УДК 577.213.3:616-006.6-08

B. Pasupula

Scientific supervisor: Associate Professor, PhD A. N Koval

Educational Establishment
"Gomel State Medical University"
Gomel, Republic of Belarus

GENOMIC LANDSCAPE OF PUTATIVE G-QUADRUPLEXES IN THE HUMAN MYC ONCOGENE: THERAPEUTIC IMPLICATIONS FOR ONCOLOGY

Introduction

The *MYC* proto-oncogene (symbol: *MYC*, Gene ID: 4609), located on chromosome 8 (NC_000008.11: 127,735,434–127,742,951), encodes a multifunctional nuclear phosphoprotein that acts as a master transcription factor regulating cell cycle progression, metabolism, apoptosis, and cellular transformation [1]. The c-Myc protein, through heterodimerization with Max, binds E-box DNA sequences to control the expression of thousands of target genes. Deregulation of *MYC* – through amplification, chromosomal translocation (e.g., t (8;14) in Burkitt lymphoma), or enhanced translation – is one of the most common oncogenic drivers in human malignancies, including lymphomas, leukemias, and solid tumors such as breast, lung, and colon cancer [2]. Consequently, *MYC* represents an attractive but challenging target for cancer therapy.

G-quadruplexes (G4s) are non-canonical secondary structures formed in guanine-rich DNA or RNA sequences, where planar arrays of four guanines (G-quartets) stack and are stabilized by monovalent cations [3]. G4 motifs are enriched in regulatory regions of the genome, including promoters, untranslated regions, and telomeres, where they can modulate transcription, replication, and translation. The *MYC* promoter contains a well-characterized G4–forming nuclease hypersensitivity element (NHE III_1) that negatively regulates transcription. Targeting *MYC* G4s with small-molecule ligands has emerged as a promising strategy to suppress oncogenic *MYC* expression. Given the critical role of *MYC* in oncology, a comprehensive *in silico* analysis of its G4 landscape may reveal new regulatory hotspots and potential therapeutic targets.

Goal

To identify and characterize putative G-quadruplex forming sequences in the human *MYC* gene using bioinformatic tools and to evaluate their potential significance in the context of targeted cancer therapy.

Material and methods of research

The genomic sequence of the human *MYC* gene (RefSeq accession NG_007161.2) was retrieved from the NCBI Datasets Gene database (https://www.ncbi.nlm.nih.gov/datasets/gene/). The sequence was downloaded in FASTA format (gene.fna) from the ncbi_dataset.zip archive. This file was submitted to the G4IPDB online prediction tool (https://people.iiti.ac.in/~amitk/bsbe/ipdb/pattern2.php), which analyzes both sense and anti-sense strands for canonical G4 motifs. The output table (MYC_G4.txt) included the strand, start/end positions, putative G4 sequence, and three confidence scores: cG (total guanine score), cC (total cytosine score), and cG/cC ratio. Sequences with a cG/cC ratio >20 or marked as "INF" (where cC=0, indicating extremely high G-richness without competing C-richness) were considered high-confidence predictions. The data were compiled and analyzed in Microsoft Excel.

The results of the research and their discussion

The G4IPDB analysis identified a total of 82 putative G-quadruplex forming sequences within the human *MYC* gene. Among these, 40 were located on the sense strand and 42 on the anti-sense

strand, indicating a nearly equal distribution. This abundance of G4 motifs underscores the potential for complex secondary structure formation throughout the gene locus.

A selection of the most significant high-confidence G4 sequences (cG/cC >20 or INF) is presented in Table 1. The cG/cC score reflects the ratio of guanine to cytosine content; values above 20, especially "INF" values, denote sequences with a very high probability of stable G4 formation *in vivo* due to minimal competition from complementary C-rich strands.

Table 1 – High-confidence putative G-quadruplex forming sequences in the human *MYC* gene

Strand	Start Position (bp)	Putative G4 Sequence (5'–>3')	Length (bp)	cG Score	cC Score	cG/cC Score
Sense	1602	GGGTTTTGGGGGGCTGGGGG	21	1010	10	101
Sense	1412	GGGGCTGGGGTGGGGGG	18	900	10	90
Sense	1936	GGCAGGGGAAAAGGGAGG	19	380	10	38
Sense	2340	GGGGACGGGGGCGGTGGAGAGGG	24	740	20	37
Sense	1469	GGGAGAGGAGAAGGCAGAGGG	22	310	10	31
Sense	3989	GGGGCGGGGTGGCAGGG	18	540	20	27
Sense	2318	GGACGGGGGCGGTACTGGGGG	22	780	30	26
Sense	1913	GGACGTGGTGTTGGGTAGG	20	240	10	24
Anti-sense	571	GGGGAAGGTGGGGAGG	16	480	0	INF
Anti-sense	588	GGGGAGGGTGGGGAGGG	17	600	0	INF
Anti-sense	3833	GGGGAAGGGATGGGAGG	17	440	0	INF
Anti-sense	4691	GGGGGTTGGAGG	12	430	0	INF
Anti-sense	6118	GGGATGGTGGTGGTTGGG	18	320	0	INF
Anti-sense	4270	GGGCGGGGGAAGG	13	490	10	49
Anti-sense	3646	GGTGGTGGGCGG	12	220	10	22

Note. "INF" denotes a cG/cC score approaching infinity due to a cC score of 0, indicating an exceptionally G-rich sequence with no competing cytosine runs.

Several observations emerge from this analysis. First, the most stable G4s on the sense strand, such as the sequence at position 1602 (cG/cC=101) and 1412 (cG/cC=90), possess extremely long G-runs and high cG scores, suggesting they are very likely to form stable structures that could impede polymerase progression or affect splicing [2]. Second, the presence of multiple "INF" sequences on the anti-sense strand (e.g., at positions 571, 588, 3833) is particularly intriguing. These sequences lack any cytosines in the motif, making them obligate G4–formers when the strand is single-stranded during transcription or replication. They may serve as transient regulatory switches [3].

Third, a notable pattern of sequence repetition was observed. Several motifs, such as the sequence at 328 bp and its identical copy at 7930 bp, appear twice in the gene, suggesting possible gene duplication events or conserved structural elements. This clustering of G4 motifs, particularly in the first intron and around the promoter-proximal region (e.g., near position 571–588 on the anti-sense strand), may create a "G4 landscape" that finely tunes *MYC* expression [4]. The well-known NHE III_1 G4 in the *MYC* promoter is a validated target for G4–stabilizing ligands. The novel high-confidence G4s identified here, especially those with "INF" scores, represent additional candidate targets for future drug design. Targeting these structures with specific ligands could provide a multi-pronged approach to silence *MYC* overexpression in cancer cells [5].

Conclusions

The human *MYC* oncogene harbors 82 putative G-quadruplex forming sequences, with a nearly equal distribution between the sense (40) and anti-sense (42) strands, indicating a dense G4 landscape.

Fifteen high-confidence G4 motifs with cG/cC scores greater than 20 or "INF" values were identified as the most promising candidates for stable structure formation *in vivo*.

The presence of five "INF" sequences on the anti-sense strand (cC=0) suggests that these regions are obligate G4–formers and may serve as critical regulatory elements during transcription.

The identified high-confidence G4 sequences, particularly those in the promoter-proximal region and first intron, represent novel potential therapeutic targets for small molecules aimed at suppressing *MYC* expression in human cancers.

LITERATURE

1. Dang, C. V. MYC on the path to cancer / C. V. Dang // Cell. – 2012. – Vol. 149, № 1. – P. 22–35.

2. Бобровник, Д. А. Сравнение содержания гуаниновых квадруплексов в генах человека, регулирующих апоптоз / Д. А. Бобровник // Проблемы и перспективы развития современной медицины : Сборник научных статей XIX Республиканской научно-практической конференции с международным участием студентов и молодых ученых: в 6 томах, Гомель, 05–06 мая 2022 года. – Гомель: Гомельский государственный медицинский университет, 2022. – С. 89–91.

3. Гринкевич, М. В. Сравнительный анализ методов биоинформатики для идентификации гуаниновых квадруплексов в митохондриальном геноме Kluyveromyces lactis / М. В. Гринкевич, А. О. Власенко // Молодые исследователи – регионам : материалы Международной научной конференции, Вологда, 22 апреля 2025 года. – Вологда: Вологодский государственный университет, 2025. – С. 353–355.

4. Бенько, А. С. Мутации в гене C9orf72, митохондриальная дисфункция и роль гуаниновых квадруплексов в нейродегенерации / А. С. Бенько, А. М. Пашкевич // Молодые исследователи – регионам : материалы Международной научной конференции, Вологда, 22 апреля 2025 года. – Вологда: Вологодский государственный университет, 2025. – С. 347–349.

5. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins / S. K. Mishra [et al.]. // Scientific reports. – 2016. – Vol. 6, № 1. – P. 38144.

УДК 577.213.3:614.876:616-006.6

J. Pathipati

Scientific supervisor: Associate Professor, PhD A. N. Koval

Educational Establishment
"Gomel State Medical University"
Gomel, Republic of Belarus

G-QUADRUPLEX STRUCTURES IN DNA REPAIR GENES: A BIOINFORMATICS ANALYSIS OF BRCA1 IN THE CONTEXT OF RADIATION ONCOLOGY

Introduction

Radiation therapy remains a cornerstone of cancer treatment, relying on the induction of DNA double-strand breaks in tumor cells to trigger apoptotic cell death or irreversible growth arrest. Cellular responses to such damage are coordinated through the DNA damage response (DDR) network, which activates DNA repair mechanisms and enforces cell cycle checkpoints. Despite this, a significant number of tumors evade radiation-induced cell death by upregulating DNA repair efficiency. The *BRCA1* tumor suppressor is encoded on chromosome 17q21, playing central role in homologous recombination repair, cell cycle regulation, and control of transcription [1]. Non-canonical DNA structures known as G-quadruplexes (G4) within genes like *BRCA1* may represent an additional regulatory element, influencing their expression and function, potentially involved into cellular sensitivity to radiation.

Goal

To perform an in-silico analysis of the human *BRCA1* gene sequence to identify and characterize putative G-quadruplex forming sequences (PQS) and to evaluate their potential relevance to the cellular response to ionizing radiation.

Material and methods of research

The reference genomic sequence for the human *BRCA1* gene (NG_005905.2) was obtained from the NCBI Datasets Gene database. The sequence was analyzed with the G4IPDB online