

**TO ENHANCE THE SECRETORY EFFICIENCY OF  
GRANULOCYTE COLONY STIMULATING FACTOR IN  
*PICHIA PASTORIS* VIA MANIPULATION OF SECRETORY  
SIGNAL SEQUENCES**

**SAKSHI AGGARWAL**



**DEPARTMENT OF BIOCHEMICAL ENGINEERING & BIOTECHNOLOGY  
INDIAN INSTITUTE OF TECHNOLOGY DELHI  
SEPTEMBER 2020**

©Indian Institute of Technology Delhi (IITD), New Delhi, 2020

**TO ENHANCE THE SECRETORY EFFICIENCY OF  
GRANULOCYTE COLONY STIMULATING FACTOR IN  
*PICHIA PASTORIS* VIA MANIPULATION OF SECRETORY  
SIGNAL SEQUENCES**

by

**SAKSHI AGGARWAL**

**DEPARTMENT OF BIOCHEMICAL ENGINEERING & BIOTECHNOLOGY**

*Submitted*

*In the fulfilment of the requirements for the degree of doctor of philosophy  
to the*



**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**SEPTEMBER 2020**

# DEDICATION

---

*This research is dedicated to my beloved  
parents and my supervisor...*

# CERTIFICATE

---

This is to certify that the thesis entitled “ **To enhance the secretory efficiency of granulocyte colony stimulating factor in *Pichia pastoris* via manipulation of secretory signal sequences**” being submitted by **Ms. Sakshi Aggarwal** 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 her, which has been prepared under our supervision and guidance 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.

**Professor Saroj Mishra**

Professor

Department of Biochemical Engineering & Biotechnology

Indian Institute of Technology Delhi

**Date:**

**Place:**

# ACKNOWLEDGMENTS

---

With ending this long journey of PhD, I would like to express my sincere gratitude and respect to my supervisor **Prof. Saroj Mishra** for her support, guidance, motivation and immense knowledge which gave me power and spirit to excel in this field. She guided me so nicely through the entire project with her wide knowledge and innovative ideas. She has always been more enthusiastic about the experimental results, her strong dedication towards research and regular lab visits fuelled me to excel in this area of research. Whenever an experiment fails, she reacts so positive and always have a way out. I am highly grateful to her for my never-ending fount of moral support. She has endlessly corrected my thesis with her discerning vision for which I am really thankful. Apart from this, she has been a counsellor with whom I could discuss all my anguishes and take advice. She is an epitome of wisdom, knowledge and perfection, an ideal person I have ever met in my career. She is very responsible and caring mentor, I feel myself lucky under her guidance.

I would like to acknowledge my SRC members Prof. Prashant Mishra, Prof. S.K. Khare and Dr. Ritu Kulshreshta for their invaluable comments and advice. I sincerely thank Dr. M.K Aggarwal and Dr. A.K Sau for always being there and helping me with all the important decisions of my life. I am very thankful to DST for providing me fellowship during PhD. upto 5-years.

I would give my special thanks to Mr. Sumeet Kapoor, for always helping me with access to instruments in the laboratory. Thanks to Rehmanji, Harilalji and Yogesh for providing me with all the lab consumables and clean glasswares, whenever necessary.

I am indebted to my seniors Sumbul Afreen, Swati Ohja, Sujata, Tushar, Tenzin kemzon and Anees K for their kind help and guidance. My sincere thanks to Anshul Sharma, who really introduced me to *P. pastoris* cultivation and process optimization experiments very nicely. A

very special thanks and love to my juniors Avijeet Jaiswal and Manju Khicher for their invaluable advice in addressing a research problem. I am indebted to Dr. Ayan Dutta, Scientist (Qiagen) for guiding me in my cloning experiments. He is such a humble person who has really helped me in troubleshooting PCR failures. I would always remember my friends Sumbul Afreen, Manju khicher and Aakanksha Ahlawat for the fun time and numerous tea break discussions on research topics to future plans. I would like to give special thanks to Neetu Singh and Kritika Goyal (sister cum best friend), for her unconditional love and support during my ups and down. Special thanks to my lab mates Rohit, Ashima, Radhika, Aishvi, Asmita for maintaining healthy environment in lab. I would give my special thanks to my friend Ruchika Bhatt for helping me in computational work. I am thankful to my batchmates Richa Sharma, Sonam Takkar, Gautam Anand for their kind help and support.

The unconditional love and support of my family members is really acknowledgeable. My heartfelt thanks to my family members: Mom, Dad, Lakshay, Akshay and Shriesty for their unconditional love and affection. My Dad, an ideal person, always motivated me to fulfil my dreams. His hard work and dedication is always an inspiration for me. My Mom (superhero) my best friend has a major role during my PhD, without her I would not be able to do it. Thanks for remembering me in your prayers for the ultimate success. In every step, she has supported and encouraged me throughout my life. Along with me, she had also spent sleepless nights during my PhD. She used to call me mid-nights when I had work in lab for the whole night. My heartiest thanks to my brother Lakshay for his numerous late-night visits to IITD, to take me home. It was all family love that refilled fresh energy after a failed experiment.

I would like to thank the Almighty for His blessings throughout my journey.

**...Sakshi Aggarwal**

# ABSTRACT

---

The methylotrophic yeast *Pichia pastoris* has been widely used for the production of human therapeutics, but production of granulocyte colony-stimulating factor (G-CSF) in this yeast is low. In this study, improved extracellular production of G-CSF was carried out by introducing mutations in the  $\alpha$ -mating type (MAT) secretory signal using a native cDNA (*WT-GCSF*) and a codon optimized *GCSF* gene (*CO-GCSF*). Mutations in the pro-region of the  $\alpha$ -MAT (deletion ( $\Delta$ ) of the amino acids 57-70), resulted in an increase in extracellular production of G-CSF in both the cases with higher production from the *CO-GCSF*. Response to  $\Delta$ 30-43 and  $\Delta$ 47-49 deletions was different from the *CO-GCSF* and the *WT-GCSF* genes indicating higher rates of synthesis in the former to overcome the regulatory control exercised by these segments. The loss of secretion occurring due to  $\Delta$ 30-43 in the *WT-GCSF* was partially restored (by 60%) when the  $\Delta$ 57-70 was added. An important role of the 47-49 amino acids was also demonstrated. The role of the P1' position of the kex2 cleavage site in the  $\alpha$ -MAT was demonstrated and specific substitutions by smaller amino acids lead to increased production of G-CSF. Secondary and tertiary structure prediction using I-TASSER indicated an important role of the 3<sup>rd</sup> alpha-helix in the pre-pro peptide. Presence of a minimum loop length and secondary structure on the pro-peptide region allowed enhanced extracellular production of G-CSF. Also, the role of several nutritional factors in controlling the morphology of recombinant *P. pastoris* was shown for the first time and the morphological switch was mediated through quorum sensing molecules.

मिथाइलोड्रॉफिक खमीर, पिक्रिया पास्टोरिस का व्यापक रूप से मानव चिकित्सा विज्ञान के उत्पादन के लिए उपयोग किया गया है, लेकिन इस खमीर में ग्रैनुलोसाइट कॉलोनी-उत्तेजक कारक (जी-सीएसएफ) का उत्पादन कम है। इस अध्ययन में,  $\alpha$ -संभोग ( $\alpha$ -मेट) प्रकार स्त्रावी संकेत को एक देशी (डब्ल्यू टी-जीसीएसएफ) और एक कोडन अनुकूलित जीन (को-जीसीएसएफ) का उपयोग करके जीसीएसएफ के बेहतर बाह्य उत्पादन को अंजाम दिया गया।  $\alpha$ -संभोग कारक के एमिनो एसिड ( $\Delta 47-70$ ) के विलोपन से जीसीएसएफ के समर्थक क्षेत्र में को-जीसीएसएफ से परिणामस्वरूप उच्च उत्पादन के साथ दोनों मामलों में जीसीएसएफ के बाह्य उत्पादन में वृद्धि हुई है।  $\Delta 30-43$  और  $\Delta 47-49$  विलोपन का जवाब डब्ल्यूटी-जीसीएसएफ और को-जीसीएसएफ जीन से अलग था, जो इन खंडों द्वारा नियंत्रित नियामक नियंत्रण को दूर करने के लिए पूर्व में संश्लेषण की उच्च दरों का संकेत देता है। डब्ल्यूटी-जीसीएसएफ में  $\Delta 30-43$  के कारण होने वाले स्त्राव की हानि आंशिक रूप से (60%) द्वारा बहाल की गई थी जब  $\Delta 47-70$  जोड़ा गया था। '47-49' एमिनो एसिड की महत्वपूर्ण भूमिका का भी प्रदर्शन किया गया था।  $\alpha$ -संभोग ( $\alpha$ -मेट) में केएक्सरपीश साइट की भूमिका का प्रदर्शन किया गया और संकेत दिया कि छोटे अमीनो एसिड जीसीएसएफ प्रोटीन उत्पादन को बढ़ावा देते हैं। आई-टेसर का उपयोग करके माध्यमिक और तृतीयक संरचना की भविष्यवाणी ने पूर्व-प्रोपेप्टाइड में तीसरी अल्फा-हेलिक्स की एक महत्वपूर्ण भूमिका का संकेत दिया। प्रो-पेप्टाइड क्षेत्र पर एक न्यूनतम लूप लंबाई और माध्यमिक संरचना की उपस्थिति ने जी-सीएसएफ के अतिरिक्त बाह्य उत्पादन की अनुमति दी। इसके अलावा, पुनः संयोजक पी. पास्टोरिस की आकृति विज्ञान को नियंत्रित करने में कई पोषण संबंधी कारकों की भूमिका पहली बार दिखाई गई और रूपात्मक संवेदी अणुओं के माध्यम से रूपात्मक स्विच की मध्यस्थता की गई।

# CONTENTS

---

<b>Certificate</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii-iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Contents</b>	<b>v-xii</b>
<b>List of figures</b>	<b>xiii-xxi</b>
<b>List of tables</b>	<b>xxii-xxiii</b>
<b>List of abbreviations</b>	<b>xxiv</b>
<b>List of symbols</b>	<b>xxv</b>
<b>1. Chapter 1: Introduction &amp; Objectives</b>	<b>1-6</b>
<b>1.1 Introduction</b>	<b>1</b>
<b>1.2 Objectives</b>	<b>5</b>
<b>2. Chapter 2: Review of Literature</b>	<b>7-42</b>
<b>2.1 Discovery of G-CSF</b>	<b>7-8</b>
<b>2.2 G-CSF and its structure</b>	<b>8-9</b>
<b>2.3 The biological role of G-CSF</b>	<b>10</b>
<b>2.4 Therapeutic applications of G-CSF</b>	<b>11-14</b>
2.4.1 Chemotherapy-induced neutropenia	
2.4.2 Hematopoietic stem and progenitor cell mobilization	
2.4.3 Acute myocardial infarction	
2.4.4 Disorders of the central nervous system	

2.4.5 Treatment of Crohn's disease	
2.4.6 Myelodysplastic syndrome, aplastic anaemia, acquired immune-deficiency syndrome (AIDS)	
<b>2.5 Pharmaceutical analogs of the recombinant human G-CSF in the market</b>	<b>15-16</b>
<b>2.6 Expression platforms for human therapeutics</b>	<b>16-22</b>
2.6.1 Prokaryotic expression system	
2.6.2 Eukaryotic expression system	
<b>2.7 <i>P. pastoris</i> as an expression platform for production of human therapeutics</b>	<b>22-34</b>
2.7.1 Important properties of <i>P. pastoris</i> as an expression platform and available strains	
2.7.2 Methanol utilization phenotype	
2.7.3 Secretion of the expressed proteins in <i>P. pastoris</i>	
2.7.4 Production of G-CSF in <i>Pichia pastoris</i>	
<b>2.8 Detailed structure of the <math>\alpha</math>-MAT secretion signal of <i>S. cerevisiae</i>, its processing and signal peptide engineering</b>	<b>35-39</b>
<b>2.9 Dimorphic transition among yeast species</b>	<b>39-42</b>
2.9.1 Production of heterologous proteins by dimorphic yeast	
<b>3. Chapter 3: To evaluate the role of amino acid segments in the <math>\alpha</math>-MAT secretion signal in secretion of human G-CSF emerging from a wild type or codon optimized copy of the gene</b>	<b>43-82</b>
<b>3.1 Introduction</b>	<b>43-46</b>

## **3.2 Materials**

**47-48**

3.2.1 Plasmids and strains used in this study for exploring the functional expression of G-CSF

3.2.2 Source of the *WT-GCSF* and *CO-GCSF* genes used in the present study

3.2.3 Media and culture conditions

3.2.4 Restriction, modifying enzymes and molecular biology kits

## **3.3 Experimental Methods**

**49-56**

3.3.1 Cloning of the codon optimized (*CO-GCSF*) gene in the pPICZ $\alpha$ B vector

3.3.2 Truncations in the  $\alpha$ -MAT pro-secretory sequences

3.3.3 Transformation of *P. pastoris* with plasmid DNA containing confirmed mutations in the pro-region of secretory signal and establishing their methanol utilisation pattern

3.3.4 High throughput screening of the *P. pastoris* transformants for the production of G-CSF in a 48-well plate

3.3.5 Methanol inducible production of G-CSF by *P. pastoris* transformants at shake flask level

3.3.6 Bioinformatic analyses

## **3.4 Analytical methods**

**56-58**

3.4.1 Measurement of the cell growth

3.4.2 Total protein estimation

3.4.3 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

3.4.4 The quantitation of the G-CSF using gel densitometry

3.4.5 Trypsin digestion and MALDI-TOF/MS

3.4.6 The Bioinformatic analysis

### **3.5 Results**

**58-76**

3.5.1 Cloning of the *WT-GCSF* and the *CO-GCSF* in pPICZ $\alpha$ B vector and verification of deletion mutations in the pro-segment of the  $\alpha$ -MAT signal sequence through sequencing

3.5.2 Truncation in pro-secretory sequences

3.5.3 PCR optimization for site-directed mutagenesis

3.5.4 Transformation in *E. coli* and sequence verification of the deletion constructs of the  $\alpha$ -MAT secretory signal sequence

3.5.5 Electroporation into *P. pastoris*, confirmation of the integration event and screening for methanol utilization pattern

3.5.6 High throughput screening of *P. pastoris* transformants for production of extracellular G-CSF with truncated pro-secretory sequences preceding the *WT-GCSF* gene

3.5.7 Comparative analysis of G-CSF production among *P. pastoris* transformants containing different truncated pro-secretory sequences fused to the *WT-GCSF* gene

3.5.8 High throughput screening of the *P. pastoris* transformants for production of extracellular G-CSF with truncated pro-secretory sequences preceding the *CO-GCSF* gene

3.5.9 Comparative analysis of G-CSF production among the *P. pastoris* transformants containing different truncated pro-secretory sequences fused to the *CO-GCSF* gene

3.5.10 MALDI TOF/MS	
3.5.11 Secondary structures and the predicted 3D model of truncated secretory signal sequence	
<b>3.6 Discussion</b>	<b>76-82</b>
<b>4. Chapter 4: To evaluate the role of small amino acids at the kex2 P1' cleavage site of the <math>\alpha</math>-MAT secretion signal on the yield of extracellular G-CSF in <i>P. pastoris</i></b>	<b>83-123</b>
<b>4.1 Introduction</b>	<b>83-86</b>
<b>4.2 Materials</b>	<b>86-87</b>
4.2.1 Strains, plasmids, and expression vector used for the production of G-CSF	
4.2.2 Cultivation media for the expression of G-CSF	
4.2.3 Enzymes and molecular biology kits	
<b>4.3 Experimental Methods</b>	<b>87-95</b>
4.3.1 Construction of the expression vector "pPICZ $\alpha$ B-CO-GCSF" used in this study	
4.3.2 Modification of the kex2 cleavage site of the $\alpha$ -MAT signal sequence <i>via</i> site-directed mutagenesis	
4.3.3 Colony PCR for the confirmation of positive mutagenic plasmids containing the <i>G-CSF</i> insert	
4.3.4 The DNA sequencing of the mutagenic plasmid constructs containing the desired mutations at the kex2 cleavage site	
4.3.5 Electroporation of the mutagenic plasmids containing the confirmed mutations at the kex2 cleavage site	

4.3.6 Genomic DNA isolation from the recombinant *P. pastoris* and confirmation of gene integration event by PCR

4.3.7 Screening of the *P. pastoris* transformants for the methanol utilization phenotype using PCR

4.3.8 High throughput screening of the transformants in a microtiter plate

4.3.9 Methanol inducible expression of the G-CSF at the shake flask level

4.3.10 Bioinformatic analyses

#### **4.4 Analytical methods 95-97**

4.4.1 Measurement of the cell growth

4.4.2 Total protein estimation

4.4.3 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

4.4.4 The quantitation of the G-CSF using gel densitometry

#### **4.5 Results 97-113**

4.5.1 Replacement of Glutamate (Glu) at the *kex2* cleavage site with Ala, Met and Val in the  $\alpha$ -MAT secretion signal

4.5.2 Screening of the positive *E. coli* transformants containing the mutagenic plasmids and confirmation of the mutations introduced in the *kex2* site

4.5.3 Transformation of *P. pastoris* with plasmid DNA containing confirmed mutations at the 'kex2 site' and establishing their methanol utilisation pattern

4.5.4 Screening of the *P. pastoris* transformants containing the desired mutations at the *kex2* site for the production of G-CSF

4.5.5 Comparative analysis of the Mut <sup>+</sup> kex2 mutants at the shake flask level	
4.5.6 Comparative analysis of the Mut <sup>s</sup> kex2 mutants at shake flask level	
4.5.7 Secondary structure prediction of the mutations introduced at the P1' site of the $\alpha$ -MAT secretion signal	
4.5.8 Predicted 3D models	
<b>4.6 Discussion</b>	<b>114-123</b>
<b>5. Chapter 5: To investigate the influence of nutritional and environmental factors on Yeast-mycelial dimorphism in <i>Pichia pastoris</i></b>	<b>124-142</b>
<b>5.1 Introduction</b>	<b>124-126</b>
<b>5.2 Materials and Methods</b>	<b>126-128</b>
5.2.1 Strains and media used for morphological study	
5.2.2 Cultivation conditions for morphological analysis by SMD1168 strain	
5.2.3 Confocal microscopy	
5.2.4 Gas chromatography-mass spectrometry (GC-MS)	
<b>5.3 Results and Discussion</b>	<b>129-143</b>
5.3.1 Extracellular pH dictates the morphology of <i>P. pastoris</i>	
5.3.2 Effect of different nitrogen sources on morphology of <i>P. pastoris</i>	
5.3.3 Effect of different carbon sources on the morphology of <i>P. pastoris</i>	

5.3.4 Dimorphic transition in <i>P. pastoris</i> is reversible	
5.3.5 Extracellular metabolome profiling by GC-MS	
<b>6. Chapter 6: Conclusions &amp; Future prospects</b>	<b>144-147</b>
<b>6.1 Summary and Conclusions</b>	<b>144-146</b>
<b>6.2 Future prospects</b>	<b>146-147</b>
<b>7. Chapter 7: References</b>	<b>148-181</b>
<b>8. Chapter 8: Annexures</b>	<b>182-206</b>
Annexure-I	<b>182-183</b>
Annexure-II	<b>184</b>
Annexure-III	<b>185-192</b>
Annexure-IV	<b>193</b>
Annexure-V	<b>194</b>
Annexure-VI	<b>195</b>
Annexure-VII	<b>196</b>
Annexure-VIII	<b>197</b>
Annexure-IX	<b>198</b>
Annexure-X	<b>199-206</b>
<b>9. Chapter 9: Curriculum Vitae</b>	<b>207-210</b>

## LIST OF FIGURES

---

Number	Details	Page
<b>Fig 2.1</b>	Ribbon diagram of the recombinant human G-CSF produced in <i>E. coli</i> . Helix A comprises of the residues 11-39, B of 71-91, C of 100-123, and D of 143-172. The short $3^{10}$ helix extends from 44-47 and a single turn helix is located between 48-53 amino acids. Amino acids between residues 93-95, form almost one turn, reminiscent of a left-handed helix (From Hill et al., 1993).	<b>9</b>
<b>Fig 2.2</b>	Schematic representation of hematopoiesis and the colony-stimulating factors (IL, interleukin; FL, Fms-like tyrosine kinase 3 ligand (FLT3/FLK-2 ligand); SCF, stem cell factor; TPO, thrombopoietin; GM-CSF, granulocyte macrophage colony- stimulating factor; G-CSF, granulocyte colony-stimulating factor; M-CSF, macrophage colony-stimulating factor. Colony-stimulating factors stimulate the pluripotent stem cells to give rise to various types of committed progenitor cells to generate mature functional cells (Sanganalmath et al., 2011).	<b>10</b>
<b>Fig 2.3</b>	Schematic representation of the yeast secretory pathway. The correctly folded heterologous protein (eg. G-CSF) is secreted into the extracellular medium with high efficiency whereas the misfolded proteins are either retained into the ER or activate UPR and ERAD pathways.	<b>32</b>
<b>Fig 2.4</b>	Diagrammatic presentation of the $\alpha$ -MAT secretion signal and its processing during secretion.	<b>35</b>
<b>Fig 3.1</b>	Diagrammatic presentation of the $\alpha$ -MAT secretion signal and its processing during secretion.	<b>45</b>

- Fig 3.2** Scheme for cloning of *CO-GCSF* gene in pPICZ $\alpha$ B vector. *CO-GCSF* gene was amplified from pPICZ $\alpha$ B containing the fusion gene *HSADIII-CO-GCSF* (Sigar et al., 2017) using *G-CSF* specific forward and reverse primers containing appropriate restriction sites (See Table 3.1). The restriction enzyme digested (*Xho*I and *Xba*I) vector and the amplified *CO-GCSF* fragment were isolated from the gel, purified and ligated to form pPICZ $\alpha$ B-*CO-GCSF*. **49**
- Fig 3.3** Cloning of the *CO-GCSF* gene in pPICZ $\alpha$ B vector (a) PCR amplification of *CO-GCSF* using gene specific primers. Lane M- DNA ladder (1 kb); Lanes 1 to 4- PCR amplified product (*CO-GCSF*). (b) Double digestion of pPICZ $\alpha$ B and the amplified *CO-GCSF* gene. Lanes M- DNA ladder (1 kb); Lane 1- pPICZ $\alpha$ B vector digested with *Xho*I and *Xba*I; Lane 2- *CO-GCSF* fragment digested with *Xho*I and *Xba*I. (c) Products of digestion of two recombinant plasmids, with *Xho*I and *Xba*I, run on 0.7% agarose; Lane M-DNA ladder (1kb); Lanes 1, 2- double digested recombinant vector, upper band (pPICZ $\alpha$ B), lower band (*CO-GCSF*). **60**
- Fig 3.4** Schematic representation of truncations in the  $\alpha$ -MAT secretory signal. (a) The native  $\alpha$ -MAT consisting of the pre-peptide (19 a.a), followed by the pro-peptide (66a.a). Truncations in the pro-region, carried out in the present study of the  $\alpha$ -MAT secretory signal peptide are shown. The position of mutations is indicated by amino acid numbers. (b) Schematic representation of the MAT $\alpha$ :  $\Delta$ 57-70 fused with the *WT-GCSF*. (c) Schematic representation of the MAT $\alpha$ :  $\Delta$ 57-70 fused with the *CO-GCSF*. **61**
- Fig 3.5** Representative gel image of PCR optimization to carry out truncations in  $\alpha$ -MAT secretion signal. Lane M: Molecular weight ladder; Lanes 1 to 7: 5  $\mu$ l of PCR amplified product “pPICZ $\alpha$ B-*WT-GCSF*” or “pPICZ $\alpha$ B-*CO-GCSF*” (4.1 kb) obtained at different temperatures (52°C to 58°C). **62**
- Fig 3.6** Representative gel image of colony PCR. Lane M: Molecular weight ladder; Lanes 1 to 7: PCR amplified product from different colonies obtained after transformation. **63**

- Fig 3.7** (a) Screening of the methanol utilization phenotype of MAT $\alpha$ :  $\Delta$ 57-70 by PCR. Lane 1: Molecular weight ladder (1 kb); Lanes 2 to 14: PCR amplified product with upper band (2.2 kb) and lower band (1.2 kb) from several randomly picked colonies. (b) Gene insertion event at *AOXI* locus. **65**
- Fig 3.8** (a) SDS-PAGE analysis of the extracellular G-CSF produced in BMMY medium by the truncated  $\alpha$ -MAT fused to the *WT-GCSF* gene. Equal volume (25  $\mu$ l) of the culture filtrate was loaded after 48 h of cultivation in a 48-well plate. Representative image for the MAT $\alpha$ :  $\Delta$ 57-70 transformants is shown. Lanes 1 & 11: Protein ladder; Lane 2: empty vector; Lanes 3, 4, 5, 6, 8, 9, 10: Cl # s 3 to 9; Lane 12: MAT $\alpha$ : WT, Cl # 9; Lanes 13, 14, 15, 17, 18, 19, 20: Cl # s 13 to 19; Lanes 7 & 16: Standard Filgrastim (b) Average extracellular G-CSF (mg/L) production from 15 clones of each category at microplate level 72 h post-methanol induction. **66**
- Fig 3.9** (a) SDS-PAGE analysis of extracellular G-CSF produced by the transformants containing single deletion in the pro-secretory sequence fused to the *WT-GCSF* gene. Lane 1: Protein ladder; Lane 2; empty vector; Lane 3: MAT $\alpha$ : (WT) fused to *WT-GCSF* gene; Lanes 4 & 5: MAT $\alpha$ :  $\Delta$ 57-70, Cl # s 6 & 17; Lanes 6 & 7: MAT $\alpha$ :  $\Delta$ 30-43, Cl # s 10 & 16; Lanes 8 & 9: MAT $\alpha$ :  $\Delta$ 47-49, Cl # s 24, 36; Lane 10: Filgrastim as a standard (1.5  $\mu$ g). (b) SDS-PAGE analysis of G-CSF produced by transformants containing double deletions in the pro-secretory sequence fused to the *WT-GCSF* gene. Lane 11: Protein ladder; Lane 12: MAT $\alpha$ : (WT) fused to the *WT-GCSF* gene; Lanes 13 & 14: MAT $\alpha$ :  $\Delta$ 57-70, Cl # s 6, 17; Lanes 15 & 16: MAT $\alpha$ :  $\Delta$ 57-70;47-49, Cl # s 22, 25; Lanes 17 & 18: MAT $\alpha$ :  $\Delta$ 57-70;30-43, Cl # s 21, 34; Lane 19: Filgrastim (1.5  $\mu$ g). (c) Cell O.D, total extracellular protein and G-CSF titre at 120 h post-methanol induction. **68**
- Fig 3.10** (a) SDS-PAGE analysis of extracellular G-CSF produced in BMMY medium by the truncated  $\alpha$ -MAT fused to the *CO-GCSF*. Equal volume (25  $\mu$ l) of the culture filtrate was loaded after 48 h of cultivation in a 48-well plate. Representative image of MAT $\alpha$ :  $\Delta$ 57-70 transformants is **70**

shown. Lanes 1 & 11: Protein ladder; Lane 2: empty vector; Lane 3: MAT $\alpha$ : (WT) fused to the *CO-GCSF* gene (Cl # 6); Lanes 4 to 9: Cl # s 22, 23, 24, 25, 26, 27; Lanes 10 & 17: Standard Filgrastim (1.5  $\mu$ g & 4.5  $\mu$ g); Lanes 12 to 19: Cl # s 28, 29, 30, 31, 32, 34, 35. (b) Average extracellular G-CSF (mg/L) production from 15 clones of each category at microplate level, 72 h post-methanol induction.

- Fig 3.11** (a) SDS-PAGE analysis of extracellular G-CSF produced by transformants containing single deletions in the pro-secretory sequence fused to the *CO-GCSF* gene; Lane 1: Molecular weight ladder; Lane 2; empty vector; Lane 3: MAT $\alpha$ : (WT) fused to the *CO-GCSF* gene; Lanes 4 & 5: MAT $\alpha$ :  $\Delta$ 57-70, Cl # s 25, 32; Lanes 6 & 7: MAT $\alpha$ :  $\Delta$ 30-43, Cl # s 12, 32; Lane 8: Filgrastim (1.5  $\mu$ g); Lanes 9 & 10: MAT $\alpha$ :  $\Delta$ 47-49, Cl # s 44, 45. (b) SDS-PAGE analysis of G-CSF produced by transformants containing double deletions in the pro-secretory sequence fused to the *CO-GCSF* gene; Lane 11: Molecular weight ladder; Lane 12; Filgrastim (1.5 $\mu$ g); Lane 13: MAT $\alpha$ : (WT) fused to the *CO-GCSF* gene; Lanes 14 &15: MAT $\alpha$ :  $\Delta$ 57-70, Cl # s 25 & 32; Lanes 16 &17: MAT $\alpha$ :  $\Delta$ 57-70;47-49, Cl # s 17, 18; Lanes 18 & 19: MAT $\alpha$ :  $\Delta$ 57-70;30-43, Cl # s 23, 33. (c) Cell O.D, total extracellular protein and G-CSF titre at 120 h post-methanol induction. **72**
- Fig 3.12** Peptide fragments identified using Mascot search software in MALDI TOF analysis (Courtesy: Anshul Sharma, IIT Delhi, M. Tech thesis, 2014). **73**
- Fig 3.13** Secondary structure prediction of  $\alpha$ -MAT mutants (a) Wild type  $\alpha$ - mating factor, MAT $\alpha$ : (WT) (b) MAT $\alpha$ :  $\Delta$ 57-70 (c) MAT $\alpha$ :  $\Delta$ 30-43 (d) MAT $\alpha$ : $\Delta$  47-49 (e) MAT $\alpha$ :  $\Delta$ 57-70;47-49 (f) MAT $\alpha$ :  $\Delta$ 57-70;30-43. Helix (H),  $\beta$ -Strand (S), Coil (C). **74**
- Fig 3.14** Predicted 3-D models of (a) The wild type  $\alpha$ - mating factor with 89 amino acids, MAT $\alpha$ : (WT) (b) MAT $\alpha$ :  $\Delta$ 57-70 with 75 amino acids (c) MAT $\alpha$ :  $\Delta$ 30-43 with 75 amino acids (d) MAT $\alpha$ :  $\Delta$ 47-49 with 86 amino acids (e) **76**

MAT $\alpha$ :  $\Delta$ 57-70;47-49 with 72 amino acids and (f) MAT $\alpha$ :  $\Delta$ 57-70;30-43 with 61 amino acids.

- Fig 3.15** Comparative analysis of extracellular G-CSF produced in *P. pastoris* transformants containing truncated  $\alpha$ -MAT mutants preceding either the *WT-GCSF* gene or the *CO-GCSF* gene. **79**
- Fig 4.1** The flow diagram for screening of the best transformants for production of the G-CSF. a: Site-directed mutagenesis at the P1' site of the *kex2* cleavage site fused to the *CO-GCSF* gene; b: Transformation of the mutagenic plasmid constructs in the *E. coli* (DH5 $\alpha$  cells) c: Sequence confirmation of the mutant plasmids, linearization of the plasmid followed by d: The electroporation in the X-33 *P. pastoris*; e: Screening of the transformants for the integration of the *CO-GCSF* gene and determination of the Mut phenotype (Mut<sup>+</sup> or Mut<sup>S</sup>); f: Small-scale cultivation of the transformants in the 48-well plates; g: Shake flask cultivation of the selected clones. **93**
- Fig 4.2** Schematic representation of expression of G-CSF at shake flask level (Standard Invitrogen protocol). **94**
- Fig 4.3** The diagrammatic representation of substitution of Glu (E) by three different amino acids (Val or V, Ala or A & Met or M) at the *kex2* P1' site and the deletion of the Glu-Ala repeat. A. The native  $\alpha$ -MAT secretory signal fused to the *CO-GCSF* was used as a control. B. The mutated  $\alpha$ -MAT secretory signal sequences (Substitution of Glu by Val/Ala/Met) fused to the *CO-GCSF*. The *kex2* protease site is shown as a dotted line. **97**
- Fig 4.4** A representative gel image of PCR optimization to form MAT $\alpha$ : Glu/Val. Lane M: Marker; Lanes 1 to 17: 5  $\mu$ l of the PCR product. **99**
- Fig 4.5** A representative gel image of colony PCR for screening of the *E. coli* transformants from the plasmid pPICZ $\alpha$ B- $\alpha$ MAT (Val)-*CO-GCSF*. Lane **100**

M: Molecular weight marker (1 kb); Lanes 1 to 14: 10  $\mu$ l of the PCR product obtained from colony PCR of randomly picked *E. coli* colonies.

- Fig 4.6** A representative gel image of screening of the methanol utilization phenotype by PCR. Lane M: Molecular weight ladder; Lane 1: MAT $\alpha$ : Glu/Val transformants, Cl # 26 (Mut<sup>s</sup>); Lane 2 & 4: MAT $\alpha$ : Glu/Val transformants, Cl # s 24, 25 (Mut<sup>+</sup>); Lane 5 & 9: MAT $\alpha$ : Glu/Ala transformants, Cl # s 3, 4 (Mut<sup>+</sup>); Lane 6: MAT $\alpha$ : Glu/Ala transformants, Cl # 2 (Mut<sup>s</sup>); Lane 8: MAT $\alpha$ : Glu/Met transformants, Cl # 15 (Mut<sup>s</sup>); Lanes 10 & 12: MAT $\alpha$ : Glu/Met transformants, Cl # s 2 & 4 (Mut<sup>+</sup>) respectively. **102**
- Fig 4.7** (a) Gene insertion at the *AOXI* locus by single cross-over event. (b) Gene replacement event at *AOXI* locus by double cross over. **103**
- Fig 4.8** (A) SDS-PAGE analysis of extracellular G-CSF production by different transformants obtained from three category of mutants. Equal volumes (25  $\mu$ l) were loaded after 48 h of cultivation in a 48-well plate. (a) MAT $\alpha$ : Glu/Val:- Lanes 1 & 11: Protein ladder; Lane 2: empty vector; Lane 3: MAT $\alpha$ : (WT); Lanes 4 to 8, 12, 13, 15 to 20: Cl # s 21 to 34 (Mut<sup>+</sup>); Lane 9: Cl # 26 (Mut<sup>s</sup>); Lanes 10 & 14: Filgrastim (3.0  $\mu$ g and 1.5  $\mu$ g) respectively (b) MAT $\alpha$ : Glu/Ala:- Lane 1 & 11: Protein ladder; Lane 2: empty vector; Lane 3: MAT $\alpha$ : (WT); Lanes 4: Cl # 1 (Mut<sup>+</sup>); Lane 5: Cl # 2 (Mut<sup>s</sup>); Lanes 6 to 9, 12 to 15, 17 to 20: Cl # s 3 to 15 (Mut<sup>+</sup>); Lanes 10 & 16: Filgrastim (2.5  $\mu$ g and 1.5  $\mu$ g) respectively (c) MAT $\alpha$ : Glu/Met:- Lanes 1 & 11: Protein ladder; Lanes 2 & 19: Filgrastim (2.5  $\mu$ g and 1.5  $\mu$ g) respectively; Lanes 3 & 13: MAT $\alpha$ : (WT); Lanes 4 & 5: Cl # s 4 & 2 (Mut<sup>+</sup>); Lane 6: Cl # 15 (Mut<sup>s</sup>); Lanes 7 to 10: Cl # s 36, 5, 6, 7 (Mut<sup>+</sup>); Lanes 14 to 18 & 20: Cl # s 1, 16, 35, 33, 30, 8 (Mut<sup>+</sup>) (B) Average G-CSF (mg/L) production in Mut<sup>+</sup> as measured from 15 clones of each category (substitution at the P1' site) after 72 h post methanol induction. **106**

- Fig 4.9** (A) Cell O.D, total extracellular protein and G-CSF titre at 120 h post methanol induction. (B) SDS-PAGE analysis of the extracellular culture filtrate produced by the Mut<sup>+</sup> kex2 mutants at shake flask level. Lane 1: Protein ladder; Lane 2; empty vector; Lane 3: MAT $\alpha$ : (WT); Lanes 4 & 5: MAT $\alpha$ : Glu/Val, Cl # s 24, 25; Lanes 6 & 7: MAT $\alpha$ : Glu/Ala Cl # s 6, 13; Lanes 8 & 9: MAT $\alpha$ : Glu/Met, Cl # s 36, 1; Lane 10: Standard Filgrastim (1.5  $\mu$ g). **109**
- Fig 4.10** (A) Cell O.D, total extracellular protein and G-CSF titre at 120 h post methanol induction (B) SDS-PAGE analysis of the culture filtrate produced by Mut<sup>s</sup> kex2 mutants at shake flask level. Lane 1: Molecular weight ladder; Lane 2; empty vector; Lane 3: MAT $\alpha$  (WT); Lanes 4 & 5: MAT $\alpha$ : Glu/Val, Cl # 26; Lanes 6 & 7: MAT $\alpha$ : Glu/Ala Cl # 2; Lanes 8 & 9: MAT $\alpha$ : Glu/Met, Cl # 15; Lane 10: Standard Filgrastim (1.5  $\mu$ g). **111**
- Fig 4.11** Secondary structure prediction of the MAT $\alpha$ : kex2 P1' mutants (a) Wild type  $\alpha$ - mating factor with Glu at the kex2 site (b) Glu at the kex2 cleavage site replaced by Val (c) Glu at the kex2 cleavage site replaced by Ala (d) Glu at the kex2 cleavage site replaced by Met. **112**
- Fig 4.12** Predicted 3-D models of (a) the wild type  $\alpha$ - mating factor ( $\alpha$ -MAT) (b) Val replacing Glu at the kex 2 P1' site (c) Ala replacing the kex 2 P1' site (d) Met replacing the kex 2 P1' site. **113**
- Fig 4.13** Interaction of the cargo protein (G-CSF) with the native signal peptide and its variants, Green colored ribbon structure is that of the cargo protein while red stick representation is for the  $\alpha$ -MAT secretion signal. G-CSF fused to the (a) MAT $\alpha$ : (WT) (b) MAT $\alpha$ : Glu/Val (c) MAT $\alpha$ : Glu/Met or (d) MAT $\alpha$ : Glu/Ala. **120**
- Fig 5.1** Representative images of *P. pastoris* SMD1168 cell morphology on YCB plus 0.2% ammonium sulphate at different pH (3.0 to 8.0). The cells were **131**

sampled after 72 h growth on liquid media and photographed at 1000 x magnification.

- Fig 5.2** Representative images of *P. pastoris* SMD1168 cell morphology on YCB plus 0.2% of different amino acids at pH 6.0. The cells were sampled after 72 h growth on liquid media and photographed at 1000 x magnification. Inoculum was 100% yeast-like morphology. **132**
- Fig 5.3** Representative images of *P. pastoris* SMD1168 cell morphology on YNB plus 0.2% of different carbon sources. The cells were sampled after 72h growth in liquid media and photographed at 1000 x magnification. Inoculum was 100% yeast-like morphology. **134**
- Fig 5.4** Reversible pseudo-hyphae formation in *P. pastoris* during therapeutic protein production. Morphological transition was observed from 24 h to 72 h with 1% methanol induction after every 24 h. **137**
- Fig 5.5** (A) GC-MS chromatogram of cell-free culture filtrate of recombinant *P. pastoris* SMD1168. Chromatograms of (i) cells exhibiting yeast like morphology (ii) cells exhibiting hyphae-like morphology. Cell-free culture filtrate was lyophilized, extracted with benzene and derivatized by N-Methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) and Trimethylsilyl (TMS) prior to GC-MS analysis using GC-2010 coupled with a GC-MS QP-2010 Ultra (Shimadzu, Japan). (B) Relative percentage of extracellular metabolites produced by *P. pastoris* exhibiting yeast-like morphology or hyphae-like morphology. The quantity of each metabolite was calculated in terms of relative percentage by measuring the area under peak of each metabolite in relation to the total area of all other metabolites detected in the extract. **138**
- Fig 8.1** (a) Codon-optimized copy of G-CSF (*CO-GCSF*, 564 bp) fused to native  $\alpha$ -MAT secretory signal sequences; (b) Native  $\alpha$ -MAT secretory signal sequences (267 bp); *kex2* P1' site and Glu-Ala repeats are highlighted in red. **200**

<b>Fig 8.2</b>	(a) 47-49 amino acid sequences highlighted in the native $\alpha$ -MAT secretory signal sequences, (b) Deletion of ‘amino acids 47-49’ ( $\Delta$ 47-49) in the native $\alpha$ -MAT secretory signal sequences fused to <i>WT-GCSF</i> as well as <i>CO-GCSF</i> .	<b>201</b>
<b>Fig 8.3</b>	(a) 57-70 amino acid sequences highlighted in the native $\alpha$ -MAT secretory signal sequences, (b) Deletion of ‘amino acids 57-70’ ( $\Delta$ 57-70) in the native $\alpha$ -MAT secretory signal sequences fused to <i>WT-GCSF</i> as well as <i>CO-GCSF</i> .	<b>202</b>
<b>Fig 8.4</b>	(a) 30-43 amino acid sequences highlighted in the native $\alpha$ -MAT secretory signal sequences, (b) Deletion of ‘amino acids 30-43’ ( $\Delta$ 30-43) in the native $\alpha$ -MAT secretory signal sequences fused to <i>WT-GCSF</i> as well as <i>CO-GCSF</i> .	<b>202</b>
<b>Fig 8.5</b>	(a) 57-70;47-49 amino acid sequences highlighted in the native $\alpha$ -MAT secretory signal sequences, (b) Deletion of ‘amino acids 57-70 & 47-49’ ( $\Delta$ 57-70;47-49) in the native $\alpha$ -MAT secretory signal sequences fused to <i>WT-GCSF</i> as well as <i>CO-GCSF</i> .	<b>203</b>
<b>Fig 8.6</b>	(a) 57-70;30-43 amino acid sequences highlighted in the native $\alpha$ -MAT secretory signal sequences, (b) Deletion of ‘amino acids 57-70 & 30-43’ ( $\Delta$ 57-70;30-43) in the native $\alpha$ -MAT secretory signal sequences fused to <i>WT-GCSF</i> as well as <i>CO-GCSF</i> .	<b>204</b>
<b>Fig 8.7</b>	(a) Native $\alpha$ -MAT secretory signal sequence fused to <i>CO-GCSF</i> with Glu at kex2 P1’ site followed by Glu-Ala repeats, (b) Replacement of Glu by Val at kex2 P1’ site followed by deletion of Glu-Ala repeats in $\alpha$ -MAT fused to <i>CO-GCSF</i> , (c) Replacement of Glu by Met at kex2 P1’ site followed by deletion of Glu-Ala repeats in $\alpha$ -MAT fused to <i>CO-GCSF</i> (d) Replacement of Glu by Ala at kex2 P1’ site followed by deletion of Glu-Ala repeats in $\alpha$ -MAT fused to <i>CO-GCSF</i> .	<b>205</b>

---

## LIST OF TABLES

Number	Details	Page
<b>Table 2.1</b>	A summary of reported literature on production of G-CSF in prokaryotic systems	<b>18</b>
<b>Table 2.2</b>	A summary of reported literature on the production of G-CSF in eukaryotic system other than yeasts	<b>21</b>
<b>Table 2.3</b>	<i>Pichia pastoris</i> expression strains	<b>27</b>
<b>Table 2.4</b>	Expression of therapeutic proteins in <i>P. pastoris</i> at shake flask and bioreactor level	<b>28</b>
<b>Table 2.5</b>	Different secretion signals used to express heterologous proteins in <i>P. pastoris</i>	<b>30</b>
<b>Table 2.6</b>	Secreted expression of different heterologous proteins using $\alpha$ -MAT secretion signal	<b>31</b>
<b>Table 2.7</b>	A summary of reported literature on production of G-CSF in <i>P. pastoris</i>	<b>34</b>
<b>Table 3.1</b>	The primers used for gene amplification and the sequencing of the <i>CO-GCSF</i> construct	<b>49</b>
<b>Table 3.2</b>	The PCR reaction components for the amplification of the <i>CO-GCSF</i> gene	<b>50</b>
<b>Table 3.3</b>	The ligation of the double digested vector (pPICZ $\alpha$ B vector) and ( <i>CO-GCSF</i> )	<b>51</b>
<b>Table 3.4</b>	Mutagenic primers used to carry out truncations in the $\alpha$ -MAT secretory signal	<b>52</b>
<b>Table 3.5</b>	Optimization parameters for different truncations in $\alpha$ -MAT	<b>62</b>
<b>Table 3.6</b>	List of clones selected from each category of mutants after 72 h post methanol induction	<b>67</b>
<b>Table 3.7</b>	Helix, strand and coil percentage of truncated $\alpha$ - mating factor	<b>75</b>
<b>Table 4.1</b>	The mutagenic primers used to modify the <i>kex2</i> cleavage site of the $\alpha$ -MAT secretory signal	<b>88</b>
<b>Table 4.2</b>	The PCR reaction mixture for the colony PCR	<b>89</b>
<b>Table 4.3</b>	The primers used for amplification of the signal sequence	<b>90</b>

<b>Table 4.4</b>	The typical reaction mixture for the linearization of the circular plasmid	<b>90</b>
<b>Table 4.5</b>	The components of the PCR reaction mixture	<b>92</b>
<b>Table 4.6</b>	List of transformants based on the methanol utilization phenotype	<b>104</b>
<b>Table 4.7</b>	Average G-CSF (mg/L) titre from 15 clones from each category (substitution at the P1' site), 72 h post methanol induction	<b>107</b>
<b>Table 4.8</b>	Comparative analysis of extracellular G-CSF production in the Mut <sup>+</sup> and the Mut <sup>s</sup> <i>kex2</i> cleavage site mutants (Shake flask studies)	<b>118</b>
<b>Table 5.1</b>	<i>P. pastoris</i> SMD1168 cell morphology under different nutrient deficient media	<b>135</b>
<b>Table 5.2</b>	Extracellular metabolome among two morphological forms of SMD1168 strain of <i>P. pastoris</i> by GC-MS	<b>141</b>
<b>Table 8.1</b>	The double digestion of the empty vector (pPICZ $\alpha$ B vector) and the insert ( <i>CO-GCSF</i> )	<b>192</b>

## List of Abbreviations

Abbreviations	Full forms
APS	Ammonium per sulphate
BME	$\beta$ -Mercaptoethanol
BSA	Bovine serum albumin
BMGY	Buffered complex glycerol medium
BMMY	Buffered complex methanol medium
bp	Base pair
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
MALDI	Matrix assisted laser desorption ionization
PCR	Polymerase chain reaction
PDB	Protein data base
SDS	Sodium dodecyl sulphate
TEMED	N,N,N',N'-Tetramethylethylenediamine
UV	Ultra violet

## List of symbols

Symbols	Meanings
~	Approximately
bp	base pair
°C	Degree Celsius
μM	Micromolar
μg	Microgram
μl	Microliter
g	Gram
h	hours
Kb	Kilobase pair
kDa	Kilodalton
mg	Milligram
min	Minute
ml	Millilitre
mM	Millimolar
ng	Nanogram
Nm	Nano meter