is denoted by a certain position in the Cartesian plane, and at every step the change of its trajectory is measured. In order to obtain insight of the valid change in the structural conformation under investigation, we can vary the root mean square deviation (RMSD) of the distance between two atoms with variable paths at each timepoint of the simulation. Structural conformation changes are subject to fluctuations, which are commonly measured as RMS. RMS is an important factor in detecting changes in the final conformation compared to the starting conformation of the ligand when complexed with the target. Lower values of RMS indicate early stabilization of ligand-target complex, while higher values indicate longer stabilization time.

Software or Analytical tools

The data were analyzed and graphical representations prepared by employing Microsoft Excel Office 2019 version 2103, SigmaPlot version 12.0, and Origin v.7.0 software's. For the statistical analysis, we employed the use of GraphPad Prism version 8.4.2.

3.5 References

1. Kinzler, K. W. *et al.* Amplification units containing human N-myc and c-MYC genes. *Proc Natl Acad Sci U S A* **83**, 1031–1035 (1986).
2. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat Rev Cancer* **8**, 976–990 (2008).
3. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat Rev Cancer* **2**, 764–776 (2002).
4. Beaulieu, M.-E. & Soucek, L. Finding MYCure. *Mol Cell Oncol* **6**, e1618178 (2019).

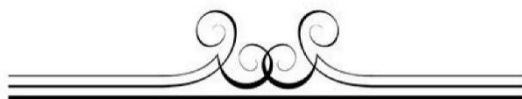
5. Niu, Z. *et al.* Knockdown of c-MYC inhibits cell proliferation by negatively regulating the Cdk/Rb/E2F pathway in nasopharyngeal carcinoma cells. *Acta Biochim Biophys Sin (Shanghai)* **47**, 183–191 (2015).
6. Weinstein, I. B. & Joe, A. Oncogene addiction. *Cancer Res* **68**, 3077–3080; discussion 3080 (2008).
7. Wierstra, I. & Alves, J. The c-MYC promoter: still Mystery and challenge. *Adv Cancer Res* **99**, 113–333 (2008).
8. Balaratnam, S. & Schneekloth, J. S. Chapter Eleven - Transcriptional regulation of MYC through G-quadruplex structures. in *Annual Reports in Medicinal Chemistry* (ed. Neidle, S.) vol. 54 361–407 (Academic Press, 2020).
9. González, V., Guo, K., Hurley, L. & Sun, D. Identification and Characterization of Nucleolin as a c-MYC G-quadruplex-binding Protein. *J Biol Chem* **284**, 23622–23635 (2009).
10. Mathad, R. I., Hatzakis, E., Dai, J. & Yang, D. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res* **39**, 9023–9033 (2011).
11. Brooks, T. A. & Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **1**, 641–649 (2010).
12. Ghimire, C. *et al.* Direct Quantification of Loop Interaction and π - π Stacking for G-Quadruplex Stability at the Submolecular Level. *J. Am. Chem. Soc.* **136**, 15537–15544 (2014).
13. Chelerythrine down regulates expression of VEGFA, BCL2 and KRAS by arresting G-Quadruplex structures at their promoter regions | Scientific Reports.
<https://www.nature.com/articles/srep40706>.
14. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol* **19**, 414–422 (2009).
15. Kouzine, F. *et al.* In Vivo Chemical Probing for G-Quadruplex Formation. *Methods Mol Biol* **2035**, 369–382 (2019).
16. Summers, P. A. *et al.* Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat Commun* **12**, 162 (2021).
17. Hurley, L. H., Von Hoff, D. D., Siddiqui-Jain, A. & Yang, D. Drug Targeting of the c-MYC Promoter to Repress Gene Expression via a G-Quadruplex Silencer Element. *Seminars in Oncology* **33**, 498–512 (2006).
18. Han, F. X., Wheelhouse, R. T. & Hurley, L. H. Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *J. Am. Chem. Soc.* **121**, 3561–3570 (1999).

19. Kim, M.-Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *J Am Chem Soc* **124**, 2098–2099 (2002).
20. Jana, J. *et al.* Chelerythrine down regulates expression of VEGFA, BCL2 and KRAS by arresting G-Quadruplex structures at their promoter regions. *Sci Rep* **7**, 40706 (2017).
21. Dickerhoff, J., Brundridge, N., McLuckey, S. A. & Yang, D. Berberine Molecular Recognition of the Parallel MYC G-Quadruplex in Solution. *J. Med. Chem.* **64**, 16205–16212 (2021).
22. Bhadra, K. & Kumar, G. S. Interaction of berberine, palmatine, coralyne, and sanguinarine to quadruplex DNA: a comparative spectroscopic and calorimetric study. *Biochim Biophys Acta* **1810**, 485–496 (2011).
23. Liu, L.-Y., Ma, T.-Z., Zeng, Y.-L., Liu, W. & Mao, Z.-W. Structural Basis of Pyridostatin and Its Derivatives Specifically Binding to G-Quadruplexes. *J. Am. Chem. Soc.* **144**, 11878–11887 (2022).
24. Machireddy, B., Sullivan, H.-J. & Wu, C. Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *Molecules* **24**, 1010 (2019).
25. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew Chem Int Ed Engl* **53**, 999–1002 (2014).
26. Zhang, S., Wu, Y. & Zhang, W. G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem* **9**, 899–911 (2014).
27. Maleki, P. *et al.* Quantifying the impact of small molecule ligands on G-quadruplex stability against Bloom helicase. *Nucleic Acids Research* **47**, 10744–10753 (2019).
28. Lo Cascio, F., Marzullo, P., Kayed, R. & Palumbo Piccionello, A. Curcumin as Scaffold for Drug Discovery against Neurodegenerative Diseases. *Biomedicines* **9**, 173 (2021).
29. Priyadarsini, K. I. The chemistry of curcumin: from extraction to therapeutic agent. *Molecules* **19**, 20091–20112 (2014).
30. Amalraj, A., Pius, A., Gopi, S. & Gopi, S. Biological activities of curcuminoids, other biomolecules from turmeric and their derivatives – A review. *J Tradit Complement Med* **7**, 205–233 (2016).
31. Jha, N. S. *et al.* Targeting human telomeric G-quadruplex DNA with curcumin and its synthesized analogues under molecular crowding conditions. *RSC Adv.* **6**, 7474–7487 (2016).
32. Sengupta, P., Bhattacharya, A., Sa, G., Das, T. & Chatterjee, S. Truncated G-Quadruplex Isomers Cross-Talk with the Transcription Factors To Maintain Homeostatic Equilibria in c-MYC Transcription. *Biochemistry* **58**, 1975–1991 (2019).

33. Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* **138**, 14138–14151 (2016).
34. Sengupta, P., Bose, D. & Chatterjee, S. The Molecular Tête-à-Tête between G-Quadruplexes and the i-motif in the Human Genome. *Chembiochem* **22**, 1517–1537 (2021).
35. Abou Assi, H., Garavís, M., González, C. & Damha, M. J. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res* **46**, 8038–8056 (2018).
36. Liang, P. *et al.* The expression and proangiogenic effect of nucleolin during the recovery of heat-denatured HUVECs. *Biochim Biophys Acta* **1830**, 4500–4512 (2013).
37. Chen, Z. & Xu, X. Roles of nucleolin. Focus on cancer and anti-cancer therapy. *Saudi Med J* **37**, 1312–1318 (2016).
38. Banerjee, S., Jha, H. C. & Robertson, E. S. Regulation of the metastasis suppressor Nm23-H1 by tumor viruses. *Naunyn Schmiedebergs Arch Pharmacol* **388**, 207–224 (2015).
39. Ray, S., Chattopadhyay, N., Mitra, A., Siddiqi, M. & Chatterjee, A. Curcumin exhibits antimetastatic properties by modulating integrin receptors, collagenase activity, and expression of Nm23 and E-cadherin. *J Environ Pathol Toxicol Oncol* **22**, 49–58 (2003).
40. Mehta, K., Pantazis, P., McQueen, T. & Aggarwal, B. B. Antiproliferative effect of curcumin (diferuloylmethane) against human breast tumor cell lines. *Anticancer Drugs* **8**, 470–481 (1997).
41. Rivera, M. *et al.* Targeting multiple pro-apoptotic signaling pathways with curcumin in prostate cancer cells. *PLOS ONE* **12**, e0179587 (2017).
42. Yoysungnoen, P., Wirachwong, P., Changtam, C., Suksamrarn, A. & Patumraj, S. Anti-cancer and anti-angiogenic effects of curcumin and tetrahydrocurcumin on implanted hepatocellular carcinoma in nude mice. *World J Gastroenterol* **14**, 2003–2009 (2008).
43. Wilken, R., Veena, M. S., Wang, M. B. & Srivatsan, E. S. Curcumin: A review of anti-cancer properties and therapeutic activity in head and neck squamous cell carcinoma. *Mol Cancer* **10**, 12 (2011).
44. Waseem, M., Pandey, P., Tomar, B., Raisuddin, S. & Parvez, S. Ameliorative action of curcumin in cisplatin-mediated hepatotoxicity: an in vivo study in Wistar rats. *Arch Med Res* **45**, 462–468 (2014).
45. Jacob, A., Wu, R., Zhou, M. & Wang, P. Mechanism of the Anti-inflammatory Effect of Curcumin: PPAR-gamma Activation. *PPAR Res* **2007**, 89369 (2007).
46. Quenching of Fluorescence. in *Principles of Fluorescence Spectroscopy* (ed. Lakowicz, J. R.) 277–330 (Springer US, 2006). doi:10.1007/978-0-387-46312-4_8.

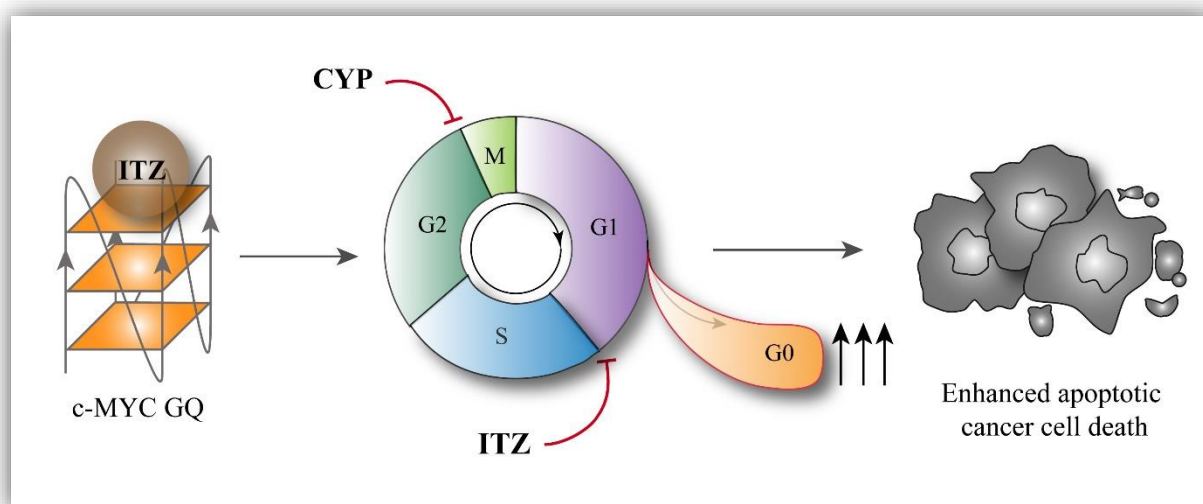
47. Koepfel, F. *et al.* Ethidium derivatives bind to G-quartets, inhibit telomerase and act as fluorescent probes for quadruplexes. *Nucleic Acids Res* **29**, 1087–1096 (2001).
48. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences* **99**, 11593–11598 (2002).
49. Phan, A. T., Modi, Y. S. & Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-MYC Promoter. *J. Am. Chem. Soc.* **126**, 8710–8716 (2004).
50. Pattanayak, R., Basak, P., Sen, S. & Bhattacharyya, M. Interaction of KRAS G-quadruplex with natural polyphenols: A spectroscopic analysis with molecular modeling. *Int J Biol Macromol* **89**, 228–237 (2016).
51. PubChem. Curcumin. <https://pubchem.ncbi.nlm.nih.gov/compound/969516>.
52. Liu, Q., Loo, W. T. Y., Sze, S. C. W. & Tong, Y. Curcumin inhibits cell proliferation of MDA-MB-231 and BT-483 breast cancer cells mediated by down-regulation of NFκB, cyclinD and MMP-1 transcription. *Phytomedicine* **16**, 916–922 (2009).
53. Hosseini, S. A., Zand, H. & Cheraghpour, M. The Influence of Curcumin on the Downregulation of MYC, Insulin and IGF-1 Receptors: A possible Mechanism Underlying the Anti-Growth and Anti-Migration in Chemoresistant Colorectal Cancer Cells. *Medicina (Kaunas)* **55**, 90 (2019).
54. Ji, W. *et al.* c-MYC regulates the sensitivity of breast cancer cells to palbociclib via c-MYC/miR-29b-3p/CDK6 axis. *Cell Death Dis* **11**, 760 (2020).
55. Dexheimer, T. S. *et al.* NM23-H2 may play an indirect role in transcriptional activation of c-MYC gene expression but does not cleave the nuclease hypersensitive element III(1). *Mol Cancer Ther* **8**, 1363–1377 (2009).
56. New Myc-interacting proteins: a second Myc network emerges | Oncogene. <https://www.nature.com/articles/1202725>.
57. Wagner, A. J., Kokontis, J. M. & Hay, N. Myc-mediated apoptosis requires wild-type p53 in a manner independent of cell cycle arrest and the ability of p53 to induce p21waf1/cip1. *Genes Dev* **8**, 2817–2830 (1994).
58. Strasser, A., Harris, A. W., Bath, M. L. & Cory, S. Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2. *Nature* **348**, 331–333 (1990).
59. Vaux, D. L., Cory, S. & Adams, J. M. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-MYC to immortalize pre-B cells. *Nature* **335**, 440–442 (1988).
60. Xu, J., Chen, Y. & Olopade, O. I. MYC and Breast Cancer. *Genes Cancer* **1**, 629–640 (2010).

61. Site-specific amino acid substitution in dodecameric peptides determines the stability and unfolding of c-MYC quadruplex promoting apoptosis in cancer cells | *Nucleic Acids Research* | Oxford Academic. <https://academic.oup.com/nar/article/46/19/9932/5098598>.
62. Gogada, R. *et al.* Curcumin induces Apaf-1-dependent, p21-mediated caspase activation and apoptosis. *Cell Cycle* **10**, 4128–4137 (2011).
63. Di, G., Li, H., Shen, Z. & Shao, Z. [Analysis of anti-proliferation of curcumin on human breast cancer cells and its mechanism]. *Zhonghua Yi Xue Za Zhi* **83**, 1764–1768 (2003).
64. Chaires, J. B. Analysis and interpretation of ligand-DNA binding isotherms. *Methods Enzymol* **340**, 3–22 (2001).
65. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876–2890 (2006).



Chapter 4

Itraconazole and Cyclophosphamide combinational treatment enhances cancer cell death *via c-MYC-GQ- ITZ* interaction at the NHE_{III1} element in MDA-MB-231



4.1 Introduction

Breast cancer is the leading and fatal type of cancer that causes millions of deaths worldwide^{1,2}. Despite an enormous effort by the medical community to combat cancer, triple negative breast cancer (TNBC) remains life-threatening, jeopardizing TNBC diagnosed patients^{3,4}. From discovering a new drug to its clinical use is a long journey with huge financial investment. Health sector research budgets in developing countries like India cannot support the high cost of origin of a new anti-cancer drug. An alternative strategy to evade the long development time of a drug, extortionate budget and risk of high failure rates is repurposing of FDA-approved drugs and combinatorial treatments. Itraconazole (ITZ) is an FDA-approved anti-fungal drug. ITZ has been in clinical use for more than 30 years⁵⁻⁷. Repurposing of ITZ proved its anti-

cancer potential by suppressing angiogenesis. The co-administration of ITZ with paclitaxel had a synergistic anti-cancer effect on HT-29 human colorectal cancer *in vitro* and *in vivo*⁸. ITZ blocks VEGF from binding to its receptor VEGFR2 and inhibits VEGFR2 signaling in HUVEC cell line⁹. ITZ alone or in combination with cyclopamine has recently been found to suppress the progression of medulloblastoma *in vivo* by downregulation of hedgehog (Hh) signaling pathway⁷. The combinatorial treatment of ITZ with rapamycin is reported to act synergistically to arrest MDA-MB-231 TNBC cells in G₀/G₁ and restrict G₁/S phase progression but the combinatorial treatment did not enhance apoptosis¹⁰. In non-small cell lung cancer (NSCLC), ITZ reportedly enhanced the anticancer effect of cisplatin¹¹. ITZ is reported to inhibit cancer cell motility and proliferative rate in TNBC *via* negatively regulating Akt/mTOR pathway¹⁰. ITZ induced apoptosis in MCF-7 cell line by altering the mitochondrial membrane potential, downregulating BCL-2 and upregulating caspase-3 activity⁵. The reported repression of VEGFR2, mTOR, and Hh signaling by ITZ attribute to the suppression of angiogenesis in cancer⁵. ITZ is reported to suppress metastatic progression in pancreatic cancer by downregulation of TGF β /SMAD2/3 signaling and suppression of EMT¹². Evidences on the combinatorial treatment of ITZ with chemotherapy drugs in treating gastric cancer and NSCLC patients is promising for TNBC patients¹⁰. Cyclophosphamide (CYP) is a cytostatic DNA crosslinking agent used in standard breast cancer treatment regimen^{13–15}. It is an alkylating antineoplastic agent. A combination of rapamycin and cyclophosphamide is reported to promote apoptosis and suppress tumor cell proliferation in TNBC mice model¹⁶. Cumulative dosage is the major risk factor of CYP associated nephrotoxicity and hepatotoxicity¹⁷.

In the previous chapter, I discussed about the stabilizing effect of CUR on c-MYC GQ. In this chapter, we have studied the binding interaction of ITZ with the master regulator GQ element in the *c-MYC* gene promoter *biophysically and in vitro* and noticed a thermodynamically favorable binding interaction between ITZ and *c-MYC* GQ. *c-MYC* GQ-ITZ interaction was

confirmed in vitro by dual luciferase assay, as we observed the targeted repression of *c-MYC* promoter activity on treatment with ITZ. Next, we assessed the impact of ITZ treatment on *c-MYC* oncogene RNA expression and tried to unveil the regulatory event taking place at the NHE_{III1} element of *c-MYC* in ITZ's presence. We studied the endogenous modulatory effect of ITZ at the NHE_{III1} element encompassing the *Pu-27* GQ sequence by performing chromatin immunoprecipitation assay, which revealed ITZ increases the occupancy level of Sp1 and nucleolin (NCL) at the NHE_{III1} element that resulted in downregulation of *c-MYC*.

Based on the previous reports on the potentiality of ITZ as an anticancer drug through the inhibition of cancer cell growth, migration and invasion in different tumor types. We investigated the effect of the combinatorial treatment of MDA-MB-231 TNBC cell-line with ITZ and CYP. Our target was to observe the effect of this combinatorial drug formulation on triple-negative metastatic breast cancer cells. The observed higher efficacy of the low dose combinational treatment may help us to reduce the adverse toxic effects of the chemotherapy drug cyclophosphamide.

4.2 Results and Discussion

4.2.1 ITZ suppresses cell proliferation and reduces the adherence property of MDA-MB-231 cells

Cell staining with Crystal Violet revealed ITZ to decrease cell proliferation and reduce the adherence property of MDA-MB-231 cells with increasing concentration of ITZ [Figure 1A]. ITZ affected cells are presumed to lose their adherence and washed off from the culture plate during crystal violet staining. Although we detected a significant reduction in crystal violet staining indicating ITZ triggered cell death. further assessment is essential to determine the cause of reduced crystal violet staining.

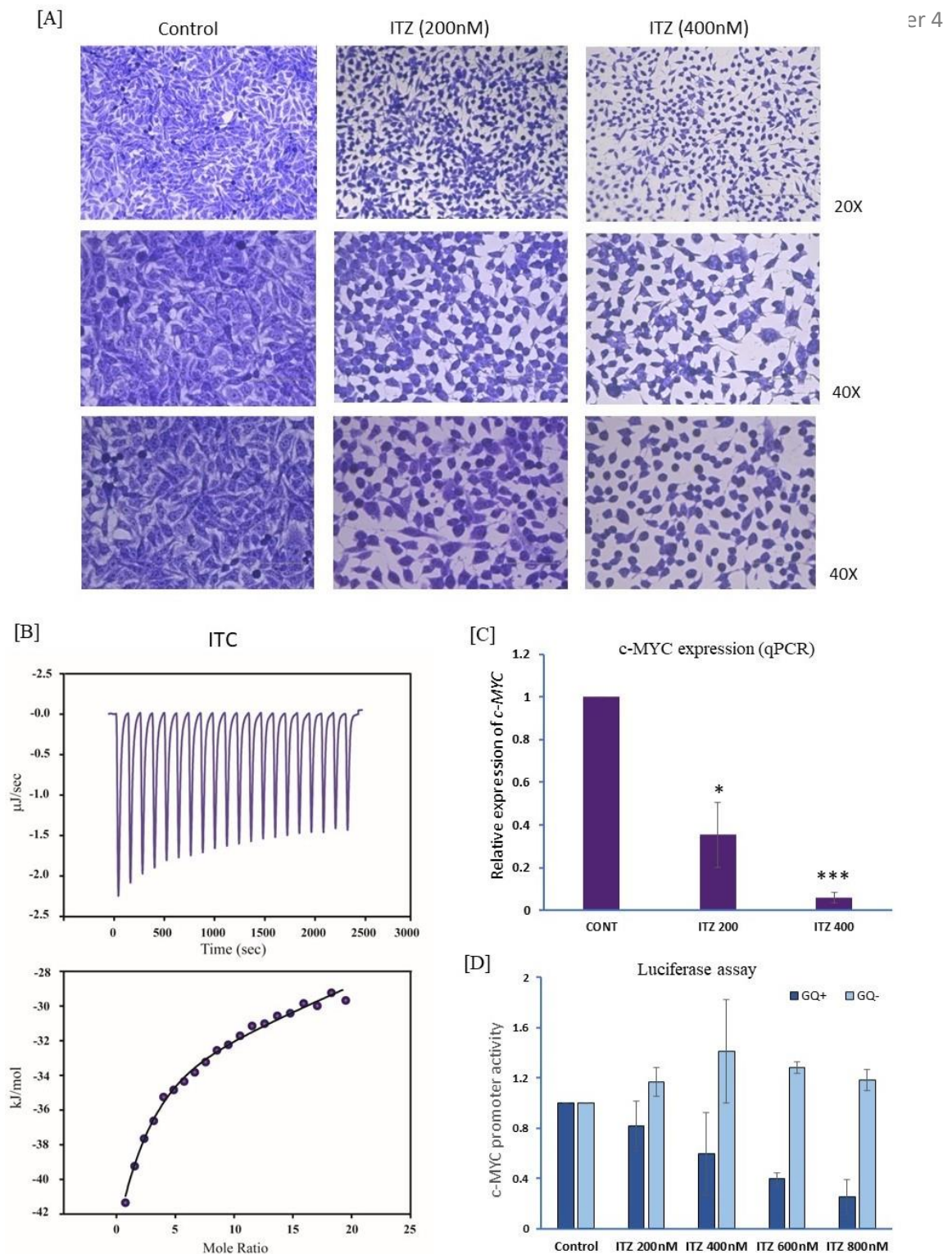


Figure 1. [A] Crystal violet staining of MDA-MB-231 cells after ITZ (200nM and 400nM) treatment. Cancer cells with initiated apoptosis detaches from the surface upon crystal violet staining of control and ITZ treated MDA-MB-231 cells. **[B] Calorimetry data (ITC)** of MYC22 with ITZ. The upper portion shows the isothermal plot of MYC GQ+ITZ complex formation. **[C] Relative RNA expression** of c-MYC upon treatment with ITZ (200nM and 400nM). GAPDH is considered the positive control. **[D] Dual Luciferase assay** of ITZ treated MDA-MB-231 cells transfected with MYC GQ+ and MYC GQ- promoter region inserted clones. Upon ITZ treatment, constructs with the quadruplex motif (GQ+) upstream of the c-MYC promoter exhibit significantly reduced luciferase activity compared to the GQ null clone with the deletion construct. Error bars in the histogram plots display the mean of normalized values \pm SE from triplicate data.

4.2.2 ITZ downregulates *c-MYC* gene expression at the RNA-level in MDA-MB-231 cell line

Since, we observed decrease in cell proliferation upon ITZ treatment in MDA-MB-231 cells, we evaluated the transcriptional level of the proliferative marker gene *c-MYC*. We observed a significant downregulation of *c-MYC* expression upon treatment with increasing concentration of ITZ. At 200nM ITZ *c-MYC* was observed to be downregulated by 2.8-fold and at 400nM ITZ concentration, *c-MYC* expression reduced by 17-fold [Figure 1C]. The fold change observed indicates *c-MYC* as the plausible target of ITZ.

4.2.3 Thermodynamically favorable interaction of ITZ with *c-MYC* GQ

We performed ITC profiling to inspect the binding interaction of *c-MYC* GQ and ITZ. The thermogram profile revealed the thermodynamically favorable interaction between ITZ and *c-MYC* GQ [Figure 1B]. We obtained a positive value for change in entropy (ΔS), a negative value for change in enthalpy (ΔH), and an overall negative Gibbs free energy change (ΔG), indicating the spontaneity of the interaction. Table 1 contains the thermodynamic parametric values that were obtained from the calorimetric study.

Sequential 3 site binding parameters			
ΔG (kJ)	ΔH (kJ/mol)	ΔS (J/mol·K)	K_d (M)
-26.4	-4407	-1.47E+04	3.74E-03
-12.4	4845	1.63E+04	9.12E-05
-19.4	-2165	-7.20E+03	3.85E-04

Table 1. Thermodynamic parameters of ITZ-*c-MYC*-GQ interaction. ΔG represents Gibbs free energy change, ΔH is change in enthalpy, ΔS is entropic change, K_d is the dissociation constant of the binding interaction. ITC data was best fit in sequential-3-site fitting model.

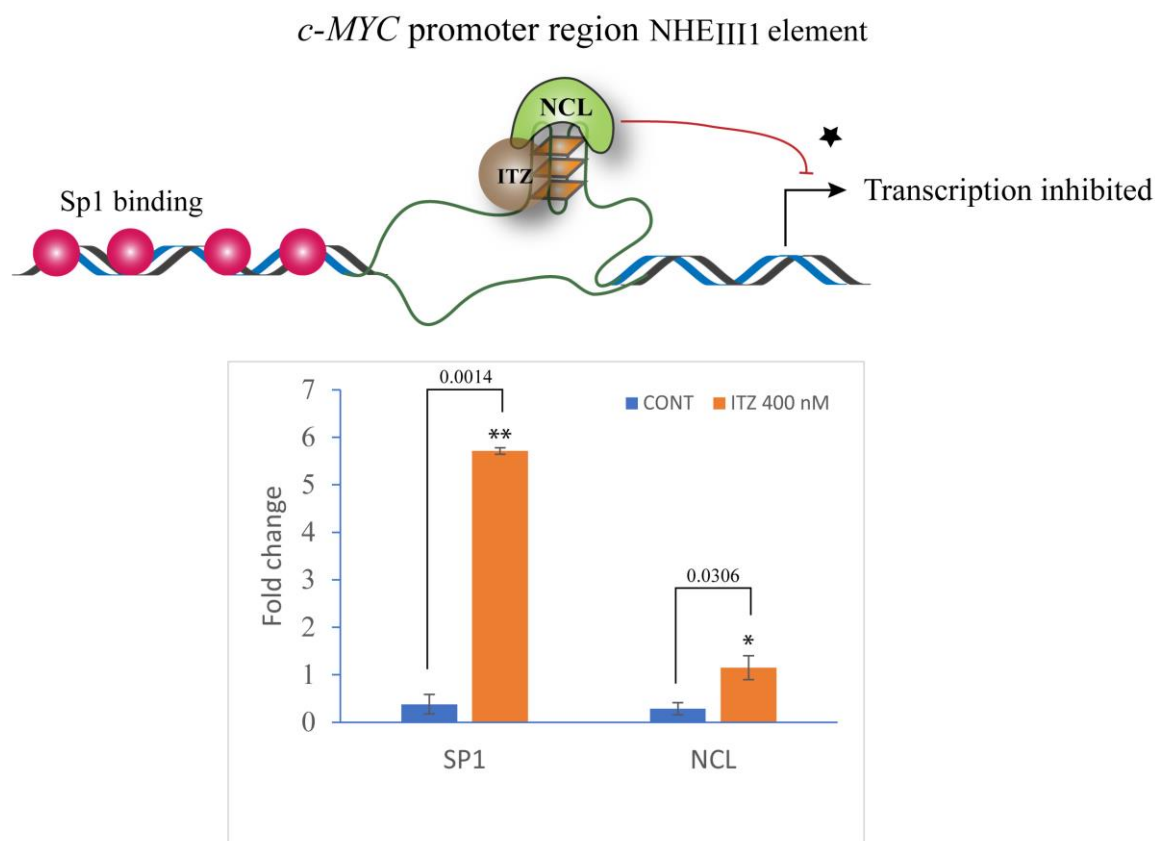


Figure 2. ChIP assay. Chromatin immunoprecipitation of *c-MYC* region encompassing the NHEIII1 element upon ITZ treatment of 400nM. SP1 and NCL TF occupancy increased upon ITZ treatment. ITZ is speculated to bind *c-MYC* GQ, facilitating the recruitment of NCL that further stabilizes the silencer element, leading to the downregulation of *c-MYC* transcription. Error bars show means \pm SE of triplicate data. Asterisks (*) indicate significance (* $P < 0.05$, ** $P < 0.01$) from paired t-test denoting significant difference in occupancy level or binding of the transcription factors.

4.2.4 ITZ specifically targets *c-MYC* GQ and downregulates its promoter function

We assessed the promoter activity of *c-MYC* upon treatment with various doses of ITZ. We observed significant repression of luciferase activity in MDA-MB-231 with MYC-GQ-positive clones upon treatment with increasing concentration of ITZ w.r.t. that of MYC-GQ-null clones. Our conducted dual luciferase assay confirms the efficiency of ITZ in targeting *c-MYC*-GQ in-cellulo leading to the downregulation of the gene promoter activity. We observed around 2-fold repression at 400nM, 2.5-fold repression at 600nM, and 4-fold at 800nM [Figure 1D]. ITZ attenuated *c-MYC* gene promoter activity with the GQ+ insert in a dose-dependent manner in

comparison to the treated deletion inserts. Thus, ITZ selectively binds and stabilizes MYC-GQ resulting in downregulation of the *c-MYC* promoter activity.

4.2.5 ITZ stabilizes endogenous *c-MYC* GQ by facilitating the recruitment to NCL

We speculated ITZ to interact with the GQ structure in the NHE_{III1} element of *c-MYC* and downregulate its expression. We observed the enrichment in the occupancy level of Sp1 and NCL TFs at the NHE_{III1} element of *c-MYC* upon treatment with ITZ, indicating the stabilization of the silencer regulatory switch *Pu-27* and downregulation of *c-MYC* oncogene [Figure 2]. Increased binding of Sp1 induces greater torsional stress that releases more of ssDNA unbiasedly facilitating the formation of GQ or iM. We hypothesized the favored evolution and stabilization of GQ by the combined action of ITZ, Sp1 and NCL that resulted in the downregulation of *c-MYC* gene.

4.2.6 ITZ and CYP arrest cells at different phases that synergizes to enhance cancer cell death

The proportion of cells in different cell cycle phases [G₀-G₁, S, G₂-M phases] were analyzed using FACS Verse, BD Biosciences. MDA-MB-231 cells on ITZ and CYP treatment singly and in combination, significantly changed the cell cycle phase distribution. We observed a potent inhibitory effect of both the drugs ITZ and CYP. Cell cycle got arrested at G₀/G₁ phase by ITZ, while upon CYP treatment, the cell cycle got arrested at G₂/M phase. The inhibitory effect on MDA-MB-231 was observed to be enhanced upon the combinatorial treatment with ITZ and CYP. The combinational treatment drastically arrested the cell population at sub-G₁ phase indicating massive apoptotic cell death [Figure 3 and 4]. We speculated that ITZ sensitized the cells to apoptotic cell death and facilitated CYP induced DNA damage. DNA multimers leakage during ethanol fixation reduced the DNA content resulting in low PI staining in the CYP treated state (alone and in combination). The heightened sub G₀/G₁ population to

the left of G0/G1 peak represents elevated apoptosis upon the combinational treatment [Figure 4].

4.2.7 ITZ and CYP combinational treatment enhances cancer cell death

We investigated the mechanism of cell death initiated by CYP and ITZ+CYP by performing flow cytometric apoptosis assay. We observed substantially increased apoptosis in MDA-MB-231 upon treatment with ITZ+CYP in comparison to CYP treatment [Figure 5]. [CYP (40 μ M) + ITZ (400nM)] treatment resulted in 2.5-fold increase in apoptotic cell death compared to CYP (40 μ M) treatment, while [CYP (60 μ M) + ITZ (600nM)] treatment resulted in 4-fold increase in apoptotic cell death compared to CYP (60 μ M) in the highly metastatic cells. Hence, we assumed the combinational drug treatment to enhance the efficiency of CYP in killing of the highly metastatic breast cancer cells. The approval of the (ITZ+CYP) combinatorial treatment shall reduce the dose of cyclophosphamide to treat breast cancer patients curbing down the adverse toxic side effects of cyclophosphamide.

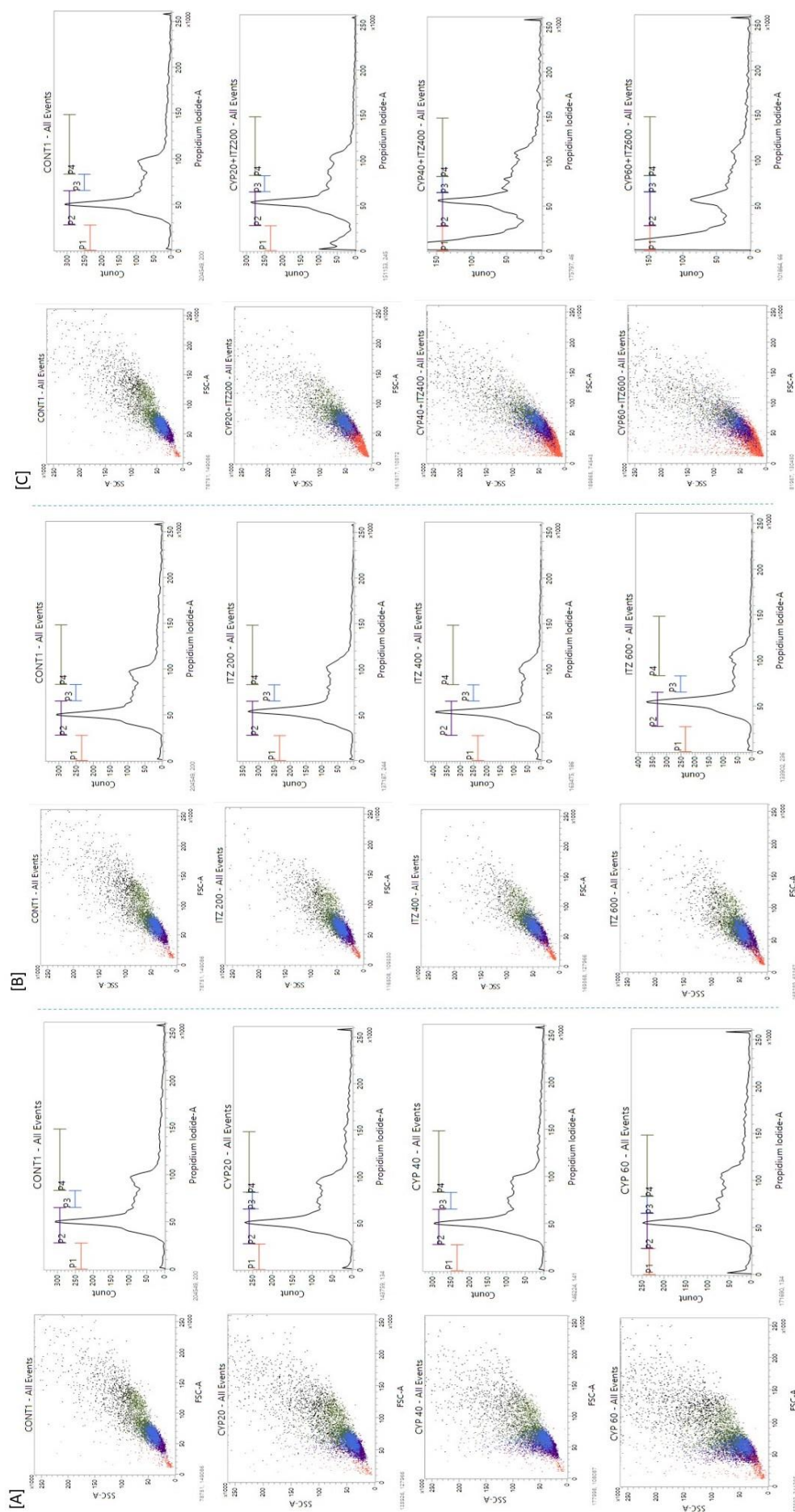


Figure 3. Cell cycle stage analysis of drug-treated MDA-MB-231 cells: [A] cyclophosphamide (CYP) 20, 40, and 60 μ M, [B] with different concentrations of Itraconazole (ITZ) 200, 400, and 600nM, [C] with the combinational treatment (CYP and ITZ) 20 μ M+200nM, 40 μ M+400nM, 60 μ M+600nM. CYP treatment increased the granularity of the cancer cells as indicated by the side scattering of light (SSC) suggesting loss of cell membrane integrity upon treatment with CYP. The combinational treatment also increased the cells' side scattering (SSC). P1 comprises sub-G0 phase, P2 is the G0-G1 phase, P3 represents synthetic S phase, and P4 shapes the G2-M phase.

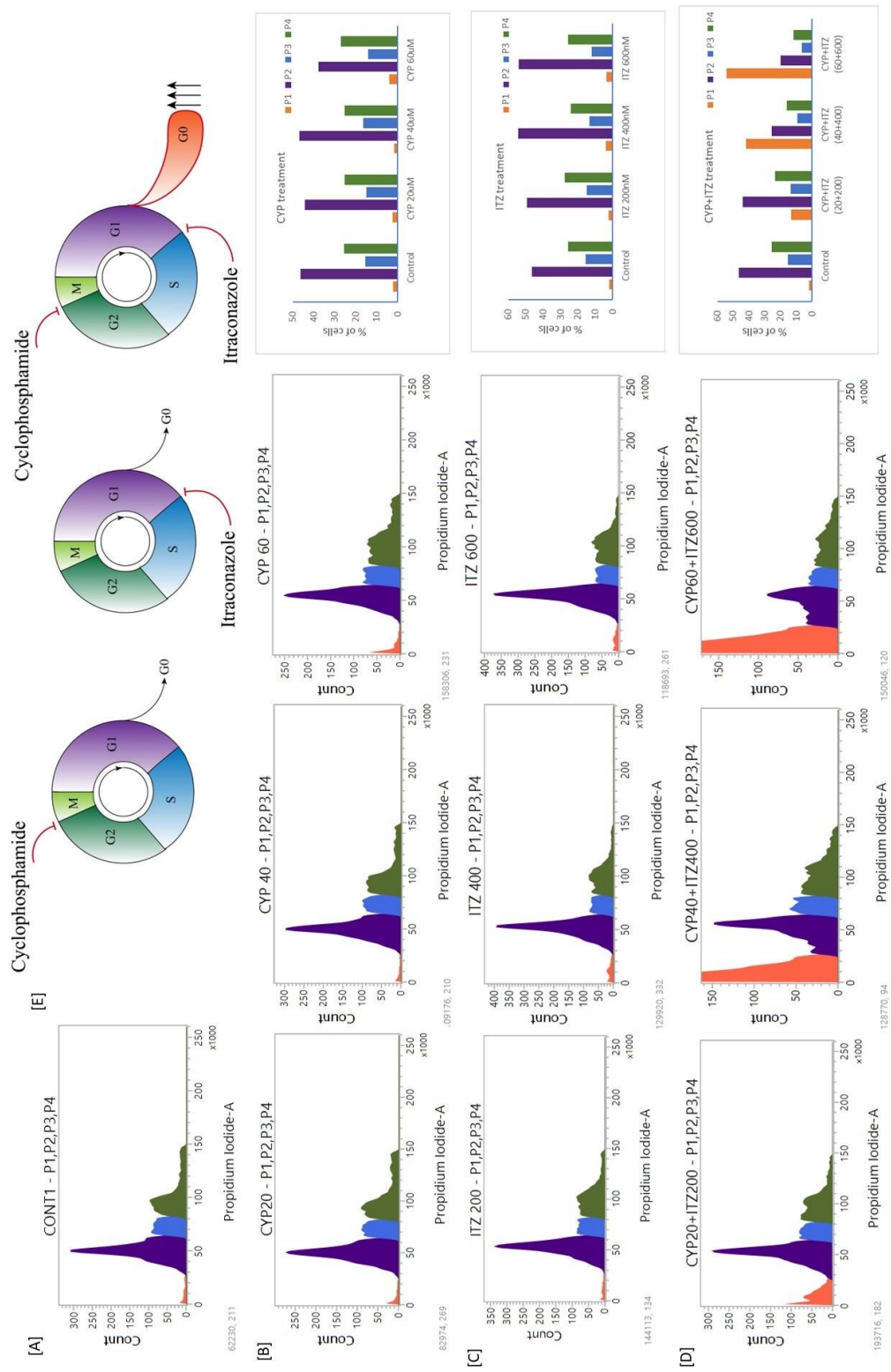


Figure 4. Comparative analysis of cell cycle phases under different conditions. [A] Control, [B] Cell cycle phases upon treatment with Cyclophosphamide (CYP) 200nM, 400nM, and 600nM. CYP results in G2/M phase arrest, [C] Cell cycle phases upon treatment with Itraconazole (ITZ) 200nM, 400nM, and 600nM. ITZ results in G0/G1 phase arrest. [D] Cell cycle phases upon treatment with CYP+ITZ (200nM+200nM, 400nM+400nM, and 600nM+600nM). [E] Schematic representation of cell cycle upon treatment with ITZ, CYP, AND (ITZ+CYP). The combinational treatment results in drastic accumulation (or arrest) of the cancer-cells at G0- and sub-G1 phase indicating enormous rate of apoptotic cell death upon the combinational drug treatment.

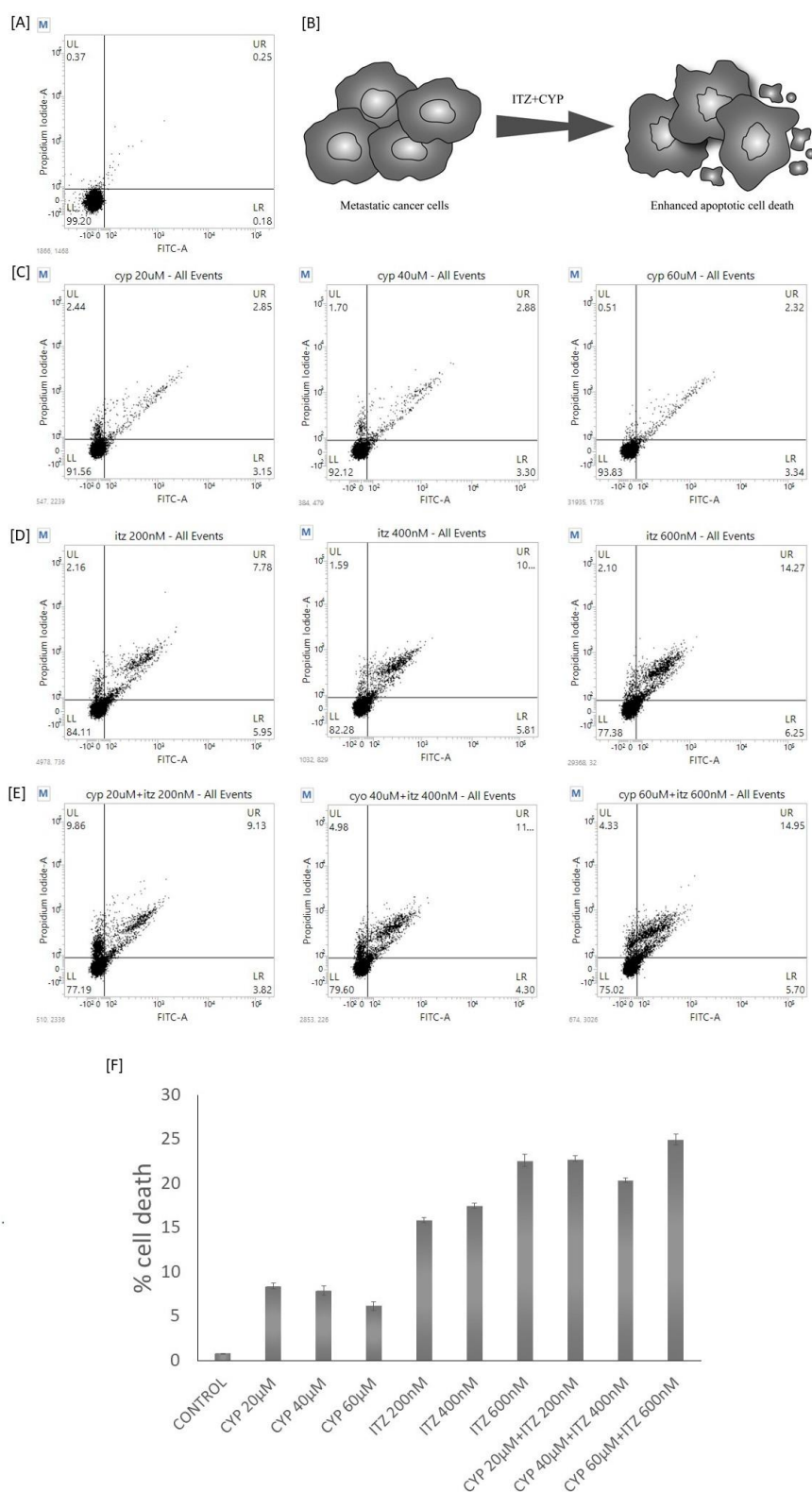
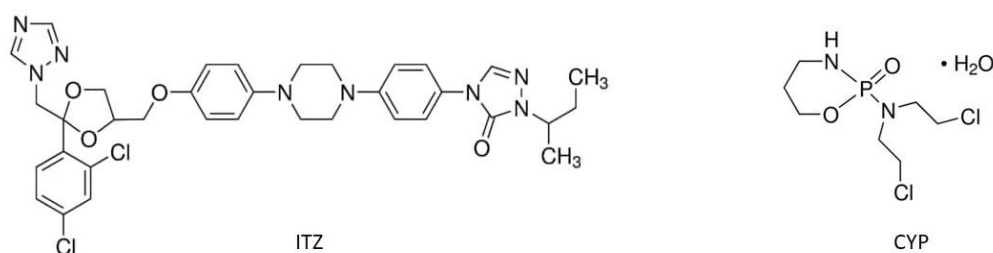


Figure 5. Enhanced apoptotic cell death of cancer cells upon combinational treatment of cyclophosphamide (CYP) and Itraconazole (ITZ) as indicated by increased FITC-A and PI-stained cells. [A] control without drug treatment, [B] schematic representation of apoptotic cell death upon CYP+ITZ treatment, [C] cell death analysis upon CYP (20μM, 40μM, and 60μM) treatment, [D] cell death analysis upon ITZ (200nM, 400nM, and 600nM) treatment. [E] cell death analysis upon (CYP+ITZ) combinational drug (20μM+200nM, 40μM+400nM, 60μM+600nM) treatment. [F] Bar plot representing the percentage of cell death under different drug concentrations alone and in combination.

4.3 Conclusion

Chemotherapy is the major therapeutic treatment available for TNBC patients. High aggressiveness, high metastasizing potential, chemoresistance and cancer relapse are responsible for the high mortality rate among cancer patients. ITZ is extremely hydrophobic, its insolubility limits the efficacy as a drug. ITZ at higher doses inhibits hepatocyte CYP3A4, the main cytochrome P450 in the human liver that is responsible for metabolizing xenobiotics¹⁸. Inhibition of CYP3A4 by ITZ that hinders the metabolism of most chemo drugs is of a major concern. Although CYP is a highly effective drug, CYP chemo treatment comes with several major side effects which includes leukopenia, haemorrhagic cystitis, pulmonary and hepatotoxicity^{19,20}. Our study proposes combining Itraconazole (ITZ) and cyclophosphamide (CYP) for TNBC treatment. The combinatorial treatment resulted in significant level of apoptosis at a low concentration level of both the drugs in MDA-MB-231 cell line. ITZ acts by targeted downregulation of the proliferative gene *c-MYC*, thereby sensitizing the cancer cells to apoptotic cell death. The ITZ+CYP combinational treatment led to massive cell death rate which is a promising lead to the use of low concentration of the drugs but with enhanced efficiency.

4.4 Material and methods



Sample preparation

Both the drugs were purchased from Sigma Aldrich. Itraconazole (ITZ) (I6657) and Cyclophosphamide (CYP) (C7397). Itraconazole (ITZ) was dissolved in DMSO at 50mg/ml stock concentration as per the company's instruction sheet. Cyclophosphamide was dissolved in water at a stock concentration of 40mg/ml.

Cell culture

MDA-MB-231 breast cancer cells were cultured and maintained in Gibco Dulbecco's Modified Eagle Medium DMEM (Cat. # 11995-065) supplemented with 10% (v/v) FBS, 1% Pen-Strep, and 2.5 µg/ml Amphotericin B, and 50 µg/ml Gentamicin. The model number of the CO₂ Incubator is same as mentioned in chapter 3. A stable temperature of 37°C with 5% CO₂ supply and humid environment was maintained in the incubator for proper mammalian cell culture. Cell passages were maintained in T-25 culture flasks suitable for adherent cells. Cells were passaged every 3 to 4 days with media replenishment every 24 hours to maintain healthy growth of cells.

Crystal violet staining

MDA-MB-231 cells were cultured in 6 well plate at a concentration of 1×10^5 cells and treated with ITZ (200nM and 400 nM) for 24 hours. The media was discarded and wells were washed twice with 1X PBS and fixed in ice-cold 100% methanol. Fixation was carried out for 10 minutes on ice. After proper fixation, the cell- staining was done with 0.5% crystal violet prepared in 25% methanol, and incubated for 10 minutes. The cells were carefully washed several times to remove the extra dye. The cells were allowed to dry at RT before imaging in EVOS M5000 Invitrogen.

Quantitative PCR (qPCR)

qPCR samples were run in QuantStudio5 instrument to evaluate the change in RNA expression of c-MYC gene. MDA-MB-231 cells were seeded at a density of 1.5×10^5 cells/well in 6 well plate. Upon reaching 60% confluency, cells were treated with ITZ at selected doses and incubated for 24 hours. RNA isolation done using TRIzol® reagent (Invitrogen) followed by 1 µg of cDNA preparation for each sample. Specific qPCR amplification of c-MYC and GAPDH were done using POWERUp SYBR Green master mix (Cat# A25742) as per the given

instructions and gene specific primers. GAPDH was used as an internal control to normalize c-MYC mRNA expression variability. Relative fold change was calculated using $2^{-\Delta\Delta CT}$ method.

$$\Delta CT = CT(\text{Target}) - CT(\text{GAPDH})$$

$$\Delta\Delta CT = \Delta CT(\text{Untreated}) - \Delta CT(\text{Treated})$$

$$\text{Fold change (R)} = 2^{-\Delta\Delta CT}$$

Primers	Oligo Sequences (5'-3')	Annealing Temperature (°C)
c-MYC FP	F: AATGAAAAGGCCCCCAAGGTAGTTATCC	55
c-MYC RP	R: GTCGTTTCCGCAACAAGTCCTCTTC	
GAPDH Forward Primer	F: GATGCTGGCGCTGAGTACGTCGTG	60
GAPDH Reverse Primer	R: AGTGATGGCATGGACTGTGGTCATGAG	

ITC

ITC experiment was performed in iTC Microcalorimeter. Temperature was maintained at 298K throughout the experiment. 20 μM of c-MYC GQ was loaded in cell and 600 μM of ITC was loaded in syringe. The buffer environment of KP-KCl was kept the same in both the cell and the syringe to avoid the heat of dilution during the titration. The samples were properly degassed prior to the titration. 20 injections, each at an interval of 120 seconds with 4.5 μl per injection was set. For proper curve fit data, a blank titration was also performed where ITZ was injected into cell containing only buffer (no oligo), both with same percentage of DMSO to normalize the heat of dilution. The heat of interaction was generated after the subtraction of the heat of dilution. The corrected values were plotted against molar ration. NanoAnalyzer with inbuilt Origin 7.0 software was used to generate the best fir values of change in entropy (ΔS),

enthalpic change (ΔH), dissociation constant (K_d), and binding stoichiometry (n) which were further used to derive the Gibbs free energy change of the interaction study.

$$\Delta G = \Delta H - T\Delta S$$

Dual luciferase assay

The intracellular selectivity of ITZ for MYC- promoter GQ study was done by performing dual luciferase experiment. MDA-MB-231 cells were seeded in 96 well plate at a density of 0.05×10^5 cells/well. Upon reaching 70% confluency, the cells were transfected with pGL4.72 vectors with GQ⁺ or GQ⁻ inserts along with the reference control plasmid pGL3. The same clones as used in chapter 3 has been used here. Transfection was done using Lipofectamine 2000. For each well, 200 ng of plasmid was dissolved in 25 μ l of OPTI-MEM media and kept for 5 minutes for incubation. 0.5 μ l of lipofectamine was diluted in 25 μ l of OPTI-MEM and incubated for 5 minutes. Diluted plasmid and diluted lipofectamine were mixed and further incubated at 37°C for 30 mins. 50 μ l of the incubated solution was then pipetted into each well. Transfection was carried out for 4 hours, followed by growth media change and drug treatment. After 24 hours of ITZ treatment at 37C and 5% CO₂ level, the dual luciferase assay was performed using Promega dual luciferase kit as per the kit guidelines. 20 μ l of 1X PLB (passive lysis buffer) was added to each well and incubated for 20 minutes under rocking condition. Prior to luminescence measurement, 50 μ l of LARII was added and luminescence measured for firefly, then 50 μ l of STOP&GLOW was added and luminescence measured for Renilla. Firefly/Renilla luminescence were normalized and data plotted with respect to control set to evaluate the change in promoter activity.

Chromatin immunoprecipitation assay

The physical interaction of SP1 and NCL was investigated at the NHE_{III} element of *c-MYC* promoter region by performing chromatin immunoprecipitation assay. We used Pierce

Magnetic ChIP Kit (Cat no. 26157) to conduct the experiment. The reagents, antibodies, primers, and the procedure followed was same as described in chapter 3.

Cell cycle analysis

MDA-MB-231 cells were seeded at 8×10^4 cells/well of a 6 well plate. Cells received treatment of CYP (20 μ M, 40 μ M, and 60 μ M), ITZ (200 nM, 400 nM, 600 nM), and CYP+ITZ [(20 μ M +400 nM), (40 μ M +400 nM), and (60 μ M +600 nM)] for 24 hours. The cells then underwent 1X PBS wash followed by fixation with 70% ethanol. The ethanol was added dropwise to ensure proper fixation and incubated at 4°C for 45 minutes. The fixed cells again underwent 1X PBS wash twice and then treated with 50 μ l of 100 μ g/ml of RNase to ensure DNA staining. 10 μ l of 500 μ g/ml or 748.5 μ M propidium iodide (PI) staining for 10 minutes at 4°C with volume made upto 500 μ l before flow cytometric measurements [forward scatter (FS) and side scatter (SS)].

Apoptosis assay

Apoptosis detection kit (Catalog no. V13242) was used to assay apoptosis. MDA-MB-231 cells were seeded at a concentration of 8×10^4 cells/ well plate and treated with CYP (20 μ M, 40 μ M, and 60 μ M), ITZ (200 nM, 400 nM, 600 nM), and CYP+ITZ [(20 μ M +400 nM), (40 μ M +400 nM), and (60 μ M +600 nM)] for 24 hours. Next, the cells underwent trypsinization and collection (harvest) followed by wash with 1X PBS. The cell pellet was resuspended in 1X annexin binding buffer as per the kit manufacturer's protocol. Next, 5 μ l of FITC Annexin V followed by 100 μ g/ml of PI and incubated at RT for 30 mins. The volume of each sample was made upto 500 μ l before flow cytometric analysis of the stained cells.

4.5 References

1. Łukasiewicz, S. *et al.* Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers (Basel)* **13**, 4287 (2021).

2. Ataollahi, M., Sharifi, J., Paknahad, M. & Paknahad, A. Breast cancer and associated factors: a review. *J Med Life* **8**, 6–11 (2015).
3. Almansour, N. M. Triple-Negative Breast Cancer: A Brief Review About Epidemiology, Risk Factors, Signaling Pathways, Treatment and Role of Artificial Intelligence. *Front Mol Biosci* **9**, 836417 (2022).
4. Jitariu, A.-A., Cîmpean, A. M., Ribatti, D. & Raica, M. Triple negative breast cancer: the kiss of death. *Oncotarget* **8**, 46652–46662 (2017).
5. Li, C.-L. *et al.* Repurposed itraconazole for use in the treatment of malignancies as a promising therapeutic strategy. *Biomedicine & Pharmacotherapy* **154**, 113616 (2022).
6. Panda, N. K. Itraconazole — a potent antifungal drug. *Indian J Otolaryngol Head Neck Surg* **49**, 293–294 (1997).
7. Kim, J. *et al.* Itraconazole, a Commonly Used Antifungal that Inhibits Hedgehog Pathway Activity and Cancer Growth. *Cancer Cell* **17**, 388–399 (2010).
8. Ghadi, M., Hosseini-mehr, S. J., Amiri, F. T., Mardanshahi, A. & Noaparast, Z. Itraconazole synergistically increases therapeutic effect of paclitaxel and ^{99m}Tc-MIBI accumulation, as a probe of P-gp activity, in HT-29 tumor-bearing nude mice. *European Journal of Pharmacology* **895**, 173892 (2021).
9. Nacev, B. A., Grassi, P., Dell, A., Haslam, S. M. & Liu, J. O. The Antifungal Drug Itraconazole Inhibits Vascular Endothelial Growth Factor Receptor 2 (VEGFR2) Glycosylation, Trafficking, and Signaling in Endothelial Cells. *J Biol Chem* **286**, 44045–44056 (2011).
10. Wu, H.-T. *et al.* Induced Cell Cycle Arrest in Triple-Negative Breast Cancer by Combined Treatment of Itraconazole and Rapamycin. *Frontiers in Pharmacology* **13**, (2022).
11. Aftab, B. T., Dobromilskaya, I., Liu, J. O. & Rudin, C. M. Itraconazole inhibits angiogenesis and tumor growth in non-small cell lung cancer. *Cancer Res* **71**, 6764–6772 (2011).
12. Chen, K. *et al.* Itraconazole inhibits invasion and migration of pancreatic cancer cells by suppressing TGF- β /SMAD2/3 signaling. *Oncol Rep* **39**, 1573–1582 (2018).
13. Johnson, L. A. *et al.* Formation of Cyclophosphamide Specific DNA Adducts in Hematological Diseases. *Pediatr Blood Cancer* **58**, 708–714 (2012).
14. Cheung-Ong, K., Giaever, G. & Nislow, C. DNA-Damaging Agents in Cancer Chemotherapy: Serendipity and Chemical Biology. *Chemistry & Biology* **20**, 648–659 (2013).
15. Wang, J.-Y., Prorok, G. & Vaughan, W. P. Cytotoxicity, DNA cross-linking, and DNA single-strand breaks induced by cyclophosphamide in a rat leukemia in vivo. *Cancer Chemother. Pharmacol.* **31**, 381–386 (1993).
16. Zeng, Q. *et al.* Treating triple-negative breast cancer by a combination of rapamycin and cyclophosphamide: An in vivo bioluminescence imaging study. *European Journal of Cancer* **46**, 1132–1143 (2010).

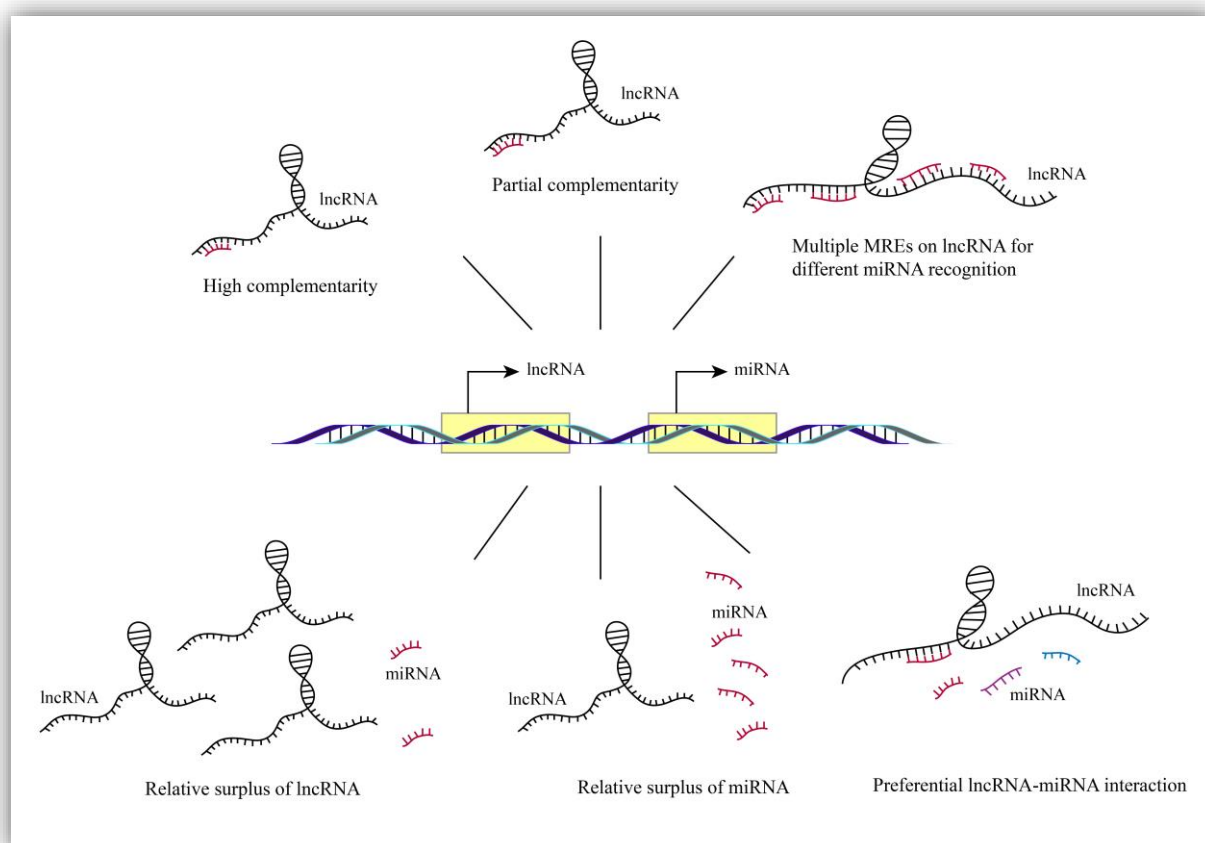
17. Ayza, M. A. *et al.* Potential Protective Effects of Antioxidants against Cyclophosphamide-Induced Nephrotoxicity. *International Journal of Nephrology* **2022**, e5096825 (2022).
18. Isoherranen, N., Kunze, K. L., Allen, K. E., Nelson, W. L. & Thummel, K. E. Role of Itraconazole Metabolites in Cyp3a4 Inhibition. *Drug Metab Dispos* **32**, 1121–1131 (2004).
19. Malik, S. W., Myers, J. L., DeRemee, R. A. & Specks, U. Lung toxicity associated with cyclophosphamide use. Two distinct patterns. *Am J Respir Crit Care Med* **154**, 1851–1856 (1996).
20. Khan, M. A. Immune potentiating and antitoxic effects of camel milk against cyclophosphamide-induced toxicity in BALB/C mice. *Int J Health Sci (Qassim)* **11**, 18–22 (2017).



Chapter 5

Insights into the role of lncRNAs and miRNAs in cancer progression and metastasis

This chapter is based on the review article 'Long noncoding RNAs in cancer immunity: a new avenue in drug discovery' Drug Discov Today 2021 Jan;26(1):264-272. doi: 10.1016/j.drudis.2020.07.026 and the communicated review 'Evaluation of epigenetic landscape in tumorigenesis: ncRNAs as new targets' Drug Discov Today



5.1 Introduction

Despite significant advancements in cancer treatment, metastasis remains the leading cause behind the mortality of cancer patients. Patients with recurrent metastasis have a maximum life span of 5 years after diagnosis, thereby challenging the medical community to effectively treat metastatic cancer¹. Metastasis promoting genes expressed in cancer cells in the primary tumor,

seeds the cells with the ability to disseminate and metastasize to distant organs¹. Phenotypic plasticity is a clever property of the malignant cells that allow these cells to flexibly adopt diverse phenotypic states to bypass metastatic barriers and drug resistance according to favorable cellular and environmental stimuli¹. Few cancer cells acquire multiple oncogenic mutations to become cancer stem cells (CSCs). With activation of an additional set of metastasis-promoting genes, CSCs acquire metastasis initiating capability (MICs), while retaining their stemness property. In invasive carcinomas like breast and lung, epithelial cells acquire malignant properties with highly plastic regenerative phenotype, ability to evade anoikis and immune surveillance^{1,2}. These cells disseminate from the primary tumor sites and enter the blood circulation as individual cells or cluster of cells termed circulating tumor cells (CTCs). CTCs express EMT markers that primes them to colonize and metastasize to distant organs¹. Epithelial to mesenchymal transition (EMT) is a phenotypic plasticity process where epithelial cells develop properties intermittent between epithelial and mesenchymal phenotype [Figure 1]. The trans-differentiation of cancer cells toward a mesenchymal state contributes to tumor heterogeneity. Cancer cells with activated EMT are featured at the invasive front of carcinoma facilitating cancer cell migration, invasion and dissemination^{1,2}. EMT is driven by SNAIL, ZEB and matrix metalloproteinases (MMPs) along with the involvement of a variety of pleiotropic signals like TGF β , EGF, VEGF, FGF and several lncRNAs functioning through TGF β /SMAD and RAS/MAPK12 signalling pathways^{1,2}. Most of the disseminated cancer cells do not survive in the host tissue due to the hostile environment created by the body's immune system. Bone marrow derived and peripheral immune cells, T cells and NK cells get recruited to newly seeded metastases and destroys it. Despite all these, metastatic cells manage to evade the immune system, corrupt the microenvironment and create a metastatic niche supportive of

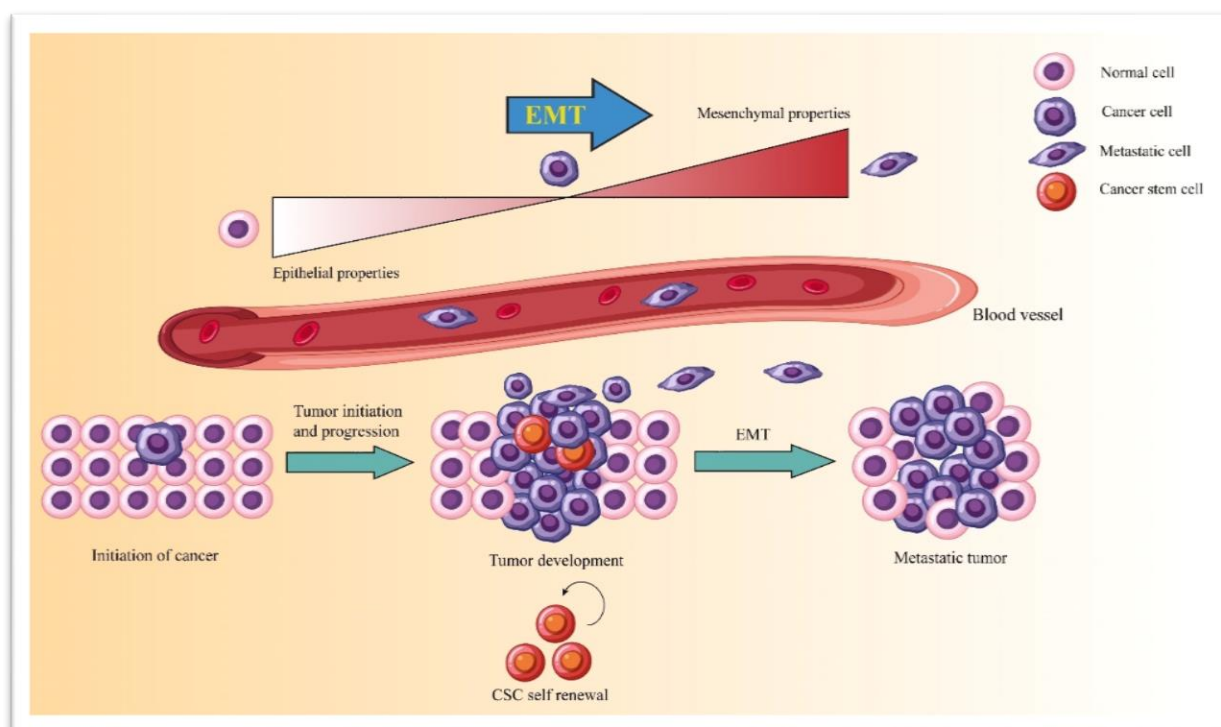


Figure 1. Schematic representation of cancer initiation and progression. Malfunctioning of a cell among the healthy cells results in tumor initiation. Few cells among the primary tumor mass undergoes several mutations to transform into cancer stem cells (CSCs), while few cells acquire mesenchymal properties through EMT. These mesenchymal like cancer cells metastasize to distant organs via the blood vessels and develop secondary tumors or cancer malignancy.

secondary tumor growth¹. Understanding the functional mechanism of cancer cell plasticity and establishment at distant organs with secondary tumor growths are critical in the development of antimetastatic drug¹. Breast cancer is the leading cause of death in females worldwide with most frequently diagnosed malignancy. The most aggressive form being the Triple negative breast cancer (TNBC) with negative HER2, PR, ER receptor signature. The aggressive progression of TNBC towards metastasis is comparatively higher than other types of cancer. Clinical management of TNBC is still a challenge to the scientific community. Deep exploration of the regulatory mechanism of tumorigenesis in TNBC cells will result in identification of effective biomarkers for better diagnosis and prognosis of TNBC patient. Globally breast cancer tops in terms of fatality while lung cancer is the largest contributor to

newly diagnosed cancer³. We have mainly focused our study on TNBC MDA-MB-231 cell line along with a comparative assessment of NSCLC metastatic A-549 lung cancer cell line.

The information described in this chapter is supplementary as it is unrelated to the main body of the thesis work. However, the explanation provided in this chapter is crucial for the understanding of the next chapter that is chapter 6.

5.1.1 Role of non-coding RNAs in cancer

Non-protein coding transcripts like lncRNAs and miRNAs play crucial role in cell differentiation, proliferation, cellular motility and in regulation of apoptotic death. Accumulating studies have identified several lncRNAs and miRNAs to be deregulated in cancer and is directly correlated with the idiosyncrasy of cancer cells.

- **Long non-coding RNAs (lncRNAs)**

Non-coding transcripts greater than 200 nucleotides falls under the group of lncRNAs. These RNAs occupying a large fraction of the human genome, are differentially expressed in tissues with alteration in their behavior in different types of cancers⁴. LncRNAs are generally synthesized by RNA Pol II and are capped, spliced and polyadenylated similar to mRNAs, although most lncRNAs do not function as templates for protein synthesis⁵. Location of lncRNAs is both nuclear and cytosolic depending on their functional target nucleic acids. Advanced sequencing platforms and algorithms have been utilized to collect transcripts from deep RNA-sequencing reads and the number of total lncRNAs are found to be in the range of approximately 20000 transcripts. However, only a limited amount of lncRNAs have been studied in animal models so far⁵.

Among the important functions of lncRNAs revealed till date comprises their role in regulating genomic imprinting, dosage compensation and nuclear organization, gene expression, RNA splicing and in translational control⁵. LncRNAs also act as molecular scaffolds participating in

List of important lncRNAs significantly affecting cancer metastasis.

lncRNAs	Expression in cancer	Effect	Reference
lncRNA HOTAIR	Upregulated	Drives NSCLC cancer cells and breast cancer towards malignancy.	⁶
lncRNA MALAT1	Upregulated	In breast cancer and lung cancer is positively correlated with immune suppression. MALAT1 is also reported to trigger EMT and CSC activity, thereby promoting chemoresistance and radio-resistance in cancer cells.	⁷
lncRNA MEG3	Upregulated	In lung cancer is responsible for increased chemoresistance.	⁸
lncRNA HAS2-AS1	Upregulated	It is a marker of tumor hypoxic condition. It assists in enhancing migratory and metastatic property of cancer cells.	^{9,10}
lncRNA ZEB2NAT	Upregulated	Upregulates EMT markers: ZEB1, N-cadherin, matrix metalloproteinases (MMP2 and MMP9) while suppressing the epithelial marker E-cadherin.	¹¹
lncRNA GAS5	Downregulated	Upregulated EMT markers in NSCLC.	¹²
lncRNA NKILA	Downregulated	Elevated metastasis in breast cancer.	^{2,13}
lncRNA NEAT1, lncRNA MALAT1	Knockdown	Suppress invasiveness of cancer in breast and lung cancer.	^{2,14,15}
lncRNA COX2	Knockdown	Reduced metastatic property of lung cancer cells.	¹⁶
lncRNA ZEB2NAT	Knockdown	Suppresses EMT, metastasis and proliferation in breast cancer cells	¹¹
lncRNA PNUTS	Knockdown	Reduced metastatic property of breast cancer.	^{2,17}
lncRNA BCAR4	Knockdown	Reduced breast cancer in mice model.	¹⁸

Table 1: List of few important lncRNAs that significantly affect cancer cell properties.

intermolecular interactions with RNAs and proteins. Thus, lncRNAs regulate gene expression at different hierarchical levels accentuating their censorious identity in cellular pathologies such as cancer⁴. lncRNAs have the potential to be promising prognostic and diagnostic biomarkers that can be therapeutically targeted to combat TNBC/cancer. List of important lncRNA that have significant impact on deregulation of cancer cells is provided in *Table 1*.

- **microRNAs (miRNAs)**

miRNAs are short 17-25 nucleotides in length ncRNAs that implement their function through sequence-specific RNA interference¹⁹. miRNAs control gene activity either by binding and inhibiting the target mRNA from getting translated or through binding and degradation of mRNA *via* RNA induced silencing complex (RISC). miRNA functioning as tumor suppressor miRs or oncomiRs are found to be dysregulated in cancer. Differential expression pattern of miRNAs is associated with aggressive malignancy. Example: miR-210, miR-221, and miR-21 are significantly upregulated and miR-122a, miR-10b, miR-205, miR-145, are downregulated in TNBC²⁰. Among several other factors, miRNAs have major contribution in regulation of EMT, a process central to cancer metastasis. The miR-200 family targets ZEB1 and ZEB2 which are responsible for encoding E-cadherin repressors. miR-200 family miRNAs are reported to be downregulated in breast and lung cancer. Upregulation of miR-200 is reported to assist in expression of E-cadherin, thereby suppressing the metastatic motile property of lung cancer cells²¹. miR-200c is reported to be downregulated in breast tumor which results in upregulation of ZEB transcription factors leading to progression of EMT. let-7 family of miRNAs are downregulated in lung cancer. Upregulation of let-7 miRNAs downregulates KRAS and suppresses the proliferative rate of lung cancer²¹. Upregulation of miR-193a-3p is reported to directly target and inhibit KRAS in lung tumor. miR-181a-5p and miR-148a-3p are reported to inhibit cancer cell proliferation and migration in NSCLC²¹. Genome wide methylation profile in breast cancer have resulted in identification of several deregulated

miRNAs. miR-21, miR-34a, and miR-155 that are found to be hypomethylated and the tumor suppressor miRNAs, miR-9-2, miR-129, and miR-124-2, that are found to be hypermethylated in breast cancer²⁰. All these examples indicate towards the critical role of miRNAs in maintaining the balance of gene expressions, the alteration of which leads to cancer. The mentioned miRNAs are just a fraction of the literature surveyed miRNAs. The reference repository of miRNAs: the miRBase currently holds information on around 2656 mature miRNAs. Although false positive is an issue, the number of deposited miRNAs is increasing due to the high throughput sequencing of small RNAs¹⁹. Important miRNAs deregulated in metastatic cancer are listed in *Table 2*.

miRNAs	Expression level in cancer	Effect of induced upregulation/downregulation in cancer	Ref.
miR-21	Over-expressed in breast cancer, lung cancer and promotes EMT. Upregulated miR-21 post-transcriptionally downregulates PDCD-4 and TPM1, PTEN genes, thereby inducing invasion and metastasis. MDA-MB-231 cells expressing low levels of BMP-6 is correlated with increased expression of miR-21.	Inhibition of miR-21 decreases cancer cell proliferation, migration and invasion. Ectopic expression of miR-21 in MCF7 breast cancer cell line significantly reduced cell proliferation and anchorage independent cell growth.	22–24
miR-205	Downregulated in breast tumor	Upregulation of miR-205 is reported to suppress metastatic and invasive property of MDA-MB-231 cells.	22,25
miR-1	Downregulated in lung cancer	Induced expression of miR-1 in A549 cells is reported to tumor growth and migration in nude mice model. Expression of miR-1 reduced the expression of MET, Pim-1, FoxP1 and triggered apoptosis in A549 lung cancer cells.	22,26,27

miR-126	Downregulated in NSCLC	Activation of miR-126 has pro-apoptotic and antimetastatic effect in lung cancer. miR-16 acts by blocking VEGF-A/VEGFR-2/ERK signaling pathway.	28
miR-27b	Upregulated miR-27b promotes breast cancer growth, migration and invasion. It acts by downregulating Suppression of Tumorigenicity 14 (ST14)	Targeted inhibition of miR-27b/ST14 interaction may be an effective therapeutic approach in treating advance stage breast cancer.	22
miR-200 family		Ectopic expression of miR-200 inhibits expression of ZEB at post-transcriptional level leading to upregulation of epithelial markers and reduced motility of cancer cells	29
miR-10b	Elevated expression level in MDA-MB-231 cells and promotes cancer metastasis	Targeted downregulation of miR-10b may suppress metastasis.	22,30,31
miR-155	Elevated levels reported in breast cancer tissue.	Successful knockdown of miR-155 may lead to inhibition of cancer metastasis.	22
miR-146	Expression lost in metastatic cell-lines	Overexpression of miR-126 inhibits metastasis.	22
Let-7	Act as natural tumor suppressor in lung tissue. Low expression level of Let-7 is observed in lung cancer. Decreased level of Let-7a in breast cancer is associated with elevated level of RAS oncogene.	Successful upregulation of let-7 in breast and lung cancer may lead to significant suppression of metastasis.	32,33

Table 2: List of important miRNAs deregulated in metastatic

5.2 lncRNA-miRNA interaction in regulating gene expression

lncRNAs and miRNAs being the master-regulators of gene expression are crucial functional mediators of cancer progression. Intricate crosstalk between ncRNAs form complex regulatory network of post-transcriptional gene regulation. The main molecular mechanism by which

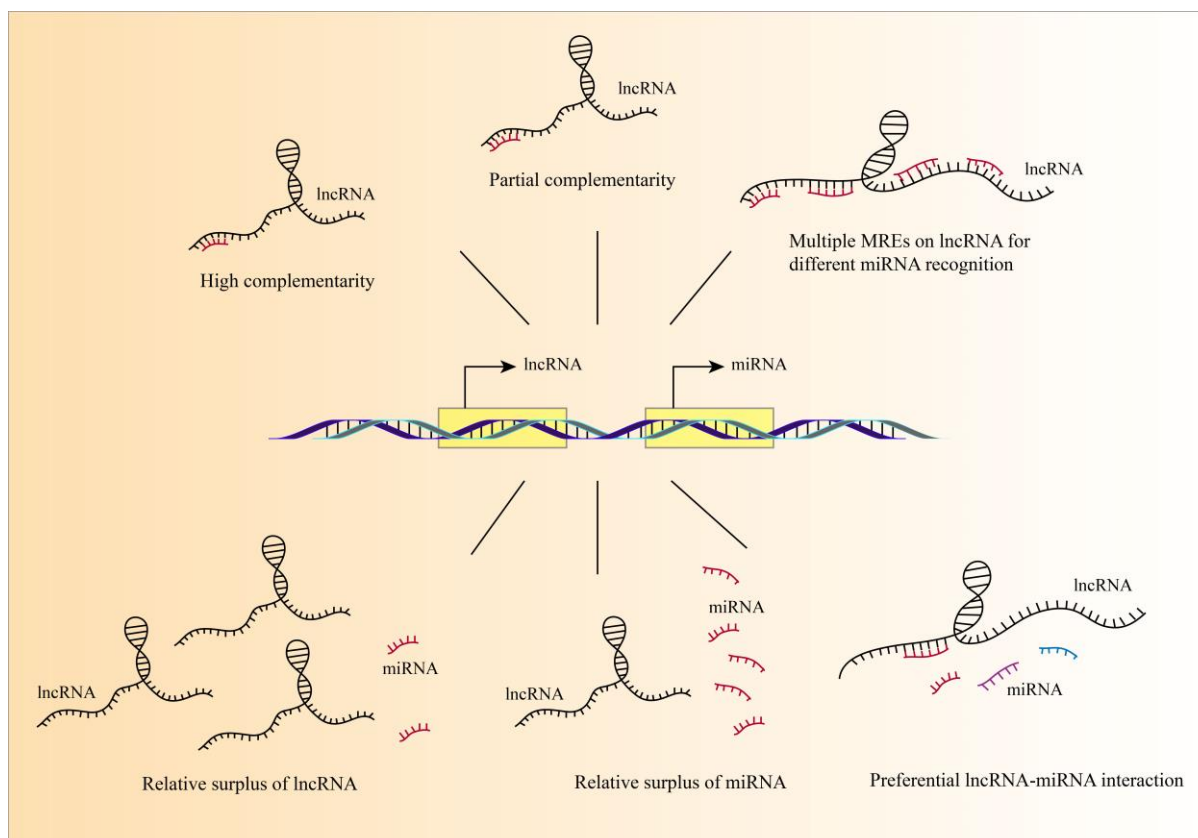


Figure 2. lncRNA-miRNA interaction. Regulatory lncRNAs have complementary site for miRNAs to bind. Efficiency of lncRNA-miRNA interaction depends on the strength of complementarity (high or partial complementarity). Higher the expression of lncRNA leads to more miRNA sequestration, thereby less available miRNAs for post-transcriptional target gene regulation and vice versa.

lncRNAs execute biological functions by acting as miRNA sponges [Figure 2]. lncRNA sequesters miRNAs, thereby affecting miRNA mediated RNA interference of genes. Depending on the specific lncRNA/miRNA interaction, lncRNA-miRNA axis may exert tumor suppressive effect or oncogenic effect²⁰. An example of a lncRNA-miRNA mediated gene regulation: lncRNA- NEAT2 (MALAT1) is identified to be abundantly expressed in metastatic breast cancer as well as NSCLC³⁴. MALAT1 is reported to promote EMT by sponging miR-1. miR-1 normally targets and degrades Slug and Vimentin restricting mesenchymal transformation of cells³⁵. Sponging of miR-1 by MALAT1 lncRNA results in upregulation of EMT markers. Knockdown of MALAT1 was correlated with miR-1 mediated inhibition of Slug and Vimentin and suppression of invasive property²⁰. Databases like starBase v2.0,

miR sponge and PeT Mbase allow selection of lncRNAs and miRNAs for studying the mechanism of ceRNA interaction³⁹. Bioinformatic tools like MechRNA, RNAHybrid, RNADuplex and RNAcofold, are developed with the aim to elucidate lncRNA-miRNA interactions⁴⁰. Important lncRNA-miRNA mediated gene regulation in cancer is enlisted in *Table 3*. Precision targeting and inhibition of lncRNA-miRNA interactions that contribute to the progression of cancer, along with the identification and characterization of novel oncogenic lncRNAs and miRNAs, may have a significant therapeutic impact with minimal off-target effects.

5.3 Potential anti-metastatic lncRNA-based therapeutics

Accumulating studies and evidence of the correlation between abnormal expression of lncRNAs and metastatic progression has impelled the researchers to develop potential lncRNA-based therapeutics. Antisense oligonucleotides (ASOs) are being used for targeted inhibition of lncRNAs. Antisense oligonucleotide (ASO1) targeted MALAT1 knockdown resulted in decelerated metastasis in breast and lung cancer xenografted mice models. Among the nucleic acid-based therapies trending now, locked nucleic acids (LNAs) are the most promising in the future development of cancer targeted drugs. Hu et al in 2019 reported the role of LINK-A LNAs in inhibition of cell proliferation in TNBC cells as well as sensitization of mammary gland tumors to immunotherapy treatment. In TNBC treatment with LINK-A LNAs, inhibition of cell proliferation was observed with reduced metastatic potential. This LNA treatment further sensitized tumors to immunotherapy by facilitating infiltration of cytotoxic CD8⁺/CD3⁺ T cells. Thus, a combinational therapy of LINK-A LNAs and immune check point blockers (ICBs) can be a promising treatment strategy for TNBC patients⁴. Inhibition of ncRNAs by RNA nanoparticle based targeted therapy is an emerging field in the development of novel vector system for cancer therapy. RNA nanoparticle mediated knockdown of onco-lncRNAs is a promising approach towards tumor specific treatment⁴.

lncRNA-miRNA-cancer crosstalk

lncRNA	State in cancer	miRNA	lncRNA mediated miRNA sponging	Effect on cancer progression	Reference
MALAT1 (NEAT2)	Elevated in breast cancer	miR-1, miR-1914-3p, miR-206	Upregulated MALAT1 sponges miR-1 that results in upregulation of SLUG and VIMENTIN. MALAT1 mediated sponging of miR-1914-3p upregulates YAP in NSCLC	Promotes EMT in breast cancer and NSCLC. Promotes chemoresistance.	20,34,36
Linc-ROR	Upregulated in breast and lung cancer	miR-205 miR-145	Upregulated linc-ROR captures miR-205, miR-145 resulting in upregulation of stemness factors and EMT mediators (ZEB1 and ZEB2).	Promotes cancer stemness and invasiveness in breast and lung cancer.	20,34
CCAT1	Upregulated in NSCLC	miR-218	Upregulated lncRNA-CCAT1 sponges miR-218 and promotes BMI1 expression.	Facilitates cancer metastasis	37
NEAT1	Upregulated in breast cancer	miR-448	Upregulated NEAT1 sponges miR-448, thereby upregulating ZEB1.	Activates EMT in cancer	20
HOTAIR	Upregulated in breast cancer and lung cancer	miR-34a	Upregulated HOTAIR sponges miR-34a and upregulates SOX2.	Promotes cancer stemness.	38

Abnormally expressed lncRNAs are characterized as biomarkers for diagnosis and prognosis of metastatic cancer. Deregulated lncRNAs in TNBC and NSCLC assist in cell division rate, EMT, metastasis and drug resistance. Till now almost all lncRNA related studies are focused on cell lines. More evidences are needed on animal models and clinical cancer patients for the development of lncRNA based therapies. ASOs, LNA and RNA nanoparticle-based targeting of lncRNAs are among the novel therapeutic approaches with the potential to reach anti-metastatic treatment in near future.

5.4 Future perspectives

Presently CTC detection remains limiting, however advancement in technology shall enable early detection of CTCs and tumor cell-free DNA (cfDNA) in the near future and could possibly determine the best course of action for treating metastatic relapses before they manifest. Incompetency of cancer targeted therapeutics has intensified the urgency for the design and development of anti-metastatic drugs⁴¹. lncRNA modulates gene expression through various pathways, setting an obstacle to unravel the mechanism of action that is essential for novel drug discovery. Current methods to target lncRNAs include double-stranded DNA-mediated interference (RNAi) and single-stranded antisense oligonucleotides (ASOs). The silencing of a lncRNA by an ASO would be sufficient to alter the expression of the downstream and associated genes since lncRNAs frequently overlap genes and regulate gene expression in cis. Targeted by ASOs, natural antisense RNAs (NATs) (also known as antago NATs) may upregulate overlapping tumor suppressor genes, suppressing the aggressiveness of cancer. But a more comprehensive approach is needed to translate these preclinical findings into clinical trials and successful medication development.

New therapeutic targets may be discovered through the detection and characterization of novel oncogenic lncRNAs and miRNAs as well as the particular lncRNA-miRNA interactions essential to the spread of cancer. Next generation sequencing (NGS) is a marvellous technology to sequence entire genomes as well as specific areas of interest to obtain population genomics and differential gene expression profiles that can accelerate the identification and characterization of crucial and yet-to-be-annotated functional lncRNAs in metastatic cancers⁴. Databases based on NGS technology such as Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) database have been used to construct ceRNA network that identified several lncRNAs differentially expressed in cancer, their correlation with protein coding genes and prognosis in cancer patients⁴. Precise targeting and inhibition of lncRNA-miRNA interactions

that contributes to cancer progression may result in great therapeutic impact with minimal off-target effects. High-throughput sequencing technology have made it possible for us to finally start comprehending the intricate control of gene expression and how it affects the development of cancer. LncRNAs' capacity to precisely regulate immune-related genes has therapeutic ramifications for the management of cancer. Further in-depth research is essential to completely comprehend their role as cancer biomarkers and as targets in cancer therapy.

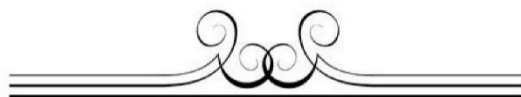
5.5 References

1. Ganesh, K. & Massagué, J. Targeting metastatic cancer. *Nat Med* **27**, 34–44 (2021).
2. Dutta, A., Roy, A. & Chatterjee, S. Long noncoding RNAs in cancer immunity: a new avenue in drug discovery. *Drug Discov Today* **26**, 264–272 (2021).
3. Lemjabbar-Alaoui, H., Hassan, O., Yang, Y.-W. & Buchanan, P. Lung cancer: biology and treatment options. *Biochim Biophys Acta* **1856**, 189–210 (2015).
4. Zhang, W., Guan, X. & Tang, J. The long non-coding RNA landscape in triple-negative breast cancer. *Cell Prolif* **54**, e12966 (2021).
5. Moran, V. A., Perera, R. J. & Khalil, A. M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res* **40**, 6391–6400 (2012).
6. Zhou, C. *et al.* Long noncoding RNA HOTAIR, a hypoxia-inducible factor-1 α activated driver of malignancy, enhances hypoxic cancer cell proliferation, migration, and invasion in non-small cell lung cancer. *Tumor Biol.* **36**, 9179–9188 (2015).
7. Jin, C., Yan, B., Lu, Q., Lin, Y. & Ma, L. The role of MALAT1/miR-1/sluc axis on radioresistance in nasopharyngeal carcinoma. *Tumor Biol.* **37**, 4025–4033 (2016).
8. Liu, J. *et al.* The Long Noncoding RNA MEG3 Contributes to Cisplatin Resistance of Human Lung Adenocarcinoma. *PLOS ONE* **10**, e0114586 (2015).
9. Zhu, G. *et al.* Long noncoding RNA HAS2-AS1 mediates hypoxia-induced invasiveness of oral squamous cell carcinoma. *Molecular Carcinogenesis* **56**, 2210–2222 (2017).
10. Lu, Y. *et al.* LncRNA HAS2-AS1 Promotes Glioblastoma Proliferation by Sponging miR-137. *Front Oncol* **11**, 634893 (2021).

11. Eroğlu Güneş, C., Güçlü, E., Vural, H. & Kurar, E. Knockdown of lncRNA ZEB2NAT suppresses epithelial mesenchymal transition, metastasis and proliferation in breast cancer cells. *Gene* **805**, 145904 (2021).
12. Wu, Y. *et al.* Downregulation of the long noncoding RNA GAS5-AS1 contributes to tumor metastasis in non-small cell lung cancer. *Sci Rep* **6**, 31093 (2016).
13. Liu, B. *et al.* A Cytoplasmic NF- κ B Interacting Long Noncoding RNA Blocks I κ B Phosphorylation and Suppresses Breast Cancer Metastasis. *Cancer Cell* **27**, 370–381 (2015).
14. Jiang, M.-C., Ni, J.-J., Cui, W.-Y., Wang, B.-Y. & Zhuo, W. Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am J Cancer Res* **9**, 1354–1366 (2019).
15. Arun, G. *et al.* Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev* **30**, 34–51 (2016).
16. Matsui, M. & Corey, D. R. Perspectives. *Nat Rev Drug Discov* **16**, 167–179 (2017).
17. Grelet, S. *et al.* A regulated PNUTS mRNA to lncRNA splice switch mediates EMT and tumour progression. *Nat Cell Biol* **19**, 1105–1115 (2017).
18. Gan, F.-J. *et al.* LncRNA BCAR4 expression predicts the clinical response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *Cancer Biomark* **32**, 339–351.
19. Alles, J. *et al.* An estimate of the total number of true human miRNAs. *Nucleic acids research* **47**, (2019).
20. Venkatesh, J., Wasson, M.-C. D., Brown, J. M., Fernando, W. & Marcato, P. LncRNA-miRNA axes in breast cancer: Novel points of interaction for strategic attack. *Cancer Letters* **509**, 81–88 (2021).
21. Wu, K.-L., Tsai, Y.-M., Lien, C.-T., Kuo, P.-L. & Hung, J.-Y. The Roles of MicroRNA in Lung Cancer. *Int J Mol Sci* **20**, 1611 (2019).
22. Baranwal, S. & Alahari, S. K. miRNA control of tumor cell invasion and metastasis. *Int J Cancer* **126**, 1283–1290 (2010).
23. Arisan, E. D. *et al.* MiR-21 Is Required for the Epithelial-Mesenchymal Transition in MDA-MB-231 Breast Cancer Cells. *Int J Mol Sci* **22**, 1557 (2021).

24. Cao, L. *et al.* Exosomal miR-21 regulates the TETs/PTENp1/PTEN pathway to promote hepatocellular carcinoma growth. *Molecular Cancer* **18**, 148 (2019).
25. Wang, L., Kang, F., Wang, J., Yang, C. & He, D. Downregulation of miR-205 contributes to epithelial–mesenchymal transition and invasion in triple-negative breast cancer by targeting HMGB1–RAGE signaling pathway. *Anticancer Drugs* **30**, 225–232 (2019).
26. Li, T., Wang, X., Jing, L. & Li, Y. MiR-1-3p Inhibits Lung Adenocarcinoma Cell Tumorigenesis via Targeting Protein Regulator of Cytokinesis 1. *Frontiers in Oncology* **9**, (2019).
27. Tominaga, E., Yuasa, K., Shimazaki, S. & Hijikata, T. MicroRNA-1 targets Slug and endows lung cancer A549 cells with epithelial and anti-tumorigenic properties. *Exp Cell Res* **319**, 77–88 (2013).
28. Chen, Q., Chen, S., Zhao, J., Zhou, Y. & Xu, L. MicroRNA-126: A new and promising player in lung cancer. *Oncol Lett* **21**, 35 (2021).
29. Brabletz, S. & Brabletz, T. The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer? *EMBO Rep* **11**, 670–677 (2010).
30. Sheedy, P. & Medarova, Z. The fundamental role of miR-10b in metastatic cancer. *Am J Cancer Res* **8**, 1674–1688 (2018).
31. Ma, L. Role of miR-10b in breast cancer metastasis. *Breast Cancer Research* **12**, 210 (2010).
32. Thammaiah, C. K. & Jayaram, S. Role of let-7 family microRNA in breast cancer. *Noncoding RNA Res* **1**, 77–82 (2016).
33. Shen, C., Li, J. & Che, G. Prognostic value of let-7 in lung cancer: systematic review and meta-analysis. *Translational Cancer Research* **9**, (2020).
34. Pan, Y. *et al.* The Emerging Roles of Long Noncoding RNA ROR (lincRNA-ROR) and its Possible Mechanisms in Human Cancers. *CPB* **40**, 219–229 (2016).
35. Safa, A. *et al.* miR-1: A comprehensive review of its role in normal development and diverse disorders. *Biomedicine & Pharmacotherapy* **132**, 110903 (2020).
36. Zhou, Q. *et al.* Novel Insights Into MALAT1 Function as a MicroRNA Sponge in NSCLC. *Frontiers in Oncology* **11**, (2021).

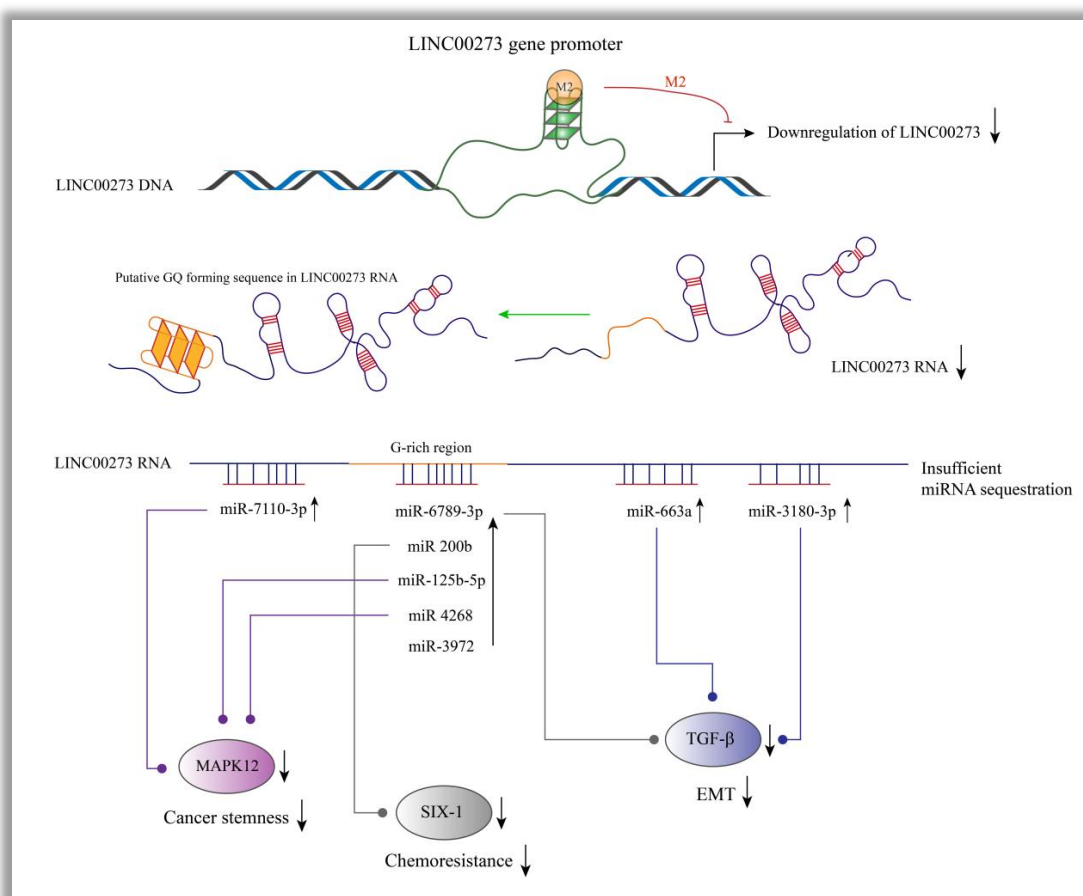
37. Hu, Q., Ma, H., Chen, H., Zhang, Z. & Xue, Q. LncRNA in tumorigenesis of non-small-cell lung cancer: From bench to bedside. *Cell Death Discov.* **8**, 1–9 (2022).
38. Deng, S., Wang, J., Zhang, L., Li, J. & Jin, Y. LncRNA HOTAIR Promotes Cancer Stem-Like Cells Properties by Sponging miR-34a to Activate the JAK2/STAT3 Pathway in Pancreatic Ductal Adenocarcinoma. *Onco Targets Ther* **14**, 1883–1893 (2021).
39. Zhang, X. *et al.* Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *Int J Mol Sci* **20**, 5573 (2019).
40. López-Urrutia, E., Bustamante Montes, L. P., Ladrón de Guevara Cervantes, D., Pérez-Plasencia, C. & Campos-Parra, A. D. Crosstalk Between Long Non-coding RNAs, Micro-RNAs and mRNAs: Deciphering Molecular Mechanisms of Master Regulators in Cancer. *Front Oncol* **9**, 669 (2019).
41. Jana, S. *et al.* LINCRNA00273 promotes cancer metastasis and its G-Quadruplex promoter can serve as a novel target to inhibit cancer invasiveness. *Oncotarget* **8**, 110234–110256 (2017).



Chapter 6

M2 mediated LINC00273 downregulation modulates lncRNA-miRNA interaction to suppress cancer metastasis

This chapter is based on “Long non-coding intergenic RNA, LINC00273 induces cancer metastasis and stemness via miRNA sponging in Triple Negative Breast Cancer” Communicated to NAR cancer 2023



6.1 Introduction

A number of biomolecular mechanisms alliance to develop cancer metastasis. Non-coding RNAs (ncRNAs) are increasingly being recognized to be associated with deregulated genes in cancer^{1,2}. The two most extensively researched ncRNAs are lncRNAs and miRNAs. Despite the expanding database of ncRNAs, the specific interaction between lncRNAs and miRNAs in

metastatic cancer remains largely unknown³. LINC00273 is a recently identified long non-coding RNA associated with various tumors. A survey of different metastatic clinical tumor samples unveiled the upregulated state of LINC00273⁴. Studies have revealed transforming growth factor beta (TGF β) mediated LINC00273 expression in lung cancer⁵. Upregulated LINC00273 in turn was observed to activate ZEB1, thereby promoting EMT in cancer. LINC00273 is reported to be involved in cancer metastasis. Knockdown of LINC00273 in AGS gastric cancer cell line suppressed cancer cell migration and invasion⁴. The discovery of the involvement of lncRNA in facilitating Wnt signalling by stabilization and activation of β -catenin confirmed the plausible role of LINC00273 in cancer stemness⁶. However, the underpinning mechanism of how *LINC00273* triggers metastatic potential and acquisition of stemness in cancer cells remains poorly understood. We have tried to evaluate the molecular mechanism of action of LINC00273 in regulating metastasis in malignant cells in our study under LINC00273 upregulated and downregulated conditions. A unique G-Quadruplex (G4) scaffold having (4n-1) topology had been mapped in the promoter region of *LINC00273* gene, that inversely regulates its transcription resulting in the reversion of the metastatic potential of cancer cells⁴. M2 is a salt molecule of a bis Schiff base that is a patent anti-metastatic drug agent identified to selectively bind to the G-quadruplex (GQ) in the promoter region of LINC00273 gene [US Patent granted (No. 9, 682, 926 B2), Australian Patent granted (AU2018201532), European Patent granted (EP2900234B1)]. M2 mediated stabilization of the GQ regulatory switch results in the downregulation of LINC00273⁴. M2 treatment in metastatic B16F₁₀ melanoma cell line and AGS gastric cancer cell line is reported to be associated with suppressed EMT mediated cell migration and metastasis⁴. lncRNAs often act as competitive endogenous RNAs (ceRNA) to sequester miRNAs that regulate protein coding genes at translational level, thereby controlling the expression of crucial genes involved in cell proliferation, invasion, metastasis and apoptosis⁷. Most lncRNAs possess miRNA response

elements (MRE) at their 3' end complementary to specific miRNAs, this complementary binding leads to sequestration of miRNAs by lncRNAs¹. Depending on the specific lncRNA/miRNA interaction, lncRNA-miRNA axis may exert tumor suppressive or oncogenic effect⁸. The identification and characterization of novel oncogenic lncRNAs and miRNAs and precise targeting and inhibition of lncRNA-miRNA interactions that contribute to cancer progression may result in great therapeutic impact with minimal off-target effects.

In our study, we have reported the upregulated state of LINC00273 in two highly metastatic cell lines [MDA-MB-231 TNBC breast cancer cell line and A549 lung cancer cell line] and studied the functional role of over-expressed LINC00273 in promoting tumorsphere formation and cell migration. We have reported the pro-metastatic function of LINC00273 and the antimetastatic effect of M2 drug treatment in the cancer cells. The transcriptomic analysis was done under LINC00273 overexpressed condition and M2 treated condition in the metastatic cell line. We observed a differential pattern of global gene expression in LINC00273 over-expressed condition that correlates with the control metastatic group with upregulated LINC00273 compared to M2 treated condition. The stemness marker genes SOX2, OCT4, and NANOG were observed to be downregulated upon M2 treatment in MDA-MB-231 and A549 metastatic cancer cell lines. Similarly, EMT and metastatic markers ZEB1, LINC00273 and NME2 were found to be overexpressed in the metastatic condition and downregulated upon M2 treatment. Chemoresistance markers TGF β and SIX1 were upregulated in metastatic condition and downregulated in the M2 treated condition. We confirmed the pro-metastatic function of LINC00273 and the antimetastatic effect of M2 drug in TNBC nude mice model as well. Similar pattern of downregulation of stemness, metastatic and chemoresistance marker genes were observed with no trace of metastasis to distant organs. We further studied the sequence of LINC00273 at the nucleotide level. The following investigation is additional or supplementary work which I have done to extensively study the mechanism of how

LINC00273 transcript functions in controlling multiple gene expression. Through bioinformatic analysis, we investigated the base-pairing between *LINC00273* and miRNAs and mapped the miRNA interacting sites in the sequence of *LINC00273*. We checked the expression of the bioinformatically suggested miRNAs in *LINC00273* upregulated condition and M2 treated condition. G4 structure formation in *LINC00273* gene promoter is supposed to negatively regulate *LINC00273* transcription resulting in only basal level of *LINC00273* transcripts, thereby allowing interference of oncogenes by miRNAs. The G-rich sequence in the *LINC00273* RNA plays an important role in the sequestration of specific miRNAs having complementary binding site within the G-rich region. We have reported TGF- β 1, SIX1 and MAPK12 to be the direct downstream targets of *LINC00273*. M2 mediated downregulation of *LINC00273* results in inefficient and low-level of miRNA sponging activity that subsequently allows the miRNAs to bind to the 3'-UTR of their target oncogenes, resulting in downregulation of metastasis, stemness and chemoresistance marker genes.

6.2 Results and discussion

6.2.1 M2 downregulates *LINC00273* in metastatic cell line MDA-MB-231 and A549

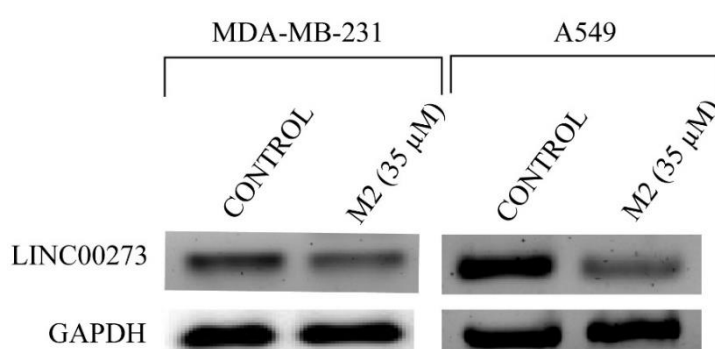


Figure 1. RNA expression of LINC00273 in MDA-MB-231 and A549 cancer cell lines. LINC00273 is upregulated in both of these highly metastatic cell lines. M2 treatment results in downregulation of LINC00273 compared to positive control GAPDH.

Initially, we checked the expression level of *LINC00273* in two highly metastatic cell lines: MDA-MB-231 triple negative breast cancer (TNBC) cell line and A549 highly metastatic non-

small cell lung cancer (NSCLC) cell line. We observed LINC00273 in an upregulated condition in both of these metastatic cell lines, while upon M2 treatment (35 μ M), LINC00273 got specifically downregulated [Figure 1]. Thus, we speculated the downregulation of LINC00273 via M2 mediated stabilization of the GQ in the promoter of LINC00273 gene.

6.2.2 M2 mediated downregulation of LINC00273 suppresses cancer cell migration and tumor formation capacity of highly metastatic cells

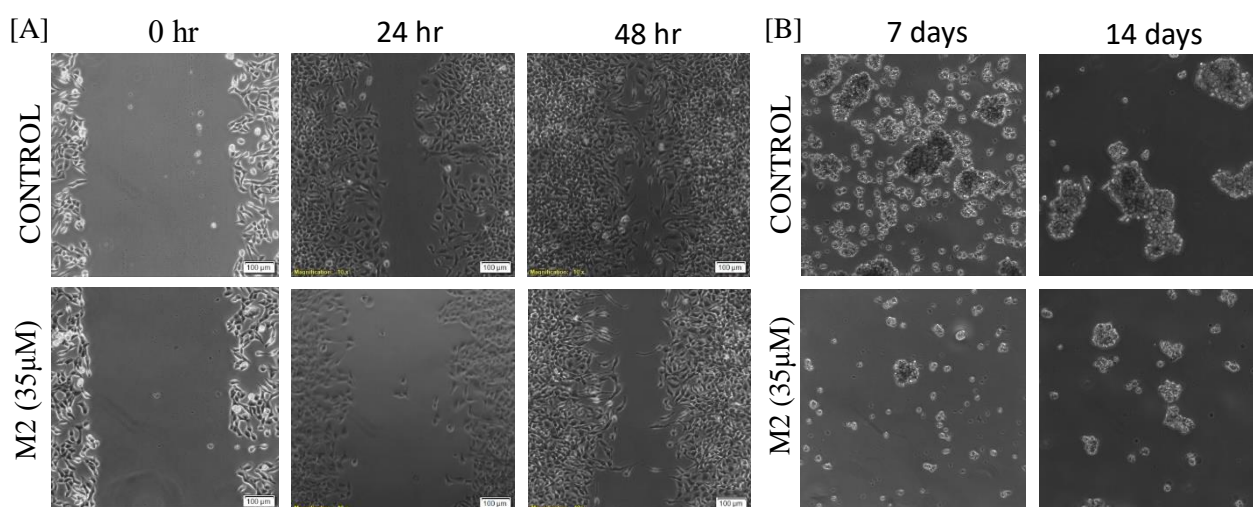


Figure 2. [A] Scratch wound assay. The scratch made at 0 hours was observed to be rapidly occupied by migrating A549 cells in the control group. Images captured at 24 hours and 48 hours. M2 treatment slowed down the migration potential of A549 cells as observed by the delayed wound healing at 24- and 48-hours image captured. **[B] Tumor sphere formation assay.** Tumor sphere formation capability of A549 cells were observed to be suppressed upon treatment with M2 compared to the control set. Images were captured on the 7th day and 14th day of tumor formation assay.

To investigate the role of M2 drug on the migration property of the metastatic cell line, we have performed scratch wound assay. We observed higher migratory potential of the metastatic cells in the control group with upregulated LINC00273 as evidenced by the rapid wound healing compared to M2 treated LINC00273 downregulated group [Figure 2A]. M2 treatment slowed down the migration potential of the metastatic A549 and MDA-MB-231 cancer cells. The scratch wound assay of MDA-MB-231 upon M2 treatment is included in the thesis of Pallabi Sengupta, the joint first author of this manuscript. From the wound healing assay, we

speculated a positive role of LINC00273 in promoting cancer metastasis with active LINC00273 expression.

Tumor formation was observed to be suppressed in MDA-MB-231 and A549 cell line upon treatment with M2 drug. The size of the tumors as well as the number of spheres were significantly less in the M2 treated set compared to the control set confirming the metastatic suppressive property of M2 drug in A549 cells [Figure 2B]. Tumor sphere formation assay confirming the suppressive action of M2 in MDA-MB-231 cell line is included in Mrs. Pallabi Sengupta's thesis. M2 assists in inhibiting the acquisition of stemness characteristic by the cancer cells.

6.2.3 A comparative analysis of the oncogenes at the transcript level in MDA-MB-231 and A549 metastatic cell lines upon M2 treatment

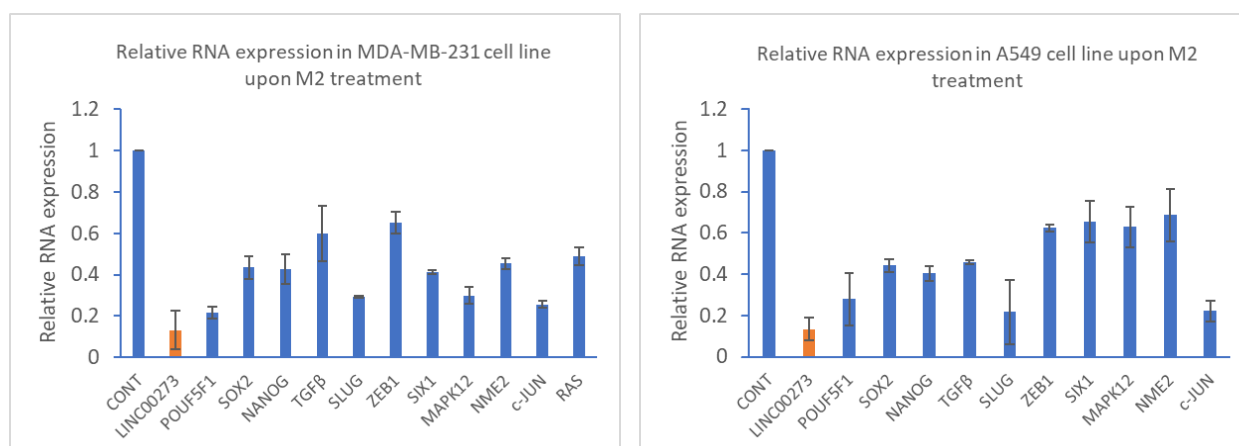


Figure 3. Relative RNA expression (fold change) of oncogenes involved in cancer cell proliferation, EMT, metastasis and chemoresistance upon M2 treatment in MDA-MB-231 cell line and A549 cell line. GAPDH was considered as the positive control for normalizing the data. LINC00273 is highlighted in orange. Error bars represent mean \pm SE (N = 3).

We checked the RNA expression level of selected genes involved in cancer stemness, EMT, metastasis and chemoresistance in MDA-MB-231 cell line and A549 cell line upon M2 treatment. Stemness promoting signature genes SOX2, OCT4 (POU5F1) and NANOG were observed to be significantly downregulated in MDA-MB-231 and A549 metastatic cell lines

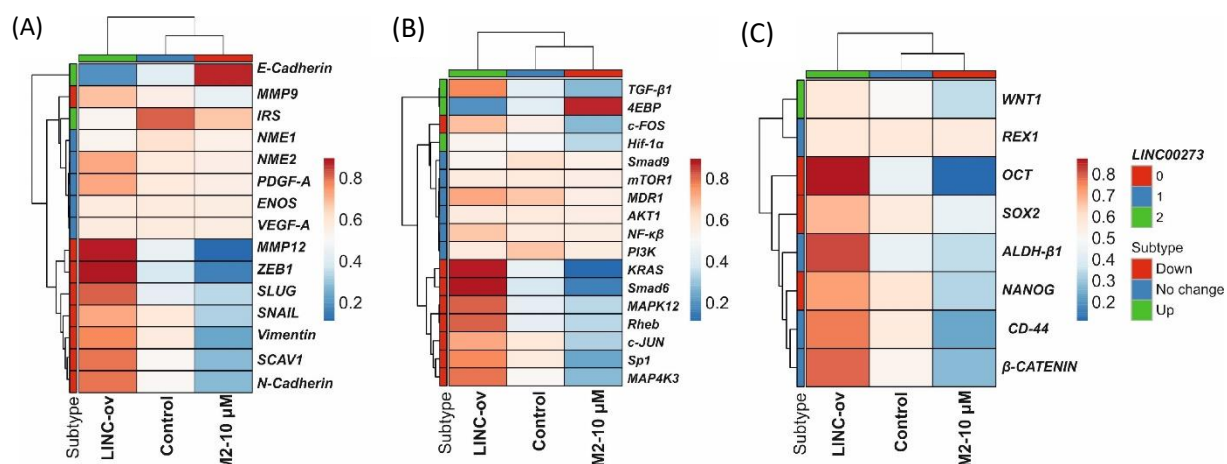


Figure 4. *LINC00273* regulates the expression of metastatic, stemness, and chemoresistance markers in MDA-MB-231 tumorspheres. Next-generation sequencing analyses of *LINC00273* overexpression and M2 treatment in MDA-MB-231 spheres. Differentially expressed genes are clustered based on the magnitude of *LINC00273* expression **[A]** RNAseq data for differentially expressed genes in metastatic regulation upon *LINC00273* overexpression and M2 treatment, compared to that of control. (*LINC00273*, *VIMENTIN*, *SNAIL*, *SLUG*, *NME1*, *NME2*, *MMP9*, *MMP12*, *ECADHERIN*, *N-CADHERIN*, *ZEB-1*, *VEGF-A*, *PDGF-A*). *GAPDH* used as housekeeping gene. **[B]** Heatmap data of biological replicates of differentially expressed genes associated with uncontrolled proliferation (*KRAS*, *c-FOS*, *c-JUN*, *Sp1*, *MAPK12*, *MAP4K3*), and chemoresistance (*NF- κ B*, *TGF- β* , *Smad9*, *Smad6*, *SIX-1*, *HIF-1 α*) and using the $\log_{10}(\text{count})$ values for each replicate upon *LINC00273* overexpression and M2 treatment, compared to that of control. **[C]** Heatmap data of biological replicates of differentially expressed genes associated with cancer stemness (*OCT-4*, *SOX2*, *Nanog*, *WNT-1*, β -Catenin, *ALDH β 1*, *CD-44*, *REX-1*) using the $\log_{10}(\text{count})$ values for each replicate. The screening threshold for the differentially expressed genes is set to: $|\log_2(\text{Fold Change})| > 1$ and $p\text{-Value} < 0.05$. Differential expression of the transcripts clustered by Euclidean correlation, with distinct upregulation, unaltered, and downregulation patterns in expression for *LINC00273* overexpression and M2 treatment, compared to that of the control set. The count values are colour-coded blue to yellow to red in increasing order. Differentially expressed genes are clustered based on the magnitude of *LINC00273* expression.

[*Figure 3*]. EMT enhancers ZEB1, SLUG and LINC00273 were also observed to be significantly downregulated upon M2 treatment in both the metastatic cell lines. c-JUN and MAPK12 genes responsible for uncontrolled cancer cell proliferation were observed to be downregulated upon M2 treatment [*Figure 3*]. SIX1 and TGF β are responsible for chemoresistance in malignant cancer cells. These chemoresistance markers were observed to be moderately downregulated upon M2 treatment. Thus, we confirm the antimetastatic property of M2 drug in MDA-MB-231 and A549 highly metastatic cell lines.

6.2.4 Differential transcriptome scenario of MDA-MB-231 tumorsphere in presence of LINC00273 and M2 treatment

To analyze the global transcriptomic scenario of MDA-MB-231 tumorspheres under metastatic condition with upregulated LINC00273, LINC00273 overexpressed (ov) and M2 treated condition, we have performed next generation sequencing (NGS). Principal Component Analysis (PCA) of RNA-sequencing data demonstrated that the global transcriptomic scenario of metastatic control group with upregulated LINC00273 and LINC00273(ov) induced group could be positively correlated. This suggests the possible major role of LINC00273 in affecting the global transcriptome in MDA-MB-231 tumorspheres. M2 mediated downregulation of LINC00273 remarkably reduced the metastatic characteristics of TNBC MDA-MB-231 cell line [[GEO accession number: GSE217692](#)]. Gene set enrichment analysis (GSEA) suggested 24 genes having distinct and overlapping combinatorial patterns, involved in the previously mentioned different oncogenic processes. Upregulation of oncogenes N-Cadherin, Snail, Slug, Vimentin, and MMP12 were found in *LINC00273* over-expressed condition that contributes to EMT induced metastasis and acquisition of invasion property [*Figure 4A*]. Upregulation of KRAS, Sp1, c-FOS, c-JUN, MAPK12, and MAP4K3 were also observed in *LINC00273* over-expressed conditions that contribute to increased cancer cell proliferation [*Figure 4B*]. Upregulation of OCT-4, SOX2, NANOG and CD44 were also observed upon overexpression

of LINC00273 that contributes to cancer stem cell (CSC) expansion in MDA-MB-231 tumorspheres [Figure 4C]. Further an increase of TGF- β 1, Smad9, SIX-1, and HIF-1 α contributed to the acquisition of chemoresistance by cancer cells in LINC00273 overexpressed cells [Figure 4B]. All these genes were downregulated upon M2 treatment as obtained from the NGS data [Figure 4]. Thus, we confirm that upregulated LINC00273 is tightly associated with higher plasticity level, higher proliferative and metastatic potential with increased resistance towards chemotherapy drugs in MDA-MB-231 TNBC cell line.

6.2.5 Antimetastatic property of M2 validated in TNBC xenograft nude mice model

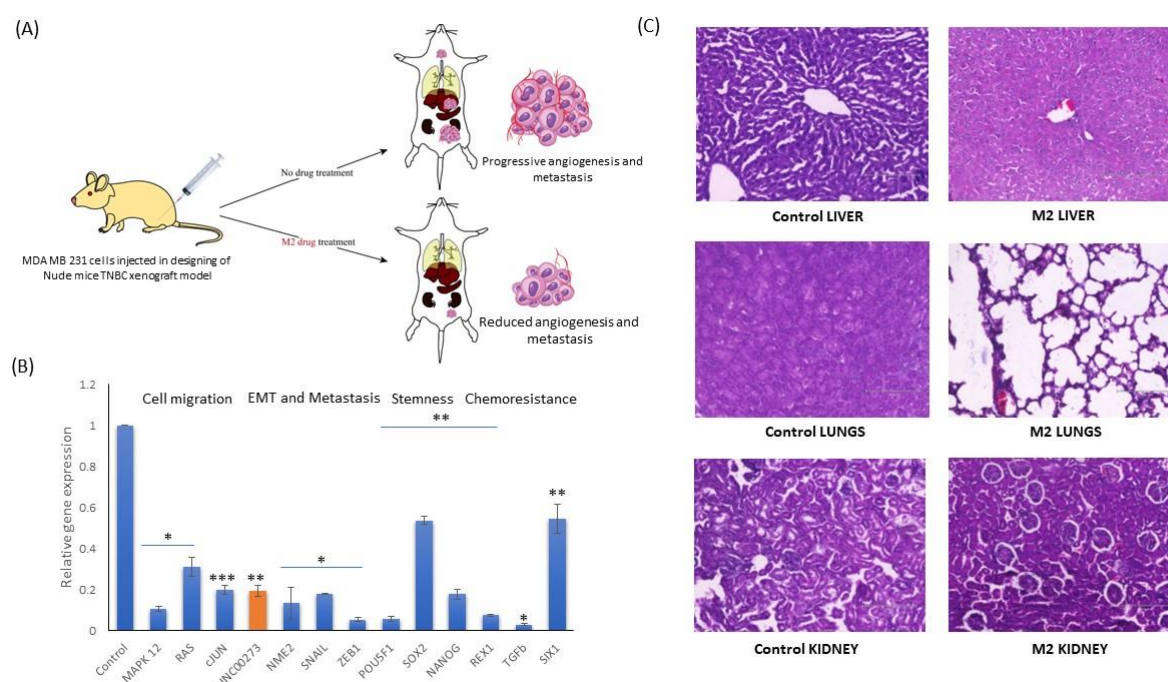


Figure 5. M2 treatment downregulates an array of oncogenes in TNBC xenograft nude mice model. [A] Schematic diagram of M2 treated TNBC xenograft tumor mice condition compared to control. [B] RT-PCR data showing relative expression level of genes involved in EMT, cancer stemness, metastasis and chemoresistance. Error bars represent mean \pm SE (N = 3). Statistical significance determined by Student's t-test. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (C) Histological tissue sections of liver, lungs and kidney of M2 treated MDA-MB-231 TNBC xenografted nude mice compared to control mice.

We have developed TNBC nude mice model and treated the mice with 35 μ M (12.76 mg/kg body weight) of M2 drug. The drug concentration was standardized in our previous study⁴. After completion of the treatment, the mice were sacrificed and the organs (liver, kidney, lung) along with the tumors were isolated for further processing. We checked the expression of EMT, stemness and chemoresistance marker genes from the isolated breast tumors. Significant downregulation of important genes' expression was observed in the M2 treated condition compared to the control [Figure 5B]. Upon histopathological analysis we observed negligible metastasis to distant organs in M2 treated mice models compared to control [Figure 5A and 5C]. Animal model data firmly validated the significance of M2 mediated *LINC00273* downregulation in constraining cancer metastasis.

However, it is difficult to comment whether *LINC00273* is directly involved in regulating the expression of these oncogenes as these genes are involved in intricate signalling cascades that crosstalk with other oncogenes enmeshed in various signalling pathways to culminate an integrated effect on cancer cell characteristics. However, our results support the underlying role of *LINC00273* as 'miRNA sponge' that modulate the expression profile of these oncogenes. *LINC00273* might have a few specific and downstream targets which modulates the expression profile of other oncogenes *via* overlapping signalling cascades.

6.2.6 Downstream targets and the plausible mechanism of *LINC00273* mediated oncogenic regulation

We have performed bioinformatic studies using miRDB database and Freiburg RNA tool to predict the base-pairing between *LINC00273* and miRNAs and map the miRNA interacting sites in the sequence of *LINC00273* [Table 1]. We observed that the target binding sites of these miRNAs were located in the 3'-UTR of TGF- β 1, MAPK12, and SIX-1 that base pair with the miRNAs. Pull down assay has also been performed that confirmed the physical interaction between the lncRNA and miRNAs. The pulldown data is included in Ms. Pallabi Sengupta's

thesis, the joined first author of this paper. We observed the seeding sequence of miR125b-5p and miR-4268 miRNAs to have complementary binding sites within the putative GQ motif of *LINC00273* RNA and also 3'-UTR of MAPK12 mRNA. The miRNA miR-7110-3p was found

<i>LINC00273</i>	MicroRNA	miRNA: <i>LINC00273</i>	<i>MAPK12</i> (3'-UTR)	miRNA: <i>MAPK12</i> (3'-UTR)
	miR-125b-5p	5' - ATACGC GGGCGGGGAGGGGGG C-3' 3' - AGUGUCAAUCCAGAGUCCCU -5'		5' - UCCCUGAGACCCUAACUUGUGAU -3' 3' - GAGGGUUCUCGGAUUGGGCAC -C-5'
	miR-4268	5' - CGC GGGCGGGGAGGGGGG CAC-3' 3' - GUGUAGGACUCUCCUCUCG -5'		5' - CAGGUUCU-AGAGGUUUUGCC-3' 3' - GUGUAGGACUCUCCUCUCG -5'
	miR-7110-5p	5' - CCCGCGGCTCCCCACCGCCT-3' 3' - GAGAGAGAGGGGUGUGGGGU -5'		5' - UGGGGGUGUGGG-GAGAGAGAG -3' : : : 3' - UCCCUACCCUCUCUCUCUC -5'

<i>LINC00273</i>	MicroRNA	miRNA: <i>LINC00273</i>	<i>SIX-1</i> (3'-UTR)	miRNA: <i>SIX-1</i> (3'-UTR)
	miR-3972	5' - GGGCGGGGAGGGGGGCA CAGAC-3' 3' - ACGGGACCUUGCCCCGACCGUC -5'		5' - UGCCUUGGCUUGGGG-UGGCAG-3' 3' - ACGGGACCUUGCCCCGACCGUC -5'
	miR-200b	5' - GGGGCA CAGACAGGCAAGGC-3' 3' - AGUAGUAAUGGUCCGUCAUA -5'		5' - CAUC-UUACUGGGCAGCAU-3' 3' - GUAGUAAUGGUCCGUCAUA -5'
	miR-4306	5' - TGCTGGTTTCTTGTCCA-3' 3' - AUGACGGAAAGAGAGGU -5'		5' - UAGUGUCCUAGAGUCUCCA-3' 3' - AUGAC-GGAA--AGAGAGGU -5'

<i>LINC00273</i>	MicroRNA	miRNA: <i>LINC00273</i>	<i>TGF-β1</i> (3'-UTR)	miRNA: <i>TGF-β1</i> (3'-UTR)
	miR-663a	5' - TCCCGCCATCGGGACCCGAAG-3' 3' - CGCCAGGGCGCCGCGGGGCGGA -5'		5' - AGG-CGGGGCG-CCGCGGGA--CCGC -3' 3' - UCCUUUGUGCCGGCGCCUAGGGUG-5'
	miR-6789-3p	5' - CGC GGGCGGGGAGGGGGGCA C-3' 3' - GACCUCCUCUGUGCCCGCGGC -5'		5' - GAGGUAGGGGCGUCCGGGCGCGCG-3' 3' - GACCUCCUCUGU---GCCCGCG-GC -5'
	miR-3180-3p	5' - CCCGCGGCGCCCCGCCTAG-3' 3' - GAGGCCUUCGAGGCGGGGU -5'		5' - GGGGCGGAGCUCCGGAG -3' 3' - CCCUGAC-CGUAGGCCUC-5'

Table 1. List of miRNAs having binding sites in *LINC00273* and the 3'-UTR of specific genes.

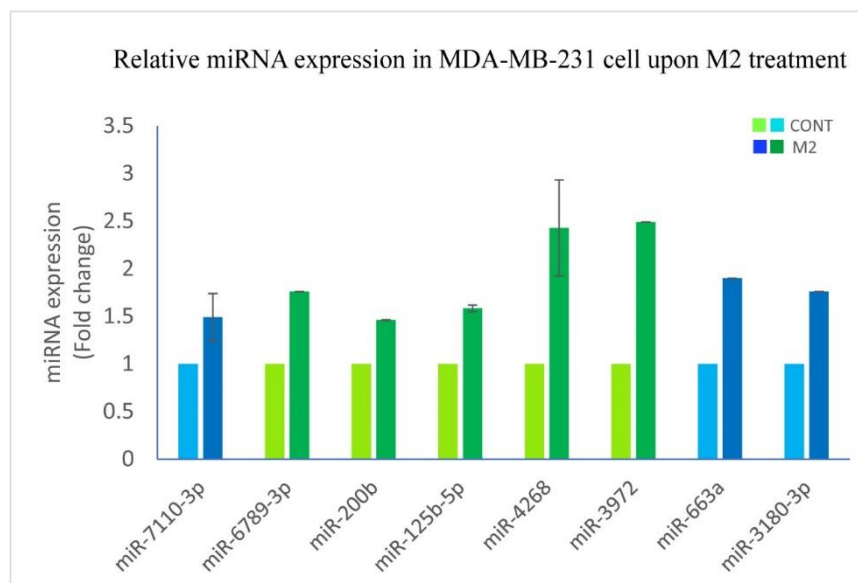


Figure 6. miRNA expression level in M2 treated MDA-MB-231 group with downregulated *LINC00273* compared to control. Each value of miRNA expression has been normalized to that of control. The bars colored green represents the miRNAs having complementary binding sites within the G rich sequence of *LINC00273* RNA. The bars colored blue represents the miRNAs with complementary binding sites upstream or downstream of *LINC00273* G rich sequence. Error bars represent mean \pm SE (N = 3).

to interact with *LINC00273* in the upstream of GQ sequence. Similarly, miR-3972 and miR-200b having target binding sites at the 3'-UTR of *SIX-1* were found to physically interact with the complementary binding sites at the putative GQ sequence of *LINC00273* RNA. The miR-4306 showed complementary binding sites at the downstream of G4 motif of *LINC00273* as well as in the 3'-UTR region of *SIX1*. miR-6789-3p is observed to bind complementarily to the G4 motif of *LINC00273* as well as at the 3'-UTR of *TGF- β 1* mRNA. The miR-663a and miR-3180-3p were found to interact with *LINC00273* downstream to G4 motif and with the 3'-UTR of *TGF- β 1* mRNA [Table 1]. Thus, our study suggests that *LINC00273* has complementary binding sites for many miRNAs that are directly involved in regulating oncogenes, especially MAPK12, *SIX1*, and *TGF- β 1*.

We checked the expression of the bioinformatically hunted miRNAs in MDA-MB-231 cell line by performing stem-loop PCR. We observed the upregulation of the mentioned miRNAs under

M2 treated condition compared to the control confirming the miRNA sponging role of LINC00273 RNA. Under overexpressed LINC00273 condition, the said miRNAs were lowly expressed or scanty due to their sequestration by LINC00273 RNA and thus, they were unavailable for downstream activity. Under M2 treated condition, M2 targeted LINC00273 gene promoter and downregulated its RNA expression, the low-level of LINC00273 RNA was inadequate to efficiently sequester the miRNAs and thus, we observed higher levels of miRNAs in M2 treated condition [*Figure 6 miRNA expressions in MDA-MB-231 bar plot*]. The available miRNAs further bind to their target oncogenes and downregulate them, thereby providing a plausible explanation for the downregulation of several oncogenes upon M2 treatment.

6.3 Conclusion

The large-scale discoveries suggest the promising role of long non-coding RNAs as diagnostic and prognostic biomarkers in cancer, fuelling the possibilities to develop cancer-therapeutic and preventive strategies by targeting the long non-coding RNAs. In this study, the underlying mechanism of *LINC00273*-induced metastasis, cancer stemness and chemoresistance had been investigated. Upregulated *LINC00273* was found to be associated with higher migratory potential of the metastatic cells and greater tumorsphere formation ability. RNA-seq data analysis of MDA-MB-231 tumorspheres identified many oncogenes, associated with metastasis, cancer stemness, and chemoresistance that exhibit differential expression upon modulation of *LINC00273* transcription. RNA sequencing data was firmly validated in MDA-MB-231 xenografted nude mice model. Breast tumor analysis was correlated with significant downregulation of metastatic, stemness, chemoresistance marker genes upon M2 treatment with negligible metastasis to distance organs in the M2 treated mice. *LINC00273* regulates certain miRNAs, referred to as onco-miRNAs that interfere with the transcription and post-transcriptional processing of their respective target oncogenes. These onco-miRNAs mediated physical interaction with *LINC00273* by imperfect base-pairing and in doing so, they were

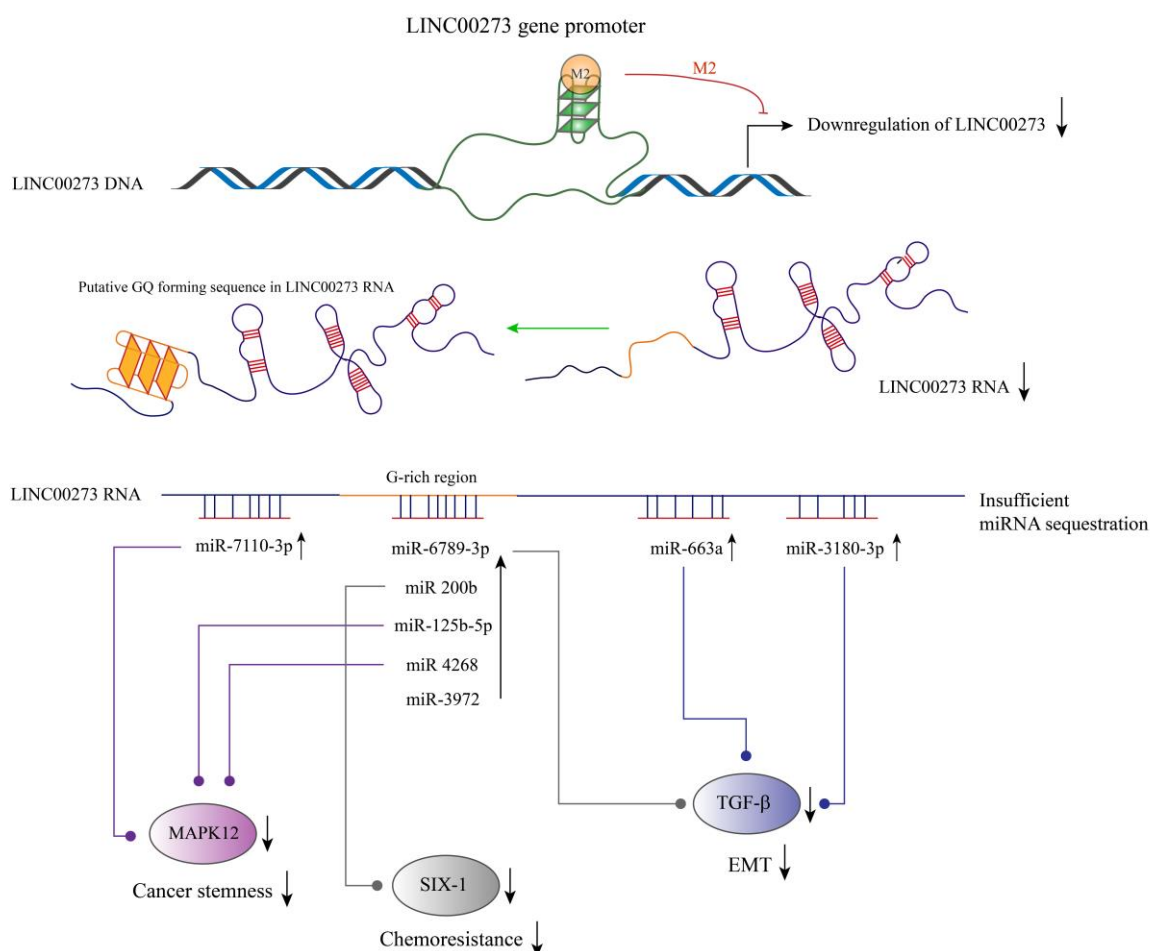


Figure 7. Schematic representation of M2 mediated downregulation of oncogenes and LINC00273-miRNA interaction. M2 binds and stabilizes the GQ in the promoter region of LINC00273 DNA sequence resulting in downregulation of LINC00273 RNA. LINC00273 RNA acts as a miRNA sponge that sequesters the mentioned miRNAs and prevents their interaction with the target oncogenes under upregulated condition. Low level of LINC00273 RNA is inadequate to sequester the miRNAs which becomes available to bind to target oncogenes and downregulate their expression.

sequestered from the functional interactions with the target mRNAs of the oncogenes regulating their subsequent expression. These findings indicated that *LINC00273* behaves as the ‘miRNA sponge’ and therefore, appears to be the master regulator of many oncogenes to regulate metastasis, chemoresistance, and cancer stemness.

Additionally, we observed the putative G4 sequence within *LINC00273* RNA to play a crucial role in determining the binding efficiency of miRNAs having complementary binding site within that region (miR-6789-5p, miR200b, miR-125b-5p, miR-4268, miR3978), which in turn

controlled the level of miRNAs available for oncogene regulation in cancer cells. TGF- β 1, SIX-1, and MAPK12 were found to be the direct downstream targets of LINC00273, whose transcription had been regulated by the miRNAs, the immediate downstream targets of LINC00273 [Figure 7]. *In-depth* investigation is required to unravel the critical role of the G4 dynamics in the *LINC00273* RNA. M2 molecule targets and stabilizes the silencer element G-quadruplex in the promoter region of *LINC00273* gene that results in the downregulation of *LINC00273* expression. Low-level of available *LINC00273* transcripts is inadequate to sequester the miRNAs, which become available to bind and downregulate their target oncogenes, while the intricate signalling cascade is assumed to crosstalk with other genes enmeshed in various signalling pathways that results in global transcriptomic change as observed in whole transcriptome profile of M2 treated tumorspheres compared to control. M2-targeted *LINC00273* downregulation showed promising results guiding us towards the development of lncRNA-based antimetastatic treatment against triple negative breast cancer. Precise targeted interference of *LINC00273*-miRNA interaction that contributes to cancer metastasis shall be of great therapeutic impact in near future to treat metastatic cancer.

6.4 Material and methods

Cell culture

The metastatic cell lines (MDA-MB 231 and A549 cells) were grown in DMEM media were cultured and maintained in DMEM (Cat. # 11995-065). The media was supplemented with 10% (v/v) FBS, 1% Pen-Strep, 50 μ g/ml Gentamicin, and 2.5 μ g/ml Amphotericin B. ESCO cell culture CO₂ Incubator was used to maintain the mammalian cells. A stable temperature of 37°C with 5% CO₂ supply and humid environment was maintained in the incubator for proper cell culture. Cell passages were maintained in T-25 flasks suitable for adherent cell culture. Cells were passaged every 3 to 4 days with media replenishment every 24 hours to maintain healthy growth of cells.

M2 drug

The bis Schiff base compound M2 was dissolved in water at a stock concentration of 5mg/ml. Sub-stocks were freshly prepared before performing the experiments. The molecular weight of M2 molecule is 426.65 kDa.

Semi-quantitative PCR

Semiquantitative Reverse transcriptase-PCR was employed to examine the change in RNA expression of LINC00273 and GAPDH upon M2 treatment. Final PCR reactions were performed using DreamTaq PCR master mix (2X) [K1071] according to the manufacturer's protocol in Proflex thermal cycler. Thermocycling conditions: Initial denaturation at 95°C (10 minutes), 40 cycles: Denaturation at 95°C (30 seconds), Annealing temperature (45 seconds), Extension at 72°C (45 seconds). To verify the presence of a single band of the anticipated size, the amplified PCR products were resolved in a 2% agarose EtBr gel. After normalizing the data using the control, the relative expression of the genes was determined. PCR primers were designed using Primer-BLAST, NCBI, and analyzed in OligoAnalyser 3.1-IDT. List of primers are provided in Table 2.

Tumorsphere formation assay (TFA)

TFA was performed in MDA-MB-231 and A549 metastatic cell lines. 3D spatial conforming property of tumor cells was mimicked by exposing the cells to heterogenous level of nutrients, oxygen, physical, chemical, and other stress factors⁹. The cells were cultured as monolayers until they reached 80–90% confluency. Following which cell media was removed and the cells washed with 1X PBS. Cells from each well were scraped, collected, and centrifuged at 300xg for 5min at RT. After aspirating off the supernatant, each pellet was resuspended in DMEM F12 (serum free) cell growth media. Media reconstitution: bFGF (20ng/ml), recombinant EGF (20ng/ml), BSA(0.4%), insulin(5µg/ml), B27 supplement (BD biosciences) (2%), and 500ml

of DMEM. The resuspended cells were seeded into ultralow-adhesion plates at a cell density of 5000 cells/well of a 6-well plate. Cells were incubated in a CO₂ incubator maintained at 5% CO₂ and 37°C temperature. Each week, half of the media was replaced with fresh culture media. On the 7th day of incubation, the formation of spheres was initiated and subsequently observed under the phase-contrast microscope.

Scratch-wound assay

This technique is also alternately called as ‘wound-healing’ assay. We observed the collective 2D migration potential of MDA-MB-231 cells¹⁰ at different time-points upon M2 treatment. MDA-MB-231 and A549 cells were seeded in 60mm adherent culture dish at a cell density of 0.5×10^5 /well and kept in the CO₂ incubator supplied with 5% CO₂, maintained humid environment and 37°C. The following day, the media was changed into serum-free media to suppress cell proliferation as it would affect correct data analysis of cell migration study. A horizontal straight-lined scratch was made in the properly adhered monolayered cell dish. The scratch out cells were washed away with 1× PBS and replenished with complete media. Centering the plate's lid edge in each well created reference points. These points enabled us to locate scratches with a straight-line tip and acquire images consistently in the same field. The cell dishes with the made scratch or wound on it, was incubated in the CO₂ incubator and imaged periodically at 12 hours, 24 hours and 48 hours’ time points. The microscopic images were analyzed to examine the effect of drug treatment on cell migration.

Real time PCR

The qPCR amplification of the LINC00273 and other genes was done upon M2 treatment. PowerUp™ SYBR™ Green Master Mix (Cat. # A25742) and the specific designed primers were used to prepare the samples for PCR. QuantStudio 5 thermal cycler was used to perform the PCR cycles. The thermal cycling conditions were as follows: Initial denaturation step:

Heating at 95°C for 10 min, 40 cycles of initial denaturation at 95°C for 30 secs, in appropriate melting temperature for 30 secs, extension at 72°C for 30 secs. Quantification of the relative expression of the genes is done by using the equation mentioned below:

$$\Delta\Delta Ct = Ct (GAPDH) - Ct (\text{target gene})$$

$$\text{Fold enrichment: } 2^{(-\Delta\Delta Ct)}$$

Each data was normalized with the control.

Oncogenes	Primer pairs
<i>LINC00273</i>	F: GCCACACAGTAGGTGACGAG R: ACTGCTTTCGGGAGAGAATG
<i>SLUG</i>	F: CTGGCCAAACACAAGCAG R: ACCCAGGCTCACATATTCCT
NME2	F: GGACTTCTGCATTGAGGTTGGC R: TGTAGTCAACCAGTTCTTCAGGC
ZEB1	F: GGCATACACCTACTCAACTACGG R: TGGGCGGTGTAGAATCAGAGTC
c-JUN	F: CCTTGAAAGCTCAGAACTCGGAG R: TGCTGCGTTAGCATGAGTTGGC
MAPK12	F: CAGTTCCTCGTGTACCAGATGC R: CACAGTCTTCGTTACAGCCAG
MAP4K3	F: CGGAAGTCTTTCATTGGCACACC R: GTGATTCCCACTGCCCAGAGAT
TGF- β 1	F: TACCTGAACCCGTGTTGCTCTC R: GTTGCTGAGGTATCGCCAGGAA
SIX-1	F: AGGTCAGCAACTGGTTTAAGAACC R: GAGGAGAGAGTTGGTTCTGCTTG
NANOG	F: CTCCAACATCCTGAACCTCAGC R: CGTCACACCATTGCTATTCTTCG
OCT4	F: CCTGAAGCAGAAGAGGATCACC R: AAAGCGGCAGATGGTCGTTTGG
SOX2	F: GCTACAGCATGATGCAGGACCA

	R: TCTGCGAGCTGGTCATGGAGTT
REX1	F: CGCAATCGCTTGTCTCAGAGT R: GCTCTCAACGAACGCTTTCCCA
GAPDH	F: ACAACTTTGGCATTGTGGAA R: GATGCAGGGATGATGTTCTG

Table 2. List of primers used to check the mRNA expression of genes.

LINC00273 overexpressing clone transfection

LINC00273 coding sequence was sub-cloned into pcDNA 3.1 (-) vector at KpnI and EcoRI restriction sites downstream CMV promoter and upstream bGH poly (A) signal to construct the *LINC00273* overexpressing plasmid. Clones were bought from BioBharati LifeScience Pvt. Ltd. 500 ng of plasmid was transfected into MDAMB-231 cells (in monolayer) using Lipofectamine 2000 reagent following manufacturer's protocol. However, for transfection into tumorspheres, Lipofectamine LTX was used as per manufacturer's recommendation in Opti-MEM (reduced serum), devoid of the antibiotics in order to preserve viability during the transfection process. 500 ng of the *LINC00273* overexpressing plasmid was then bathed over 50 µl matrigel in single well followed by the replacement of the aforementioned spheroid medium in the very next day.

Next generation sequencing

The analysis of whole transcriptome expression profile under *LINC00273* overexpressed and M2 treatment, RNA-sequencing study was done in MDA-MB- 231 tumorspheres. Using TRIzol and following standard protocol of RNA isolation, total RNA was extracted from the tumorspheres and the extracted RNA purity and concentration were checked in nanodrop spectrophotometer (the ratio of absorbance at 260/280 nm as a measure of protein contamination, and at 260/230 nm as a measure of carbohydrate contamination). RNA integrity was inspected using capillary electrophoresis (BioAnalyzer 2100, Agilent Technologies, Santa Clara, CA) to ensure the achievement of RNA integrity number (RIN) (28S to 18S ribosomal

RNA) >9. This is essential for NGS library preparation. RiboCop rRNA depletion kit of Lexogen company was used to process the total RNA and according to the manufacturer's instruction, ribosomal RNA was removed from the total RNA. Sequencing run was performed at Medgenome India Pvt. Ltd. using Illumina® HiSeq 4000/x platform. Sequencing depth for each sample was 100 million reads and we obtained 150-bp paired end reads. The raw reads underwent quality filtering and adapter trimming using Trimmomatic (v-0.36). Unwanted sequences like transfer RNAs (tRNAs), mitochondrial genome, ribosomal RNAs (rRNAs) using Bowtie2 (2.2.4). Filtered reads were aligned to Human genome (hg19) and the mapped reads of transcripts were quantified using splice aware aligner like STAR. The number of reads that were uniquely mapped was estimated utilizing Feature Count version 1.5.2. The raw read counts obtained from both treated samples and control samples were normalized using DESeq2. A normally distributed $\log_2(\text{fold change})$ values were used to identify differentially expressed genes, which were 2 standard deviations away from the mean. The mean calculation was done using the formula: (mean + 2×Standard deviation, mean- 2×Standard deviation). Reactome pathway information and GO annotations for differentially expressed genes were analyzed using the Panther database. The GEO accession number is GSE217692. The RNA-sequencing data identified 1,481 genes, with 479 upregulated and 376 downregulated. The statistical significance was (FDR > 0.05, and fold change [FC] > 2) in the *LINC00273*-overexpressed cells compared to the untreated ones. In contrast, in M2-treated cells, 294 genes were found to be differentially upregulated and 376 were differentially downregulated with the statistical significance (FDR > 0.05, and fold change [FC] > 2). Of these genes, 894 overlapped between *LINC00273*-overexpressed and M2-treated cells and 40 oncogenes were linked to EMT-induced metastasis, stemness, proliferation and drug-resistance oncogenic processes.

TNBC Nude mice model generation and M2 treatment

M₂ drug preparation

Molecular weight of the bis Schiff base molecule M₂ is 426.65 kDa. 5mg/ml stock was prepared in water and a final dose of 12.76 mg/kg body weight which is equal to 35 μ M was prepared in 1X PBS for mice treatment. The standardized treatment dosage was used from our previous investigation (<https://doi.org/10.18632/oncotarget.22622>).

Xenografted Nude mice generation and M2 treatment

Nude mice were maintained in highly sterile and 22 to 25°C ambient temperature, 12 hours of light and 12 hours of darkness, access to drinking water, and a diet of pellets. During the experiment the age of the female mice were ~ 5months of body weight ~ 20 gm. The Institute's stringent animal care ethics guidelines were followed when working with animals.

MDA-MB-231 cells were cultured in RPMI media supplemented with 10% FBS. 3 lakh cells were resuspended in 50 μ l of 1X PBS and 50 μ l of Matrigel and immediately injected into each female nude mice intraperitoneally in the breast region. Tumor growth was observed in 2 weeks. Treatment was started with a dose of 35 μ M of M2 drug dissolved in 1X PBS and injected intraperitoneally with an interval of 48 hours between each dose with a total of 4 dosages given.

Tissue collection and histological studies

All the animals were euthanized and sacrificed after the last dose of M2 treatment. Breast tumor tissues, lungs, liver, kidneys and of the mice were collected, cleaned in 0.9% saline, soaked in filter paper and processed for histological studies and gene expression analysis. Tumor, lungs, liver, and kidney tissues were fixed in freshly prepared 4% paraformaldehyde (PFA) at 4°C overnight followed by graded dehydration in alcohol and paraffin embedded for tissue sectioning. Paraffin blocks were used to obtain 5 μ m thick sections of tumor tissues

cut through the central core region. The tissue sections were rehydrated followed by staining in haematoxylin and eosin stain. Stained sections were observed and images captured under inverted light microscope EVOS M 5000.

RNA isolation from tumor tissue of Nude mice with xenografted breast tumor

100mg of tumor tissue was cut using sterile scalpel and proceeded for RNA isolation. The tissue was made into powdered form in a mortar pestle with 1ml of chilled Trizol in presence of liquid nitrogen and transferred to a 1.5 ml tube. 200 µl of chilled chloroform per ml of Trizol was added, vortexed gently and incubated for 5 minutes on ice followed by centrifugation at 12000 rpm for 15 minutes at 4°C for phase separation. The upper aqueous phase was collected in a fresh 1.5ml tube and equal volume of chilled isopropanol was added and incubated for 10 minutes followed by centrifugation at 12000 rpm for 10 minutes at 4°C. The supernatant was discarded and the pellet was washed twice with 75% of chilled ethanol with centrifugation at 7500 rpm for 5 minutes at 4°C. The RNA pellet was air dried and dissolved in 30 µl of DEPC water, RNA concentration and the purity was checked in nanodrop instrument spectrophotometrically and proceeded for cDNA synthesis and real time PCR with specific primers [Table 2]. cDNA was prepared from the isolated RNA using thermoscientific RevertAid First Strand cDNA Synthesis Kit (K1622) followed by real time PCR using specific gene primers and applied biosystems PowerUp SYBR Green Master Mix (A25742). Real time PCR was performed in QuantStudio5 instrument.

lncRNA-miRNA interaction studying bioinformatic tools

Bioinformatic studies

MicroRNAs, implicated in metastasis, stemness, and drug-resistance in breast cancer were selected from NCBI PubMed database. miRNA sequences were obtained from miRBase registry¹¹. Target prediction of miRNA for the 3'-untranslated regions (UTR) of oncogene

mRNA was determined by miRDB database¹². To predict the binding efficiency between *LINC00273* and miRNAs, Freiburg RNA tool was used¹³. To further predict the interaction between miRNA and 3'-UTR of mRNA, TargetScan software was used.

miRNA isolation and stem loop PCR

miRNA screening was done by performing stem-loop PCR: A series of stem-loop PCR¹⁴ was performed to detect and quantify the changes in miRNA expression in MDA-MB-231 cells upon *LINC00273* overexpression and M2 treatment.

- **miRNA isolation**

miRNA of control and M2 treated cells was isolated using Roche High Pure miRNA Isolation Kit (05080576001Roche). Cells were trypsinized and washed with 1X PBS before proceeding for miRNA isolation. 20% binding buffer was used to lyse the cells according to the kit's protocol. The lysate was centrifuged at 13000xg for 2 minutes and the supernatant was collected. After three steps of buffer wash and centrifugation at 13000xg, the miRNA was eluted in the elution buffer provided in the kit. The isolated pure miRNA was proceeded for stem loop PCR.

- **Stem-loop PCR**

Stem loop PCR was done to check the expression level of the miRNAs under our study¹⁵. Verso cDNA synthesis kit was used to make cDNA of the miRNAs using the stem loop primer (stem loop primer sequence is provided in *Table 3*). 50 ng of isolated pure miRNA was used for the preparation of cDNA. Thermal cycling condition was: 16°C for 30 minutes, 42°C for 30 minutes, 85°C for 5 minutes and at 95°C for 1 minute. Next, we performed real time quantitative PCR to quantify the relative expression of the miRNAs. PowerUp™ SYBR™ Green Master Mix (Cat. # A25742) was used to perform qPCR with specific forward primer and the universal reverse primer to amplify the cDNA. The list of

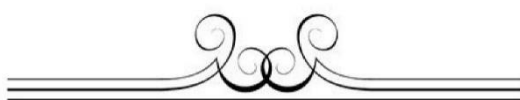
miRNAs	Primers
miR-200b	RT: TCATCA Forward: CGCTAATACTGCCTGGTAATGATGA
miR-30a	RT: CGGAAG Forward: ATAACTTACTCATGTTCTA
miR-3180-3p	RT: CGGAAG Forward: GAGCTTCCAGACGCTCCGCCCCACG
miR-4268	RT: ACATCA Forward: TTGCCCTAGCGGCTCCTCC
miR-7110-3p	RT: GAGAGA Forward: GCCCTAGCGGCTCCTCCTC
miR-125b-5p	RT: CAGCTC Forward: TAGGAGCTGGGGTGTCTTC
miR-6789-3p	RT: CAGTGA Forward: GTCCCAGGCTGGGCCCCCTC
miR-4306	RT: CAGAGG Forward: CAGAGGCATCCCTAAC
miR-3972	RT: CTGGCA Forward: GTCCTGTGGGAATGAGAG
miR-663a	RT: ACCAGG Forward: GCCCTCGTGTCTGTGGCG
U6	RT: CGCTTCACGAATTTGCGTGTCA Forward: GCTTCGGCAGCACATATACTAAAAT
Universal	GAAAGAAGGCGAGGAGCAGATCGAGGAAGAAG ACGGAAGAATGTGCGTCTCGCCTTCTTTC
Reverse	CGAGGAAGAAGACGGAAGAAT

Table 3. List of primers used in stem loop PCR. Universal primer was same for each RT primers.

primers used in stem loop PCR are provided in *Table 3*. Housekeeping gene, *U6* was used as an internal control to normalize the variability in miRNA expression levels. Relative fold change in miRNA expression was measured by $2^{-\Delta\Delta C_T}$ method.

6.5 References

1. López-Urrutia, E., Bustamante Montes, L. P., Ladrón de Guevara Cervantes, D., Pérez-Plasencia, C. & Campos-Parra, A. D. Crosstalk Between Long Non-coding RNAs, Micro-RNAs and mRNAs: Deciphering Molecular Mechanisms of Master Regulators in Cancer. *Front Oncol* **9**, 669 (2019).
2. Yan, H. & Bu, P. Non-coding RNA in cancer. *Essays Biochem* **65**, 625–639 (2021).
3. Sun, B. *et al.* Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer. *Oncol Lett* **19**, 595–605 (2020).
4. Jana, S. *et al.* LINC RNA00273 promotes cancer metastasis and its G-Quadruplex promoter can serve as a novel target to inhibit cancer invasiveness. *Oncotarget* **8**, 110234–110256 (2017).
5. Shu, C. *et al.* The downregulation of LINC00273 inhibits the proliferation, invasion, and migration of ovarian cancer cells in vivo and in vitro. *Annals of Translational Medicine* **10**, 1139–1139 (2022).
6. Chen, Z., Yao, L., Liu, Y. & Zhu, P. LncTIC1 interacts with β -catenin to drive liver TIC self-renewal and liver tumorigenesis. *Cancer Letters* **430**, 88–96 (2018).
7. Zhang, W., Guan, X. & Tang, J. The long non-coding RNA landscape in triple-negative breast cancer. *Cell Prolif* **54**, e12966 (2021).
8. Venkatesh, J., Wasson, M.-C. D., Brown, J. M., Fernando, W. & Marcato, P. LncRNA-miRNA axes in breast cancer: Novel points of interaction for strategic attack. *Cancer Letters* **509**, 81–88 (2021).
9. Dontu, G. *et al.* In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* **17**, 1253–1270 (2003).
10. Liang, C.-C., Park, A. Y. & Guan, J.-L. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat Protoc* **2**, 329–333 (2007).
11. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**, D155–D162 (2019).
12. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* **48**, D127–D131 (2020).
13. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res* **45**, W435–W439 (2017).
14. Kramer, M. F. Stem-loop RT-qPCR for miRNAs. *Curr Protoc Mol Biol* **Chapter 15**, Unit 15.10 (2011).
15. Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* **33**, e179 (2005).



Publications

- Anindya Dutta, **Ananya Roy**, Subhrangsu Chatterjee, Long noncoding RNAs in cancer immunity: a new avenue in drug discovery, **Drug Discovery Today**, Volume 26, Issue 1, 2021, Pages 264-272, ISSN 1359-6446, <https://doi.org/10.1016/j.drudis.2020.07.026>. (Equal contribution **1st Author**)
- Anindya Dutta, **Ananya Roy**, Laboni Roy, Samit Chattopadhyay, Subhrangsu Chatterjee, Immune response and possible therapeutics in COVID-19, **RSC Adv.**, 2021, **11**, 960-977, <https://doi.org/10.1039/D0RA08901J>. (Equal contribution **1st Author**)
- **Ananya Roy**, Oishika Chatterjee, Nilanjan Banerjee, Tanaya Roychowdhury, Gopa Dhar, Gopeswar Mukherjee & Subhrangsu Chatterjee (2022) Curcumin arrests G-quadruplex in the nuclear hyper-sensitive III₁ element of *c-MYC* oncogene leading to apoptosis in metastatic breast cancer cells, **Journal of Biomolecular Structure and Dynamics**, 40:20, 10203-10219, DOI: [10.1080/07391102.2021.1940284](https://doi.org/10.1080/07391102.2021.1940284) (**1st Author**)
- **Ananya Roy**, Debadrita Basu, Debopriya Bose, Anindya Dutta, Shubhra Ghosh Dastidar, Subhrangsu Chatterjee, Identification and characterization of a flexile G-quadruplex in the distal promoter region of stemness gene REX1, **International Journal of Biological Macromolecules**, Volume 231, 2023, 123263, ISSN 0141-8130, <https://doi.org/10.1016/j.ijbiomac.2023.123263>. (**1st Author**)

Communicated

- Pallabi Sengupta, **Ananya Roy**, Laboni Roy, Debopriya Bose, Souriy Chakraborty, Satyajit Halder, Kuladip Jana, Gopeswar Mukherjee and Subhrangsu Chatterjee, Long non-coding intergenic RNA, LINC00273 induces cancer metastasis and stemness *via* miRNA sponging in Triple Negative Breast Cancer. Under revision in **NAR Cancer**. (Equal contribution **1st Author**)
- Laboni Roy, Oishika Chatterjee, Debopriya Bose, **Ananya Roy**, Subhrangsu Chatterjee, Evaluation of epigenetic landscape in tumorigenesis: non-coding RNAs as new target. Under revision in **Drug Discovery Today**. (**2nd Author**)
- Laboni Roy, **Ananya Roy**, Debopriya Bose, Nilanjan Banerjee, Subhrangsu Chatterjee, Unraveling the structural aspects of the G-quadruplex in SMO promoter and elucidating its contribution as a transcriptional regulator. Communicated to **International Journal of Biological Macromolecules**. (**2nd Author**)
- Debopriya Bose, **Ananya Roy**, Nilanjan Banerjee, Pallabi Sengupta, Subhrangsu Chatterjee, Switchable tetraplex elements in the heterogeneous nuclear

ribonucleoprotein K promoter: micro-environment dictated structural transitions of G/C rich elements. Communicated to **ACS Bio & Med Chem**. (2nd Author)

- Debopriya Bose, **Ananya Roy**, Laboni Roy, Subhrangsu Chatterjee, Nucleic Acid Sensors and Logic Gates, Chapter 8, Nucleic Acid Biology and its Application in Human Diseases. Ahead of print in **Springer Nature** (Equal Author contribution)
- Debopriya Bose, Laboni Roy, **Ananya Roy**, Subhrangsu Chatterjee, Nucleic Acid in Nanotechnology, Chapter 6, Nucleic Acid Biology and its Application in Human Diseases. Ahead of print in **Springer Nature** (Equal Author contribution)



Curcumin arrests G-quadruplex in the nuclear hyper-sensitive III₁ element of *c-MYC* oncogene leading to apoptosis in metastatic breast cancer cells

Ananya Roy^a, Oishika Chatterjee^{a*}, Nilanjan Banerjee^{a*}, Tanaya Roychowdhury^{b*}, Gopa Dhar^a, Gopeswar Mukherjee^c and Subhrangsu Chatterjee^a

^aDepartment of Biophysics, Bose Institute, Kolkata, India; ^bDepartment of Cancer Biology and Inflammatory Disorder, IICB, Kolkata, India;

^cBarasat Cancer Research and welfare Centre, Barasat, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

c-MYC is deregulated in triple negative breast cancer (TNBC) pointing to be a promising biomarker for breast cancer treatment. Precise level of *MYC* expression is important in the control of cellular growth and proliferation. Designing of *c-MYC*-targeted antidotes to restore its basal level of cellular expression holds an optimistic approach towards anti-cancer treatment. *MYC* transcription is dominantly controlled by Nuclear Hypersensitive Element III-1 (NHE_{III1}) upstream of the promoter region possessing G-Quadruplex silencer element (*Pu-27*). We have investigated the selective binding-interaction profile of a natural phytophenolic compound Curcumin with native *MYC* G-quadruplex by conducting an array of biophysical experiments and *in silico* based Molecular Docking and Molecular Dynamic (MDs) simulation studies. Curcumin possesses immense anti-cancerous properties. We have observed significantly increased stability of *MYC*-G Quadruplex and thermodynamic spontaneity of Curcumin-*MYC* GQ binding with negative ΔG value. Transcription of *MYC* is tightly regulated by a complex mechanism involving promoters, enhancers and multiple transcription factors. We have used Curcumin as a model drug to understand the innate mechanism of controlling deregulated *MYC* back to its basal expression level. We have checked *MYC*-expression at transcriptional and translational level and proceeded for Chromatin Immuno-Precipitation assay (ChIP) to study the occupancy level of SP1, Heterogeneous nuclear ribonucleoprotein K (hnRNP), Nucleoside Diphosphate Kinase 2 (NM23-H2) and Nucleolin at NHE_{III1} upon Curcumin treatment of MDA-MB-231 cells. We have concluded that Curcumin binding tends to drive the equilibrium towards stable G-quadruplex formation repressing *MYC* back to its threshold-level. On retrospection of the synergistic effect of upregulated *c-MYC* and *BCL-2* in cancer, we have also reported a new pathway [*MYC-E2F-1-BCL-2*-axis] through which Curcumin trigger apoptosis in cancer cells.

ARTICLE HISTORY

Received 5 March 2021

Accepted 3 June 2021

KEYWORDS

c-MYC; *Pu-27*; G-quadruplex; NHE_{III1}; curcumin; MDA-MB-231; SP1; nucleolin; MD simulation


Introduction

Since 1986, *MYC* gene has been described to be amplified and overexpressed in human cancers, a number of studies consistently found that amplification of *MYC* is correlated with breast tumor progression and a poor clinical outcome, highlighting *MYC* as an attractive target for therapeutic inhibition (Kinzler et al., 1986; Meyer & Penn, 2008; Pelengaris et al., 2002). *MYC* knockout experiments in the animal models demonstrated that partial inactivation of the oncogene offers acute or sustained tumor regression in cancer cells (Maddison & Clarke, 2005; Niu et al., 2015). Designing of *c-MYC*-targeted antidotes to restore its basal level of cellular expression holds an optimistic approach towards anti-cancer treatment (Niu et al., 2015; Weinstein & Joe, 2008). *MYC* gene regulation is tightly controlled by a complex mechanism involving four promoters (P0-P3), enhancers/super-enhancers and multiple transcription factors, such as SP1, hnRNP, Cellular nucleic acid-binding protein

(CNBP), FUSE binding protein (FBP), T-cell factor (TCF), Nucleolin, and (Non metastatic 23) NM23-H2 (Balaratnam & Schneekloth, 2020). The promoter has been reported to have Far upstream element (FUSE) and the CT element Nuclear hypersensitive element (NHE_{III1}) as two important highly complex regulatory DNA elements located directly upstream of the *c-MYC* promoter that concert with DNA binding proteins to regulate *c-MYC* expression (Balaratnam & Schneekloth, 2020). The CT element, or NHE_{III1} element is a nucleosome depleted region which is highly sensitive to nuclease cleavage, located -142 to -115 bp upstream of the P1 promoter, much closer to the transcriptional start site and is less compacted to allow the binding of transcription factors (Carvalho et al., 2020). This region contains six runs of three or more contiguous guanines in the non-coding strand and cytosines in the complementary coding strand facilitating the formation of G-quadruplex (GQ) and i-motif (iM) respectively. The 27-nucleotide G quadruplex-forming

CONTACT Subhrangsu Chatterjee  subhrangsu@gmail.com; subhro_c@jcbosc.ac.in

*These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2021.1940284>.

© 2021 Informa UK Limited, trading as Taylor & Francis Group



Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Identification and characterization of a flexible G-quadruplex in the distal promoter region of stemness gene REX1

Ananya Roy^a, Debadrita Basu^{b,1}, Debopriya Bose^{a,1}, Anindya Dutta^a,
Shubhra Ghosh Dastidar^b, Subhrangsu Chatterjee^{a,*}^a Department of Biophysics, Bose Institute, EN-80 Sector V, Salt Lake, Unified Campus, Kolkata 700091, India^b Department of Bioinformatics, Bose Institute, EN-80 Sector V, Salt Lake, Unified Campus, Kolkata 700091, India

ARTICLE INFO

Keywords:

G-quadruplex
REX1
TMPyP4
BRACO-19
Receptor-ligand interaction
Destabilization
Molecular dynamics simulation

ABSTRACT

We have identified a parallel G-quadruplex (R1WT) in the distal promoter region (–821 base-pairs upstream of the TSS) of the pluripotent gene REX1. Through biophysical and biochemical approach, we have characterized the G-quadruplex (GQ) as a potential molecular switch that may control REX1 promoter activity to determine the transcriptional fate. Small-molecule interactive study of the monomeric form of R1WT (characterized as R1mut2) with TMPyP4 and BRACO-19 revealed GQ destabilization upon interaction with TMPyP4 and stabilization upon interaction with BRACO-19. This distinctive drug interactivity suggests the *in cellulo* R1WT to be a promising drug target. The endogenous existence of R1WT was confirmed by BG4 antibody derived chromatin immunoprecipitation experiment. Here in, we also report the endogenous interaction of GQ specific transcription factors (TFs) with R1WT region in the human chromatin of cancer cell. The wild-type G-quadruplex was found to interact with four important transcription factors, (i) specificity protein (Sp1) (ii) non-metastatic cell 2 (NM23-H2): a diphosphatase (iii) cellular nucleic acid binding protein (CNBP) and (iv) heterogeneous nuclear ribonucleoprotein K (hnRNP K) in the REX1 promoter. In contrast, nucleolin protein (NCL) binding was found to be low to the said G-quadruplex. The flexibility of R1WT between folded and unfolded states, obtained from experimental and computational analysis strongly suggests R1WT to be an important gene regulatory element in the genome. It controls promoter DNA relaxation with the coordinated interaction of transcription factors, the deregulation of which seeds stemness characteristic in cancer cells for further metastatic progression.

1. Introduction

Non-canonical DNA secondary structures play vital roles in maintaining the genomic architecture, contributing to cellular homeostasis [1]. Among the plethora of non-canonical structures reported, G-quadruplexes (GQs) have gained wide attention due to their biased representation within the human genome, strongly suggesting their pivotal regulatory role in several aspects of cell biology [2]. The existence of GQs and their regulatory roles have further been confirmed by their direct demonstration *in-cellulo* [3]. GQ forms by the participation of four guanine bases in a square co-planar array to build the tetrad core involving Hoogsteen hydrogen bonds between guanine bases [3–5]. Each G-quartet plate stacks on top of the other and are connected by loops. The central ring of hydrogen bonds involving O6 carbonyl groups renders quadruplex stability through the coordination of the central

cation (Na^+ , K^+) along with Van der Waal force of attraction and stacking interactions between the tetrad plates [6,7]. GQ stability is also determined by additional factors such as (i) the number of guanines per repeat, (ii) the length and base composition of the loops, and (iii) the folding conformation [1,4]. Various predictive algorithms identify probable GQ forming sequences, many of which use the typical sequence $\text{G}_{\geq 3}\text{N}_{1-7}\text{G}_{\geq 3}\text{N}_{1-7}\text{G}_{\geq 3}\text{N}_{1-7}\text{G}_{\geq 3}$ for postulation [8]. GQs are implicated in the regulation of vital cellular processes such as replication, transcription, and gene repair, while deregulation of their folding is synonymous with the prognosis of various diseases [9–11]. In this context, GQs have gained vast interest in the field of cancer biology as the malfunctioning of GQ-based switches is directly correlated with transcriptional inefficiency and accelerated genomic instability, leading to cancer initiation and progression [8]. The regulatory mechanism administered by GQ involves the direct recruitment of transcription factors (TFs) that

* Corresponding author.

E-mail address: subhrangsu@gmail.com (S. Chatterjee).¹ The authors contributed equally.<https://doi.org/10.1016/j.ijbiomac.2023.123263>

Received 25 October 2022; Received in revised form 6 January 2023; Accepted 8 January 2023

Available online 14 January 2023

0141-8130/© 2023 Elsevier B.V. All rights reserved.



Long noncoding RNAs in cancer immunity: a new avenue in drug discovery

Anindya Dutta¹, Ananya Roy¹ and Subhrangsu Chatterjee

Department of Biophysics, Bose Institute, P-1/12, CIT Scheme VIIM, Kankurgachi, Kolkata 700054, West Bengal, India



The central role of the nonprotein-coding portion of the genome, such as long noncoding (lnc)RNAs is emerging as a hidden player manipulating the immune system in cancer. lncRNAs, in association with their interacting partners, regulate the expression of various immune system genes, which are perturbed during cancer. The tissue-specific expression of lncRNAs and their importance in cellular proliferation, the tumor microenvironment (TME), epithelial–mesenchymal transition (EMT), and modulation of the cells of the innate and adaptive immune system have novel therapeutic implications in establishing lncRNAs as biomarkers and targets to overcome cancer-associated immunosuppression. In this review, we establish and strengthen the link between lncRNAs and cancer immunity.

Introduction

High-throughput sequencing technology has accelerated the study of biological systems and improved our understanding of gene expression to uncover the biological activities of the human genome. The International Human Genome Consortium revealed ~20,000 protein coding genes in the human genome, accounting for <3% of the total genomic sequence [1]. Several studies indicated that almost 85% of the human genome is transcribed, implying that the RNA pool must include numerous ncRNA transcripts [2] (Fig. 1a). Accumulating evidence has indicated the role of noncoding transcripts in regulating major biological functions controlling development, differentiation, cellular growth, DNA repair, EMT, tumor progression, cellular migration and metastasis [3–5] (Fig. 1b). The nonprotein-coding RNAs constitute the following groups based on size: ncRNAs, which include transfer RNA (tRNA) and rRNA; small ncRNAs (21–30 nucleotides in length), such as small nuclear RNA (snRNA), small interfering RNA (siRNA), miRNA, and circular RNA; lncRNAs (>200 nucleotides in length); and long intergenic ncRNA (lincRNA, occurring in between genes). miRNAs are a class of small, endogenous RNAs with important regulatory roles in targeting and silencing specific mRNAs. lncRNAs contain promoter regions that are generally

transcribed by RNA polymerase II, have 5' caps and 3' polyA tails, but lack open reading frames (ORFs), contain fewer exons, and are expressed less abundantly in a tissue-specific manner [6]. The database LNCipedia contains 127,802 transcripts from 56,946 long noncoding genes in humans and there are ~30,000 high-confidence lncRNA transcripts according to GENCODE reference genome annotations [7–9].

The mode of action of lncRNAs depends on their subcellular localization and their function is dependent on functional domains resulting from alternative splicing, which mediates their interaction with chromatin, RNA, and proteins [10]. lncRNAs modulate gene expression in various ways (Fig. 2); signaling lncRNAs act as molecular signals controlling the transcription of genes, such as XIST, which signals X-chromosome inactivation. Decoy lncRNAs act as molecular sinks to modulate specific transcription factors leading to gene expression/repression, examples include growth arrest-specific 5 (GAS5) and telomeric repeat-containing RNA (TERRA). Guide lncRNAs (e.g., HOTAIR) are involved in the proper localization of ribonucleoprotein complexes to specific chromatin regions. lncRNAs often modulate the gene expression of protein-coding genes by acting as *cis* and *trans*-acting elements. The enhancer lncRNA is transcribed from enhancer regions and interacts with the chromatin, whereas competing endogenous (ce)RNAs have multiple miRNA-binding sites, which protects mRNAs from degradation, such as H19 lncRNA and

Corresponding author: Chatterjee, S. (subhro_c@jbbose.ac.in)

¹ These authors contributed equally.


Cite this: *RSC Adv.*, 2021, 11, 960

Received 19th October 2020
Accepted 24th November 2020

DOI: 10.1039/d0ra08901j

rsc.li/rsc-advances

Immune response and possible therapeutics in COVID-19

Anindya Dutta,^{†a} Ananya Roy,^{†a} Laboni Roy,^{†a} Samit Chattopadhyay^{*b}
and Subhrangsu Chatterjee^{†a}

COVID-19 has emerged as a pandemic affecting about 213 countries in all the continents of the globe, resulting in more than 37.8 million individuals getting infected and over 1.08 million deaths worldwide, jeopardizing global human health and the economy. This presents an urgent need to develop therapies that target the SARS-CoV2 virus specifically. This review aims at presenting the available information on the coronavirus disease 2019 along with various drugs that are having widespread use until a vaccine candidate is available to aid in the development of therapeutic strategies against COVID-19.

1. Introduction

In December 2019, several people were diagnosed with a severe pneumonia of unknown origin in Wuhan, Hubei province of China. Analysis of the bronchoalveolar lavage fluid (BALF) from these patients led to the isolation of a novel coronavirus, SARS-CoV2 (Severe Acute Respiratory Syndrome coronavirus 2; previously named 2019-nCoV). SARS-CoV2 is a positive-sense single-stranded RNA virus that belongs to the betacoronavirus genus, believed to have zoonotic origins since it has high genetic similarity to the coronavirus RaTG13, isolated from bats. On 30th January 2020, the World Health Organization (WHO) declared the outbreak as a Public Health Emergency of International Concern and officially named the viral disease as coronavirus disease 2019 (COVID-19) on February 12, 2020 and on 11th March 2020, COVID-19 was assessed as a pandemic. As of 12th October 2020, COVID-19 has affected over 214 countries and territories around the world infecting >37.7 million people and proving fatal to >1.08 million individuals, thus endangering global health and the economy.^{1–4}

Corona viruses (order Nidovirales, family coronaviridae, subfamily coronavirinae) are a large family of spherical, enveloped, non-segmented RNA viruses with size ranging from 60 nm to 140 nm in diameter with external solar spike like projections on its surface which are the peplomers of viral S protein, gives it a crown like appearance under electron microscope (corona in Latin is *coronam*, which means crown).^{2,3,5–7} This virus possesses approximately 27–32 kb of single stranded positive sense RNA packed inside a helical

capsid and further surrounded by an envelope.^{7,8} Apart from infecting birds and mammals, these viruses have crossed the animal–human species barrier causing disease like upper respiratory tract infections (URTIs) resembling the common cold, to lower respiratory tract infections (LRTIs) such as bronchitis, pneumonia and severe acute respiratory syndrome (SARS).⁹

Corona viruses (CoV) are classified/divided into four genera, *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. Alpha and beta CoV carries the potential to infect mammals, while gamma and delta CoV infect birds although delta CoV infect both mammalian and avian species.^{1,7,8} Among them three coronaviruses, SARS-CoV, MERS-CoV and SARS-CoV2 belonging to beta-coronavirus group (Orthocoronavirinae subfamily) possess the potential to cause severe respiratory tract infection.^{1,2}

SARS-CoV originated in Guangdong, China in November 2002, spreading to Hong Kong in February 2003 and subsequently to various Asian countries. It infected 8422 people globally, mostly in China and Hong Kong causing 916 deaths (mortality rate being 11%) until the epidemic came to an end in June 2003.² MERS (Middle-East Respiratory Syndrome) outbreak took place in April 2012 in Jordan. As of February 29 2020, there have been 2494 confirmed cases, with majority of cases reported in Saudi Arabia, causing 859 deaths worldwide (34% fatality rate).¹⁰

The ongoing SARS CoV-2 outbreak has rapidly evolved and spreading globally. On January 7 2020, this virus was identified to have >95% homology with the bat coronavirus HKU9-1 and this virus is identified to have 79.5% genome sequence identity to SAR-CoV >70% similarity with the SARS-CoV.¹ To date, the closest relative of SARS-CoV-2 is RaTG13, identified from a *Rhinolophus affinis* sampled in Yunnan province in 2013. This virus shared 96.1% nucleotide identity and 92.9% identity in the S gene, suggesting the role of bats as coronavirus reservoirs.¹¹ The

^aDepartment of Biophysics, Bose Institute, Centenary Campus, P-1/12 C.I.T. Scheme VIIIM, Kolkata-700054, India. E-mail: subhro_c@jcbose.ac.in; subhrangsu@gmail.com; Tel: +91-033-25693340

^bBITS Pilani Goa Campus, NH-17B, Zuarinagar, Goa 403726, India. E-mail: samitc@goa.bitspilani.ac.in; Tel: +91-0832-2580856

[†] Equally contributed.

