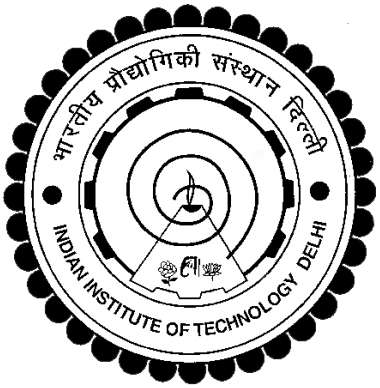
ROLE OF G-QUADRUPLEXES IN MYCOBACTERIAL GENOMES

ASHISH KUMAR



KUSUMA SCHOOL OF BIOLOGICAL SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY, DELHI

FEBRUARY 2024

ROLE OF G-QUADRUPLEXES IN MYCOBACTERIAL GENOMES

by

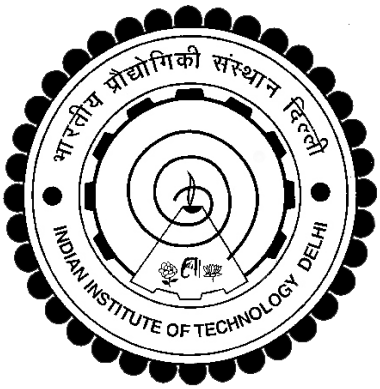
ASHISH KUMAR

Kusuma School of Biological Sciences

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY, DELHI

FEBRUARY 2024

CERTIFICATE

This is to certify that the thesis titled "Role of G-Quadruplexes in Mycobacterial Genomes", being submitted by Mr. Ashish Kumar in the Indian Institute of Technology Delhi for the award of the degree of "Doctor of Philosophy" is a record of the bonafide research carried out by him, which has been prepared under my supervision and guidance in conformity with rules and regulation of the Indian Institute of Technology Delhi, India. The results prescribed in it have not been submitted in part or full to any other University or Institute for the award of any Degree/Diploma.

Dr. Vivekanandan Perumal

Professor,

Kusuma School of Biological Sciences

Indian Institute of Technology Delhi

New Delhi-110016, India

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Prof. Vivekanandan Perumal, for his support, guidance, and mentorship throughout the entirety of my doctoral journey. His feedback has been invaluable in shaping the direction of my research and enhancing the quality of this thesis.

I am also indebted to the members of my SRC committee, Prof. Bishwajit Kundu, Prof. Archana Chugh, and Prof. Ritu Kulshreshtha, for their insightful comments, suggestions, and expertise that have significantly contributed to the refinement of this work. I am grateful to all the KSBS faculty members who have always been available for scientific discussions and suggestions. I thank them for letting me use the facilities available in their labs.

I thank my seniors Banhi, Shivani, and Nandhini, who helped me understand the G-quadruplex work. I would like to thank my lab mates and seniors, Kiruthika, Prachi, Priya, Sneha, Rishika, Arijit, Shruti, Belal, Iswarya, Sanket, Dipannita, Shivaksh, and Auroni, for bearing with me and ensuring the smooth functioning of the lab. I am grateful to my wonderful friends from the institute, Akanksha, Aquib, Debashruti, Devanshi, Firdos, Khushboo, Kritika (BK lab), Pragya, Prachi Verma, Priyanka Nair, Ramesh, Shubham, Sonalika, Tapabrata, and Uzma, for the scientific and ‘not so scientific’ discussions over tea, food, and trips. Your camaraderie has made this academic journey more enriching.

I extend my appreciation to the Kusuma School of Biological Sciences for providing the necessary resources and a conducive research environment. I thank

KSBS staff members and Ankit for their administrative and technical support. I want to thank CSIR for the fellowship.

I am grateful to my family for their unwavering love, encouragement, and understanding. Their support has been a source of patience, strength, and motivation, enabling me to navigate the complexities of doctoral research.

This thesis would not have been possible without the collective support of these individuals and institutions. Thank you for being an integral part of my academic and personal journey.

Ashish Kumar

ABSTRACT

Tuberculosis is a highly transmissible disease caused by *Mycobacterium tuberculosis* (Mtb), a member of the mycobacteriaceae family, that can endure in a latent state. Mtb is a slow-growing bacteria with a mycolic acid cell wall contributing to drug resistance, making tuberculosis treatment complex. Intracellular growth within host cells and the ability to persist in the environment further complicate the control of this infectious disease, making it a significant global health concern. G-quadruplexes have emerged as significant players in regulation at all levels of the central dogma. This study compares the distribution of putative quadruplex sequences (PQS) in mycobacteria with that in other bacteria. In contrast to the other bacteria, mycobacteria stand out and contain almost ten times more quadruplex densities in both the genomic and transcript sequences. Interestingly, our data shows that the PQS were enriched in transcript sequences (PQS density in transcripts is more than PQS density in the genome) in the case of slow-growing pathogenic mycobacteria. rG4s enrichment in just the slow-growing pathogenic cannot be a mere coincidence. Notably, the distribution of G-quadruplexes in Mtb is not uniform among the classified gene families and PE/PPE genes contain almost half the PQS found in Mtb. PE/PPE genes are two large families of genes found only in pathogenic mycobacteria and are associated with pathogenesis. Such RNA G-quadruplexes (rG4s) enrichment in any gene family has not been reported in bacteria. Our CD spectroscopy results confirm that the RNA oligonucleotides of the selected PE/PPE genes form the GQ structures, which are stabilized upon the addition of BRACO19. The results

indicate that stabilizing GQs by BRACO19 in PE/PPE genes (*PPE56, PPE67, PPE68, PE_PGRS39,* and *PE_PGRS41*) leads to the downregulation of transcription. BRACO19 also shows selective inhibition of Mtb growth at low micromolar concentrations, while the other bacteria were not affected even at much higher concentrations. Heterologous expression of the selected genes indicates that the GQ-mediated downregulation is not limited to the transcription level. Ectopic expression in both *M. smegmatis* and *E. coli* shows the downregulation of PE/PPE protein levels due to rG4s. In addition, the rG4-mediated reduction in PE/PPE protein levels attenuates proinflammatory response upon infection of THP-1 cells. Thus, our findings on the role of G-quadruplexes in regulating transcription of the PE/PPE genes, the post-transcriptional translation inhibition by rG4s in the transcripts, and its impact on macrophage pro-inflammatory response provide new insights into our current understanding of Mtb adaption, stress response and survival within the host.

सार

तपेदिक/क्षयरोग एक संक्रामक रोग है जो माइकोबैक्टेरियम ट्यूबरक्युलोसिस (एमटीबी) द्वारा प्रकारित होता है, जो माइकोबैक्टेरियमासिये परिवार का एक सदस्य है, जो एक लैटेंट स्थिति सहन कर सकता है। एमटीबी एक धीमी विकास वाला जीवाणु है जिसकी माइकोलिक एसिड सेल दीवार इसकी औषधि प्रतिरोधिता में योगदान करती है, जिससे कि क्षयरोग का इलाज जटिल हो जाता है। मेज्जातीय कोशिकाओं में अंतर्निहित वृद्धि और टिकने की क्षमता, इस संक्रामक रोग के नियंत्रण को और अधिक जटिल बनाती है, जिससे यह एक वैश्विक स्वास्थ्य संबंधी चिंता बनता है। जी-क्वाड्रुप्लेक्स केंद्रीय धर्म के सभी स्तरों पर नियमन में महत्वपूर्ण खिलाड़ियों के रूप में प्रकट हो रहे हैं। इस अध्ययन में माइकोबैक्टीरिया में पूर्वानुमानित क्वाड्रुप्लेक्स क्रम (PQS) का वितरण अन्य जीवाणुओं के साथ तुलना किया गया है। अन्य जीवाणुओं के विपरीत, माइकोबैक्टीरिया जीनोमिक और प्रतिलेख क्रमों में लगभग दस गुणा अधिक क्वाड्रुप्लेक्स घनत्व शामिल हैं। दिलचस्प बात यह है कि हमारा डेटा दिखाता है कि पूर्वानुमानित क्वाड्रुप्लेक्स क्रम (प्रतिलेखों में rG4 घनत्व जीनोम में PQS घनत्व की तुलना में अधिक है) को धीरे से विकसित होने वाले रोगजनक माइकोबैक्टीरिया के मामले में प्रतिलेखों में समृद्ध किया गया था। एमटीबी में जी-क्वाड्रुप्लेक्स का वितरण वर्गीकृत जीन परिवारों के बीच यौन नहीं है और *PE/PPE* जीनों में एमटीबी में पाए गए rG4s का लगभग आधा हिस्सा है। *PE/PPE* जीन रोगजनक माइकोबैक्टीरिया में ही पाए जाने वाले दो बड़े जीन परिवार हैं और रोगजनन से संबंधित हैं। किसी भी जीन परिवार में ऐसा आरएनए जी-क्वाड्रुप्लेक्स संवर्धन बैक्टीरिया में रिपोर्ट नहीं किया गया है। हमारे सीडी स्पेक्ट्रोस्कोपी परिणाम पुष्टि करते हैं कि चयनित PE/PPE जीनों के आरएनए अणुसंख्यात्र जी-क्यू संरचनाएं बनाते हैं, जो BRACO19 के जोड़ने पर स्थिर होती हैं। परिणाम सुझाव देते हैं कि *PE/PPE* जीनों में BRACO19 द्वारा जी-क्यू को स्थिर करने से (*PPE56, PPE67, PPE68, PE_PGRS39,* और *PE_PGRS41*) प्रतिलेखन का निम्न अधिनियमन होता है। BRACO19 कम माइक्रोमोलार अवस्थितियों पर भी एमटीबी की वृद्धि का चयनात्मक अवरोध प्रदर्शित करता है, जबकि अन्य जीवाणुओं को बहुत अधिक अवस्थितियों पर भी प्रभावित नहीं किया जाता है। चयनित जीन की विषम अभिव्यक्ति इंगित करती है कि जीक्यू-मध्यस्थता डाउनरेगुलेशन प्रतिलेखन स्तर तक सीमित नहीं है। माइकोबैक्टेरियम स्मेगमाटिस और ई. कोलाई, दोनों में अन्यत्र अभिव्यक्ति द्वारा PE/PPE प्रोटीन स्तरों की निम्नीकरण का प्रदर्शन rG4 के कारण होता है। इसके अलावा, THP-1 कोशिकाओं के संक्रमण पर rG4 के

माध्यम से PE/PPE प्रोटीन स्तरों में कमी में प्रो-इन्फ्लामेटरी प्रतिक्रिया में कमी होती है। इस प्रकार, *PE/PPE* जीन के प्रतिलेखन को विनियमित करने में जी-क्वाड्रुप्लेक्स की भूमिका पर हमारे निष्कर्ष, प्रतिलेखों में rG4 द्वारा पोस्ट-ट्रांसक्रिप्शनल अनुवाद निषेध, और मैक्रोफेज प्रो-इंफ्लेमेटरी प्रतिक्रिया पर इसका प्रभाव हमारी मेजबान के भीतर एमटीबी अनुकूलन, तनाव प्रतिक्रिया और अस्तित्ववर्तमान की समझ में नई अंतर्दृष्टि प्रदान करता है। ।

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BRACO19	N,N'-(9-{[4-(dimethylamino)phenyl]amino}acridine-3,6-diyl)bis(3-pyrrolidin-1-ylpropanamide)
c-exNDI	Core Extended Naphthalene Diimide
CFUs	Colony forming units
DHX36	DEAH Box Helicase 36
ESAT6	6 kDa Early secretory antigenic target
EBNA-1	Epstein Barr Virus Nuclear Antigen 1
EBV	Epstein Barr Virus
FRET	Fluorescence Resonance Energy Transfer
GQ	G-quadruplex
HIV	Human immunodeficiency virus
IL-10	Interleukin-10
IL-6	Interleukin-6
IRES	Internal Ribosome Entry Site
LANA	Latency-Associated Nuclear Antigen
MDR	Multi drug resistant
MPTR	Major polymorphic tandem repeats
MOI	Multiplicity of Infection
Mtb	*Mycobacterium tuberculosis*
M. smeg	*Mycobacterium smegmatis*
NGS	Next Generation Sequencing
NMR	Nuclear Magnetic Resonance

ABBREVIATIONS

OD	Optical Density
PE	Proline-Glutamate
PPE	Proline-Proline-Glutamate
PGRS	Polymorphic GC-rich repeat sequences
PQS	Putative Quadruplex Sequence
QUMA-1	(E)-2-(2-(7-(diethylamino)-2-oxo-2H-chromen-3-yl) vinyl)-6-fluoro-1-methyl-7-(4-methyl piperazin-1-yl) quinolin-1-ium iodide
rG4	RNA G-quadruplex
TB	Tuberculosi s
TERRA	Telomeric Repeat-Containing RNA
TLRs	Toll like receptors
TmPyP4	5,10,15,20-Tetrakis-(N-Methyl-4-Pyridyl)Porphine
TNF	Tumor necrosis factor
XDR	Extensive drug resistant
WHO	World Health Organization