ENHANCING SECRETORY EXPRESSION OF GRANULOCYTE COLONY STIMULATING FACTOR IN *PICHIA PASTORIS* USING GENETIC APPROACHES

MOOL CHAND

DEPARTMENT OF BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2020

ENHANCING SECRETORY EXPRESSION OF GRANULOCYTE COLONY STIMULATING FACTOR IN *PICHIA PASTORIS* USING GENETIC APPROACHES

by

Mool Chand

Department of Biochemical Engineering and Biotechnology

Submitted

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

to the

Indian Institute of Technology Delhi

October 2020

Certificate

This is to certify that the thesis entitled **'Enhancing secretory expression of granulocyte colony stimulating factor in *Pichia pastoris* using genetic approaches'** being submitted by **Mr. Mool Chand** to the Indian Institute of Technology Delhi for the award of the degree of **'Doctor of Philosophy'**, is a record of the bonafide research work carried out by him, which has been prepared under my supervision in conformity with the rules and regulations of the Indian Institute of Technology Delhi. The research reports and the results presented in this thesis have not been submitted for any degree or diploma in any other University or Institute.

Dr Saroj Mishra
Professor
Department of Biochemical
Engineering and Biotechnology
Indian Institute of Technology

Dr D. Sundar
Professor
Department of Biochemical
Engineering and Biotechnology
Indian Institute of Technology

Acknowledgements

The amazing journey of PhD has finally come to an end and there are so many people because of whom this dream has come to reality. First of all, I want to thank the **Almighty God**, the best planner who blessed me, inspired me by different sources and gave me strength to accomplish this goal.

Next, I would like to express my deepest thanks and gratitude to my PhD supervisor **Prof. Saroj Mishra.** I really admire her for her patience, foresightedness, knowledge, sincerity towards her duties and passion for science. She is an inspiration for me and one of her sentence that I will always remember and appreciate her for "I always have time for my students"

I would like to extend my warmest thanks and gratitude to my co-supervisor **Prof. D. Sundar.** I am really thankful for his guidance and critical review that always guided me towards the right direction. I admire him for his systematic and happy go lucky nature.

I would also like to thank my research committee members, **Prof. Atul Narang, Prof. Bishwajit kundu** and **Dr. Preeti Srivastava** for serving as my committee members. I am greatful to them for their brilliant comments and suggestions. I would also like to thank **Dr. Ritu Kulshreshtha** for providing access to cell culture facility.

I am thankful to all my seniors who taught me selflessly and helped me at various steps during PhD, **Dr. Anees Kaprakkaden, Dr. Nitu Maity, Dr. Ashwani Gautam, Dr. Sumit Kapoor, Dr. Arpita Vats, Dr. Rishabh, Dr. Swati, Anshul and other BRL lab members**. I would also like to thank **Dr. Jaspreet Kaur, Vidhi Malik, Nevneetan, Anjani, Shashank and Dhwani** from MMD lab for their constant support and cheerful presence. I would like to extend warm gratitude to **Rehmaan ji, Harilal ji, Mr. Rajeev Dahiya, Yogesh and Sant Ram ji** for the unconditional and prompt help they provided all through. They all are responsible for saving half of our time with their technical expertise. I also want to thank **Avijeet Singh** for being a wonderful junior. I really enjoyed working with him. I am thankful to **Vivek Anand** for helping in transcriptome data analysis.

I thank members of the **cricket team, coach and whole BSA fraternity** of IIT Delhi for the wonderful time I spent with them. It was really an amazing and learning experience to be part of them.

I am thankful to **Mr. Sumit Singh and Dr. Deepak** for being supportive seniors and for the wonderful time I spent in their company. I am really glad that I made some wonderful friends during this journey. I thank **Jitender singh verma** and **Sanjay Singh,** partners in crime for all the masti we had in these years. I also thank **Sonam Takkar (Don)** for being a supportive friend and helping in cell culture experiments.

I thank all the members of Karokoram hostel particularly, **Satish Kumar Yadav, Ishank Yadav, Anil Chaudhary, Sajal Jain, Anshul Bang, Aman and Devilal**, for all the funfilled moments and their unconditional love and support. I would also like to thank **Dev and Atul**. They are two dependable friends who were always available whenever needed. I would also like to thank **Sameer Mishra** with whom I made a special bond in a short span of time. His thoughts and poetry have always been refreshing and inspiring.

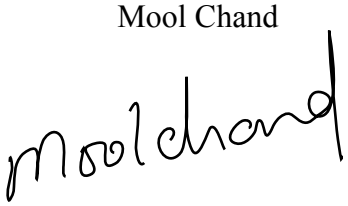
I am fortunate to have so many selfless teachers who carved my professional journey with their valuable expertise and selfless support.

There are no words to express my gratitude towards my family members, the supporting pillars of my life. All the sacrifices they made for me, their unconditional love, constant support and believe in me are the reason that I have come so far. There are no words to express what they mean to me.

I would like to give special thanks to my wife **Anjali,** who stood by me in thick and thin and showered unconditional love. She believed in me more than anyone else in my life.

I want to dedicate this thesis to all these people who have made this day possible.

Mool Chand



Abstract

Granulocyte colony stimulating factor (G-CSF) is one of the first cytokines to be discovered that finds wide applications in the treatment of neutropenia, central nervous system disorders, myeloablative therapy, myocardial infarction, hepatic damage and others. Recombinant G-CSF, filgrastim obtained from *E. coli* was approved by FDA in 1991 for the treatment of neutropenia in cancer patients following which considerable efforts have been made to produce this protein for therapeutic use. Among the various available hosts for therapeutic protein production, methylotrophic yeast *Pichia pastoris* evolved as an attractive alternative owing to its methylotrophic nature, shorter doubling time, ease of manipulation, ability to perform post translational modification, presence of strong inducible promoter and optimum cell growth over a wide pH range (pH 3 – 7) at low cost. The ultimate of aim of therapeutic protein production is to produce biologically active and stable protein in large amounts in a cost effective manner. The aim of the presented work was to enhance the production of G-CSF in *P. pastoris* by utilizing different strategies including – fusion of G-CSF with HSA and codon harmonization. Two fusion constructs of G-CSF with HSA domain III containing flexible linker and cleavable linker were generated. The fusion of G-CSF with HSADIII enhanced G-CSF production by 5-8 folds. G-CSF present in fusion protein retained its alpha helical structure and biological activity *in vitro.* Maximum protein production was obtained when cells were grown at 22°C in baffled flask. Interestingly, Mut^S strain showed higher protein production as compared to Mut^+ strain. Transcriptome data analysis was performed to reveal the underlying mechanism involved in this Mut phenotype dependent variation in heterologous protein production. Higher G-CSF production in Mut^S phenotype was accompanied by differential expression of genes involved in methanol metabolism, sterol biosynthesis, β-oxidation, amino acid metabolism, transcription, translation, unfolded protein response and biosynthesis of co-factors. In particular for methanol metabolism, *FGH1, AOX2, RPE1, SHB17*

and *RPE1-2* genes were upregulated while *DAL7* was downregulated indicating higher G-CSF production involved utilization of methanol with upregulation of assimilatory and dissimilatory pathways to prevent formaldehyde toxicity. We also explored codon harmonization to fine tune G-CSF sequence in light of specie specific codon bias and occurrence of alpha helical and link regions. Two codon harmonized sequences were generated and transformed in *P. pastoris*. Codon harmonization aided in increase in production of G-CSF protein by 10 fold with respect to native cDNA. The resulting protein showed less aggregation, retained structural and biological identity. Future studies focusing on exploring these strategies at fermenter level and evaluating the *in vivo* efficacy of resulting protein will help to establish the best possible strategy for G-CSF production in *P. pastoris* at large scale in a cost effective manner.

सार

ग्रैनुलोसाइट कॉलोनी उत्तेजक कारक (जी-सीएसएफ) खोजे जाने वाले पहले साइटोकिन्स में से एक है जो न्यूट्रोपेनिया, केंद्रीय तंत्रिका तंत्र के विकारों, मायलोब्लेटिव थेरेपी, मायोकार्डियल इन्फेक्शन, यकृत क्षति और अन्य के उपचार में व्यापक अनुप्रयोग पाता है। ई कोलाई से प्राप्त फाइब्रॉस्टिमेंट जी-सीएसएफ, 1991 में कैंसर रोगियों में न्यूट्रोपेनिया के उपचार के लिए एफडीए द्वारा अनुमोदित किया गया था जिसके बाद चिकित्सीय उपयोग के लिए इस प्रोटीन के उत्पादन के लिए काफी प्रयास किए गए हैं। चिकित्सीय प्रोटीन उत्पादन के लिए विभिन्न उपलब्ध मेजबानों में, मेथिलोट्रॉफ़िक खमीर पिचिया पस्टोरिस अपने मेथिलोट्रॉफ़िक प्रकृति के कारण एक आकर्षक विकल्प के रूप में विकसित हुआ, कम दोहरीकरण समय, हेरफेर में आसानी, अनुवाद में संशोधन करने की क्षमता, मजबूत प्रेरक प्रमोटर की उपस्थिति और इष्टतम सेल विकास से अधिक। कम लागत पर एक विस्तृत पीएच रेंज (पीएच 3 - 7)। चिकित्सीय प्रोटीन उत्पादन का अंतिम उद्देश्य लागत प्रभावी तरीके से बड़ी मात्रा में जैविक रूप से सक्रिय और स्थिर प्रोटीन का उत्पादन करना है। प्रस्तुत कार्य का उद्देश्य विभिन्न रणनीतियों का उपयोग करके पी। पादरी में जी-सीएसएफ के उत्पादन को बढ़ाने के लिए था, जिसमें एचएसए और कोडन हार्मोनाइजेशन के साथ जी-सीएसएफ का संलयन शामिल है। एचएसए डोमेन III युक्त जी-सीएसएफ के दो संलयन निर्माण जिसमें लचीले लिंकर और क्लीएबल लिंकर शामिल थे, उत्पन्न हुए थे। HSADIII के साथ G-CSF के संलयन ने G-CSF उत्पादन को 5-8 गुना बढ़ा दिया। संलयन प्रोटीन में मौजूद जी-सीएसएफ ने अपने अल्फा हेलिकल संरचना और इन विट्रो में जैविक गतिविधि को बनाए रखा। अधिकतम प्रोटीन उत्पादन प्राप्त किया गया था जब कोशिकाओं को चक्राकार फ्लास्क में 22 डिग्री सेल्सियस पर उगाया गया था। दिलचस्प बात यह है कि, MutS स्ट्रेन ने Mut + स्ट्रेन की तुलना में अधिक प्रोटीन उत्पादन दिखाया। ट्रांसपेरोमेट डेटा विश्लेषण को इस म्यूट फेनोटाइप में शामिल अंतर्निहित तंत्र को प्रकट करने के लिए किया गया था जो हेटेरोग्लस प्रोटीन उत्पादन में निर्भर है।

MutS फेनोटाइप में उच्च G-CSF उत्पादन मेथनॉल चयापचय, स्टेरोल बायोसिंथेसिस, β-ऑक्सीकरण, अमीनो एसिड चयापचय, प्रतिलेखन, अनुवाद, अनफोल्डेड प्रोटीन प्रतिक्रिया और सह-कारकों के संश्लेषण में शामिल जीनों के अंतर अभिव्यक्ति के साथ था। मेथनॉल चयापचय के लिए विशेष रूप से, FGH1, AOX2, RPE1, SHB17 और RPE1-2 जीनों को अपग्रेड किया गया था, जबकि DAL7 को उच्च G-CSF उत्पादन का संकेत दिया गया था जिसमें फॉर्मेल्डिहाइड विषाक्तता को रोकने के लिए आत्मसात और विदारक मार्गों के अपघटन से मेथनॉल का उपयोग शामिल था। हम भी विशिष्ट विशिष्ट कोडन पूर्वाग्रह और अल्फा पेचदार और लिंक क्षेत्रों की घटना के प्रकाश में जी-सीएसएफ अनुक्रम ठीक धुन करने के लिए कोडन सामंजस्य का पता लगाया। दो कोडन सामंजस्य वाले अनुक्रम उत्पन्न हुए और पिचिया पादरी में तब्दील हो गए। देशी सीडीएनए के संबंध में जी-सीएसएफ प्रोटीन के उत्पादन में 10 गुना की वृद्धि के साथ कोडोन हार्मोनाइजेशन सहायता प्राप्त है। परिणामस्वरूप प्रोटीन ने कम एकत्रीकरण दिखाया, संरचनात्मक और जैविक पहचान को बनाए रखा। भविष्य के अध्ययन किण्वक स्तर पर इन रणनीतियों की खोज पर ध्यान केंद्रित कर रहे हैं और परिणामस्वरूप प्रोटीन की विवो प्रभावकारिता का मूल्यांकन करने से लागत प्रभावी तरीके से बड़े पैमाने पर पिचिया पास्टोरिस में जी-सीएसएफ उत्पादन के लिए सर्वोत्तम संभव रणनीति स्थापित करने में मदद मिलेगी।

Table of contents

List of Figures

List of Tables

List of Abbreviations

AA	Amino acid
AOX	Alcohol oxidase
BHK	Baby hamster kidney
BMMY	Buffered Methanol-complex medium
BMGY	Buffered Glycerol-complex medium
cDNA	Complementary D-oxy ribose nucleic acid
CDS	Coding DNA sequence
CHO	Chines hamster ovary cells
CAI	Codon adaptive index
CD	Circular dichroism
DHAP	Dihydroxy acetone phosphate
DNA	D-oxyribose nucleic acid
FDA	Food and drug administration
G-CSF	Granulocyte colony stimulating factor
G-CSFR	Granulocyte colony stimulating factor receptor
GH	Growth hormone
GO	Gene ontology
HEK	Human embryonic kidney
Hr	Hour
HSA	Human serum albumin
HSCs	Hematopoietic stem cells
HIV	Human immuno virus
IL	Interleukin
IFN-β	Interferon-β
kDa	Kilodalton

LPS	Lipopolysaccharides
Mins	Minutes
Mg	Milligram
mL	Millilitre
MUT	Methanol utilization
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
mAbs	Monoclonal antibodies
mRNA	Messenger Ribose nucleic acid
OD	Optical density
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PAGE	Polyacrylamide gel electrophoresis
SDS	Sodium dodecyl sulphate
SCID	Severe combined immunodeficiency syndrome
TNF-α	Tumor necrosis factor α
tRNA	Transfer Ribose nucleic acid
VEGF	Vascular endothelial growth factor
US	United states
YPD	Yeast peptone dextrose