GRANULOCYTE-COLONY STIMULATING FACTOR ANALOG PRODUCTION BY RECOMBINANT *Escherichia coli*

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ONUR ERSOY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOTECHNOLOGY

SEPTEMBER 2017

Approval of the thesis:

GRANULOCYTE-COLONY STIMULATING FACTOR ANALOG PRODUCTION BY RECOMBINANT *Escherichia coli*

submitted by **ONUR ERSOY** in partial fulfillment of the requirements for the degree of **Master of Science in Biotechnology Department, Middle East Technical University by,**

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Assoc. Prof. Dr. Çağdaş Devrim Son
Head of Department, **Biotechnology** _____

Prof. Dr. Pınar Çalık
Supervisor, **Chemical Engineering Dept., METU** _____

Prof. Dr. Hasan Ongun Onaran
Co-supervisor, **Medical Pharmacology Dept., Ankara Uni.** _____

Examining Committee Members:

Assoc. Prof. Dr. Çağdaş Devrim Son
Biology Dept., METU _____

Prof. Dr. Pınar Çalık
Chemical Engineering Dept., METU _____

Prof. Dr. Hasan Ongun Onaran
Medical Pharmacology Dept., Ankara Uni. _____

Asst. Prof. Dr. Erhan Bat
Chemical Engineering Dept., METU _____

Asst. Prof. Dr. Harun Koku
Chemical Engineering Dept., METU _____

Date: *15.09.2017*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name :

Signature :

ABSTRACT

GRANULOCYTE-COLONY STIMULATING FACTOR ANALOG PRODUCTION BY RECOMBINANT *Escherichia coli*

Ersoy, Onur
M.Sc., Department of Biotechnology
Supersivor : Prof. Dr.Pınar Çalık
Co-Supervisor : Prof Dr. Hasan Ongun Onaran

September 2017, 90 pages

Granulocyte-colony stimulating factor (G-CSF) is a significant pharmaceutical biologic. It is predominantly produced using *Escherichia coli* in the industry. In this study, an *E. coli* expression strain was genetically engineered to heterologously express a His-tagged mutant G-CSF with one amino acid substitution with respect to native human protein. A fed-batch strategy for this strain was developed. Biomass concentration in 7.0 L bioreactor reached a density of 58 grams of dry cell weight per liter. Recombinant protein concentration of bioreactor culture reached 4.9 grams per liter. G-CSF C17A was purified from cell culture, and was shown to have comparable bioactivity to innovator product, *in vitro*. Purified G-CSF C17A was PEGylated to achieve a bioproduct with enhanced properties. A 37% conversion of G-CSF C17A to PEG-G-CSF 17A was achieved in PEGylation reaction.

Keywords: Granulocyte colony stimulating factor; *Escherichia coli*; Rational protein stabilization; fed-batch bioreactor production; protein purification; PEGylation of protein; bioactivity assay.

ÖZ

REKOMBİNANT *Escherichia coli* İLE GRANÜLOSİT-KOLONİ UYARICI FAKTÖR ANALOGU ÜRETİMİ

Ersoy, Onur
Yüksek Lisans, Biyoteknoloji Bölümü
Tez Yöneticisi: Prof. Dr.Pınar Çalık
Ortak Tez Yöntecisi: Prof Dr. Hasan Ongun Onaran

Eylül 2017, 90 sayfa

Granülosit koloni uyarıcı faktör (G-CSF) önemli bir biyolojik famasötik üründür. G-CSF, Endüstride ağırlıklı olarak *Escherichia coli*'de rekombinant olarak üretilmektedir. Bu çalışmada bir *E. coli* ekspresyon suşu doğal bir insan proteinine göre bir amino asit değişikliği olan His-etiketli mutant G-CSF'yi heterolog olarak ifade etmek üzere genetik olarak tasarlanmıştır. Bu suş 7.0 L biyoreaktörde litre başına 58 gram kuru hücre ağırlığına yoğunlaştırılmıştır. Biyoreaktör kültürünün rekombinant protein konsantrasyonu litre başına 4.9 grama ulaşmıştır. G-CSF C17A, hücre kültüründen ayrıştırılıp ve saflaştırılmış, *in vitro* olarak ticari ürünle karşılaştırılabilir biyoaktiviteye sahip olduğu gösterilmiştir. Saflaştırılmış G-CSF C17A, gelişmiş özelliklere sahip bir biyo-ürün elde etmek üzere PEGile edilmiştir. PEGilasyon reaksiyonunda % 37 dönüşüm sağlanmıştur.

Anahtar kelimeler: Granülosit koloni uyarıcı faktörü; *Escherichia coli*; Rasyonel protein stabilizasyonu; yarı-kesikli biyoreaktör üretimi; protein saflaştırma; protein PEGilasyonu; bioaktivite ölçümü.

For İsmaile,

in memoriam

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Professor Pınar Çalık, for putting her invaluable trust in me. Without her, this thesis would not be possible.

I would also like to express my gratitude for my co-advisor, Professor Ongun Onaran, for allowing the activity assay experiments to be performed in his lab in Ankara University, and providing valuable insight on pharmacological aspect of G-CSF.

I would like to thank Assistant Professor Erhan Bat for his guidance and suggestions on PEGylation studies.

I would like to thank my current and former labmates in the Industrial Biotechnology and Metabolic Engineering Laboratory: Erdem Boy, Burcu Gündüz Ergün, Damla Hüccetoğulları, Dr. Aslan Massahi, Sibel Öztürk, Özge Kalender, Duygu Yalçınkaya Bebeta Hoxha, Özge Ata, Abdullah Keskin and Hande Güneş.

I would like to especially thank Yiğit Akgün and Begüm Akcan, for pursuig a research interest in separation of G-CSF C17A in their M.Sc. thesis works.

I am grateful to academic, administrative and technical staff of Department of Biotechnology and Department of Chemical Engineering for their support through my education.

TABLE OF CONTENTS

ABSTRACT v
ÖZ vi
ACKNOWLEDGMENTS viii
LIST OF TABLES xiii
LIST OF FIGURES xiv
NOMENCLATURE xvi
CHAPTERS
1. INTRODUCTION 1
2. LITERATURE SURVEY 3
2.1 Granulocyte Colony Stimulating Factor (G-CSF) 3
2.1.1 Cys17 As a Rational Protein Stabilization Target 5
2.2 Host Microorganism 6
2.2.1 *Escherichia coli* 6
2.2.1.2 *Escherichia coli* DH5α 7
2.2.1.2 *Escherichia coli* BL21 (DE3) 7
2.3 Choice of Expression Plasmid 7
2.3.1 The Lac Operon 7
2.3.2 PRSET A Vector 8
2.4 Inclusion body formation 9
2.5 Genetic Engineering Methods 9
2.5.1 Codon Optimization 9
2.5.2 Recombinant DNA Technology 11
2.6 Bioreactor Operation Conditions 11
2.6.1 Carbon Source for Fermentation 11
2.6.2 Temperature, pH and Dissolved Oxygen 12

2.6.3 Bioprocess Kinetics in Fed-Batch Bioreactor 12
2.6.3.1 Mass balance for biomass 12
2.6.3.2 Mass balance for substrate 14
2.6.3.3 Mass balance for product 15
2.7 Separation 17
2.8 In vitro bioactivity 18
3. MATERIALS AND METHODS 21
3.1 Chemicals 21
3.2 Buffers and Stock Solutions 21
3.3 Strains and Plasmids 21
3.4 Genetic Engineering Methods 22
3.4.1 Restriction endonuclease digestion 22
3.4.2 Agarose Gel Electrophoresis 23
3.4.3 DNA Extraction From Agarose Gel 24
3.4.4 DNA Purification 24
3.4.5 Ligation Reaction 24
3.4.6 Plasmid Isolation 25
3.4.7 Transformation 25
3.5 Growth Media 25
3.5.1 Solid Medium 25
3.5.2 Liquid Media 25
3.6 Cell Cultivation and Recombinant Protein Production 27
3.7 Protein Purification 28
3.7.1 Cell Lysis 28
3.7.2 Inclusion Body Solubilization 29

3.7.3 Cobalt Affinity Chromatography 29
3.7.4 Dialysis 30
3.7.5 Concentration 30
3.7.6 Factor Xa Digestion 30
3.8 *In vitro* Bioactivity Assay 30
3.9 PEGylation 31
3.10 Analysis 32
3.10.1 Cell concentration 32
3.10.2 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) .. 32
3.10.3 Western Blotting 33
4. RESULTS AND DISCUSSION 35
4.1 Strain Development 35
4.1.1 Synthetic Gene Construct 37
4.1.2 Expression vector construction 38
4.1.3 Transformation of cloning strain 39
4.1.4 Transformation of expression strain 40
4.1.5 Verification Steps 41
4.2 Recombinant Protein Production 43
4.2.1 Shake-flask Experiments 43
4.2.2 Bioreactor Experiments 44
4.3 Separation of G-CSF C17A 53
4.4 In vitro bioactivity assay of G-CSF C17A 56
4.5 PEGylation of G-CSF C17A 57
5. CONCLUSION 59
REFERENCES 61

APPENDICES

A. BUFFERS AND STOCK SOLUTIONS 69
B. COLUMN PURIFICATION PROTOCOL 71
C. DIALYSIS TUBING PREPARATION PROTOCOL 73
D. PROTOCOLS FOR FACTOR XA DIGESTION AND FACTOR XA REMOVAL 75
E. OPTICAL DENSITY CALIBRATION CURVES 77
F. NUCLEOTIDE SEQUENCES 79
G. SDS-PAGE GELS 85
H. BIOACTIVITY DATA 89

LIST OF TABLES

Table 2.1: Factory prices for pharmaceutical G-CSF products available in Turkish market.....4

Table 3.1: Strains and Plasmids Used in This Study.....22

Table 3.2: Double digestion reaction of gene fragment.....22

Table 3.3: Double digestion reaction of pRSET A.....23

Table 3.4: Ligation mixture.....24

Table 3.5: Chemically defined batch and feeding media for bioreactor experiments 26

Table 3.6: Reaction mixture for PEGylation of GCSF C17A. 31

Table 4.1: Media modifications done while adapting Korz's medium to this study.....45

Table 4.2: Adopted constant values for exponential fed-batch bioreactor experiments.....47

Table 4.3: Experimental specific growth rate, rate, specific production rate and yield values for BR-1.....51

Table 4.4: Experimental specific growth rate, rate, specific production rate and yield values for BR-2.....52

LIST OF FIGURES

Figure 2.1 : Tertiary structure of G-CSF. 3
Figure 2.2: Primary structure of recombinant human G-CSF 5
Figure 2.3: Lac operon scheme. 8
Figure 2.4: Vector map of 2897 bp long pRSET A. 8
Figure 2.5: Formation of leptin IBs in *E. coli* 9
Figure 2.6: Codon bias for *E. coli* B strain 10
Figure 2.7: 96-well plate containing samples from M-NFS-60 cells grown in various G-CSF concentrations 18
Figure 2.8: Space filling model of PEG-G-CSF. 19
Figure 3.1: Illustration of employed 7.0-L bioreactor system 28
Figure 4.1: Flowchart for strain development 36
Figure 4.2: Double digested pRSET A plasmid against Lambda DNA/HindIII Marker on 1.5% agarose gel. 38
Figure 4.3A: Undigested plasmid isolations from transformants of *E. coli* DH5α PRSET A::gcsf c17a. 39
Figure 4.3B: Double digestions of plasmids from selected samples 49
Figure 4.4: Amplified PCR products of selected colonies.. 40
Figure 4.5A: Undigested and digested plasmid isolations from transformants of *E. coli* BL21 (DE3) pRSETA::gcsf c17a. 41
Figure 4.5B: PCR product of gene amplicate from colony 9 41
Figure 4.6: Western Blot image using Anti-G-CSF antibody. 42
Figure 4.7: Growth curve for glucose shaker experiment. 43
Figure 4.8: Growth curves for glucose vs glycerol shake-flask experiment. 44
Figure 4.9: Process Flowchart for BR-2 46
Figure 4.10: Feeding flow rate for bioreactor experiments; 47
Figure 4.11: Cell concentration profiles in bioreactor experiments. 48
Figure 4.12: Variaton of 6xHistag-G-CSF C17A concentration with fed-batch cultivation time 49
Figure 4.13: Variation of biomass formation rates in bioreactor experiments 50

Figure 4.14: Variation of product formation rates in bioreactor experiments 50
Figure 4.15: Flowchart for Separation of G-CSF C17A. 53
Figure 4.16: SDS-PAGE for separation. 54
Figure 4.17: 15% SDS-PAGE gel for checking disulfide bridges. 55
Figure 4.18: Activity data at the end of the assay (t=96h).. 56
Figure 4.19: PEGylation reaction mixture on gel. 57
Figure E.1: Calibration curve for dry cell weight *E. coli* BL21 (DE3) pRSET A:: GCSF C17A cells at 550 nm.....77
Figure E.2: Calibration curve for dry cell weight *E. coli* BL21 (DE3) pRSET A:: GCSF C17A cells at 600 nm.....78
Figure F.1: Legend for nucleotide sequence. 79
Figure F.2: Sequencing Data for G-CSF C17A gene cloned into pRSET A vector. . 84
Figure G.1: SDS-PAGE gel for BR-1. 85
Figure G.2: SDS-PAGE gel for BR-1. 86
Figure G.3: SDS-PAGE gel for BR-2. 86
Figure G.4: SDS-PAGE gel for BR-2. 87
Figure H.1: Dose and time response of Neupogen®. 89
Figure H.2: Dose and time response of G-CSF C17A. 90

NOMENCLATURE

C	Concentration	$g\ L^{-1}$ or $mol\ L^{-1}$
DO	Dissolved oxygen	%
N	Agitation rate	min^{-1}
r	Rate of formation or consumption	$g\ L^{-1}\ h^{-1}$
Y	Yield	$g\ g^{-1}$
t	Cultivation time	h

Greek Letters

μ	Specific growth rate	h^{-1}
μ_{exp}	Experimental specific growth rate	h^{-1}

Subscripts

DO	Refers to dissolved oxygen
P	Refers to product
P/S	Refers to product over substrate
P/X	Refers to product over biomass
S	Refers to substrate
X	Refers to biomass
X/S	Refers to biomass over substrate

Abbreviations

bp	Base pair.
DMSO	Dimethyl sulfoxide.
G-CSF	Granulocyte-colony stimulant factor.
IPTG	Isopropyl β-D-1-thiogalactopyranoside.
LSLB	Low salt lysogeny broth.
MTT	3-(3,5-dimethylythiazol-2-yl)-2,5-diphenyltetrazolium bromide.
PEG	Polyethylene glycol.
SDS-PAGE	Sodium dodecyl sulfate polayacrylamide gel electrophoresis.

CHAPTER 1

INTRODUCTION

The origins of biotechnology can be traced back to the dawn of civilization. Fermentation of grain and fruit juices with yeast provided ancient people a way to conserve excess crops by turning it into beer and wine. It was millenia later for scientists to figure out another reason to grow microorganisms: to fight disease. Within the last century, since the conceptualization of the magic bullet, science has been looking to discover or to invent new molecules for diseases. Some of these drugs are synthesized chemically while an important part is being produced by living microorganisms.

Recombinant DNA technology is an essential tool for producing therapeutic proteins. Over the last 40 years thousands of different proteins have been successfully expressed in bacterial, yeast, animal, insect cell lines and in transgenic animals. One of these proteins is granulocyte-colony stimulating factor(G-CSF). G-CSF is a hematopoiteic factor responsible for the profileration and maturation of granulocytes. Chemotherapy induced neutropenia, congenital neutropenia and febrile neutropenia are the chief medical indications of G-CSF. Moreover, G-CSF is also an important stem-cell mobilization agent used in medicine.

Wild type human G-CSF is 175 amino acids long with a single glycosylation site at the threonine residue at the 133rd position. This glycosylation is not essential for activity. After G-CSF's discovery and establishment of function in late 1980s, the original bacterial production method was patented by Amgen in 1991 under the trade

name Neupogen® (Souza, 1991). Since the patent's expiry, it has been produced and commercialized by different companies around the world. G-CSF is also an economically significant product in the sense that it is the first biosimilar drug approved by FDA under the trade name Zarzio® in 2015. G-CSF is the first domestic biosimilar pharmaceutical produced in Turkey, and it started being marketed under the trade name Fraven™ by Arven Pharmaceutical Company in 2017.

Escherichia coli is a significant bacterium for biologists and biotechnologists alike. It has been referred to as the work-horse of molecular biology. It has multiple strains that show vast genetic diversity. Genomes of some *E. coli* strains have been fully sequenced and available from online sources. G-CSF as a pharmaceutical was first originally expressed in *E. coli.* While mammalian cell line based expression for pharmaceutical grade G-CSF is relevant, the current industrial production is predominantly *E. coli* based.

Native human G-CSF amino acid sequence is modified in this study. An amino acid native to human G-CSF is replaced with another to avoid expression difficulties. Furthermore six histidines were added to the N-terminal of the protein to facilitate metal affinity purification.

G-CSF is a high value low volume product. While established industrial processes for production of G-CSF exist, there is still room for improvement for novel biosimilars of G-CSF.

The aim of this study is to design and construct a lab-scale process for producing and purifying a His-tagged G-CSF analog in *E. coli.* A strain with a codon optimized synthetic sequence for said analog was constructed, and this strain was cultivated in a 7.0 liter bioreactor. The produced protein was purified and tested for G-CSF activity. Furthermore, a downstream PEGylation experiment was performed to achieve a product with better characteristics.

CHAPTER 2

LITERATURE SURVEY

2.1 Granulocyte Colony Stimulating Factor (G-CSF)

Mature human G-CSF is a 174 amino acid long four helix bundle cytokine that has two disulfide bridges and a free cysteine. (Souza et al., 1986) G-CSF stimulates bone marrow precursor cells to form granulocytic colonies. (Layton et al, 1991) Native human G-CSF has a O-glycolysation site at Thr133. (Oheda et al., 1988) Glycosylation is shown to protect the protein against aggregation and subsequent loss of activity, yet it is not essential for activity. (Oheda et al., 1990) Human G-CSF is a glycoprotein 19,000 Da in size with a isoelectric point between 5.5-6.1. (Nicola, 1987)

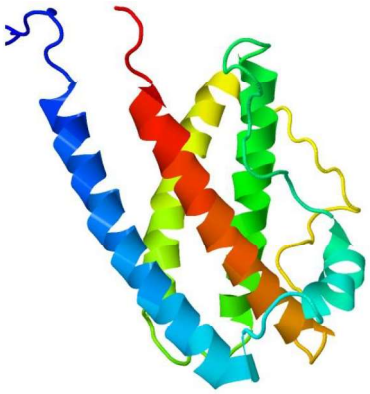


Figure 2.1: Tertiary structure of G-CSF. PDB ID:1GNC

(Zink et al., 1995)

Granulocytes are a type of white blood cells that take part in the natural defense system of the human body. Neutropenia, defined by abnormally low plasma granulocyte count, is the predominant indication of G-CSF. Another medical use for G-CSF is stem cell mobilization, particularly of CD34 cells. (Bensinger et al., 1995) Modern medicine has shifted from bone marrow transplants to leukapheresis of donated blood to harvest and transplant stem cells. Chief advantages are reduced risk to donor and improved transplantation success. (Deotare et al., 2015)

Filgrastim is the non-proprietary name for non-glycosylated methionyl human G-CSF. The production method for filgrastim was patented by Amgen. (Souza, 1991)

Table 2.1: Factory prices for pharmaceutical G-CSF products available in Turkish market. (Retrieved from http://www.titck.gov.tr/ReferansBazliFiyatListesi at 28.8.2017)

Trademark	Active Ingredient	Producer	Country of Origin	Price per mg (TL)
Neupogen	filgrastim	Amgen	Greece	259.85
Leucostim	filgrastim	Dong-A	South Korea	252.51
Leukoplus	filgrastim	CIM	Cuba	207.53
Tevagrastim	filgrastim	Teva	Greece	103.48
Granocyte	lenograstim	Chugai	France	265.23
Fraven	filgrastim	Arven	Turkey	252.51
Neulastim	pegfilgrastim	Amgen	Greece	144.93

2.1.1 Cys17 As a Rational Protein Stabilization Target

Human G-CSF contains 101 hydrophobic amino acids. This hydrophobicity leads to aggregation difficulties while expressing it intracellularly in bacterial cells. An experimental difficulty on top of this problem is the free cysteine residue at the 17th position of the amino acid sequence. This free cysteine is somewhat exposed on the N-terminal and further stabilizes the hydrophobic interaction between G-CSF molecules with disulfide bridges.

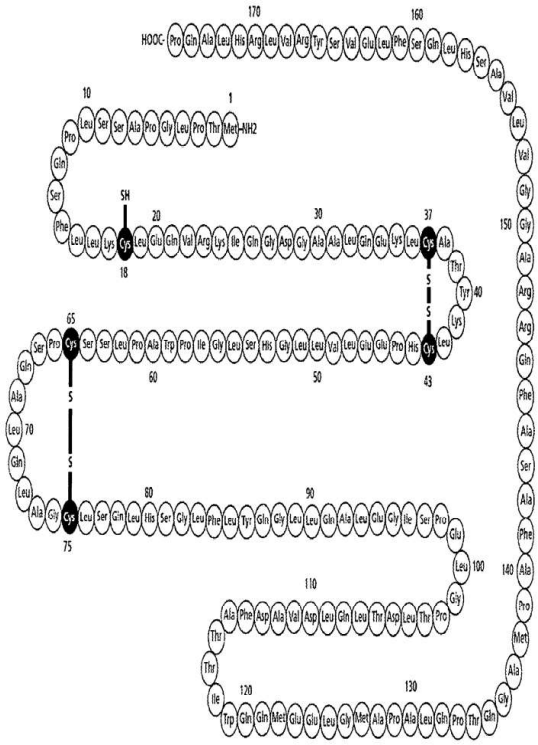


Figure 2.2: Primary structure of recombinant methionyl human G-CSF showing disulfide bridges and a free cysteine that are also found in the native protein. (Toksöz et al., 2011)

Raso and coworkers, (2005) constructed a mutant G-CSF C17A that was less prone to aggregate than native G-CSF. The same study was not able to separate native G-CSF dimers with non-reducing agents, while monomer form of G-CSF C17A was attained easily. Jiang and coworkers, 2011, also constructed the mutant G-CSF C17A with additional residues at the N-terminal to improve solubility, and named the protein G-CSFa. They presented better potency per mass protein *in vitro* and *in vivo.* Nartograstim is a commercialized G-CSF analog produced by the company Kyowa Hakko Kirin. It contains N-terminal amino acid changes and a C17S mutation with respect to filgrastim (Gomes et al., 2012).

2.2 Host Microorganism

Since G-CSF contains a glycosylation site within its structure, mammalian expression systems have an advantage over yeast as host organisms due to their similar glycosylation mechanisms to that of humans. Bacteria grow faster than yeast and other eukaryotic expression systems, however may require extra process steps in order to purify the bioactive form of the protein. In this study, *Escherichia coli* BL21 (DE3) was chosen as host microorganism.

2.2.1 *Escherichia coli*

Escherichia coli shows a vast diversity among its strains. Within the scope of recombinant protein production, some *E.coli* strains are better cloning hosts, while others are better expression hosts.

2.2.1.1 *Escherichia coli* DH5α

This strain is a commonly used cloning host. It shows a superior transformation efficiency when compared to some other cloning strains. (Taylor *et al*.,1993) This strain also does not express the genes endA and recA, which express endonuclease

and recombinase. As a result, the genetic construct to be cloned is kept readily stable in its host.

2.2.1.2 *Escherichia coli* BL21 (DE3)

This strain is first developed and described by Studier and Moffat (1985).

Genotype of *E. coli* BL21 (DE3) is as follows:

E. coli str. B F^- *ompT gal dcm lon* $hsdS_B(r_B^- m_B^-)$ λ(DE3[*lacI lacUV5-T7p07 ind1 sam7 nin5*]) $[malB^+]_{K\text{-}12}(\lambda^S)$*

E. coli BL21 (DE3) naturally lacks proteases OmpT and lon, and thus protein degradation during expression is minimalized. This strain also has the T7 RNA polymerase controlled under lacUV5 promoter, therefore is inducible by lactose or IPTG, an analog of allolactose. By controlling target protein expression under an inducible promoter, normally toxic gene products can be expressed in *E. coli* BL21 (DE3).

2.3 Choice of Expression Plasmid

2.3.1 The Lac Operon

Lactose utilization is under biochemical regulation in *E. coli* and other bacterial cells. Depending on the availabilities of glucose and lactose, the cells control and regulate gene expression to utilize lactose with genes for β-galactosidase (*lacZ*), permease (*lacY*) and transacetylase (*lacA*) enzymes as shown in figure 2.3.

---

*Retrieved from http://www.openwetware.org/wiki/E._coli_genotypes#BL21.28DE3.29 on 25.5.2017.

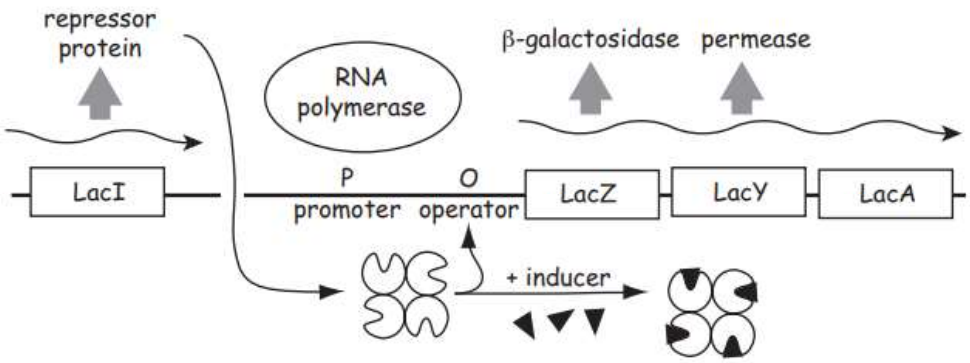


Figure 2.3: Lac operon scheme.*

2.3.2 PRSET A Vector

PRSET A is a pUC-derived, high copy vector for expression of proteins under T7 promoter. According to its manufacturer (Thermofisher, USA), the powerful T7 promoter almost always carry with itself some level of basal expression. However, if the expressed protein is non-toxic to *E. coli*, *e.g.* G-CSF, some basal expression can be tolerated.

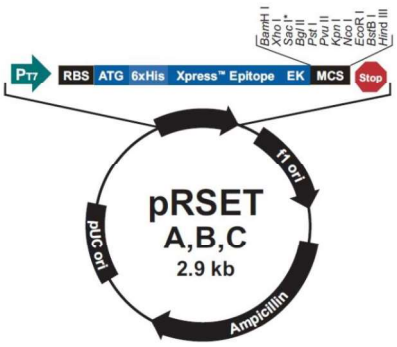


Figure 2.4: Vector map of 2897 bp long pRSET A.

---

*Retrieved from https://ocw.mit.edu/courses/ biology/7-03 -genetics - fall-2004/lecture- notes/lecture16.pdf

2.4 Inclusion body formation

Microbial intracellular production of proteins with high hydrophobic amino acid content may cause the formation of inclusion bodies. Inclusion bodies (IBs) are formations of protein aggregates within the host microorganism. The size and biochemical nature of IBs are dependant on many factors (recombinant protein, host microorganism, fermentation and purification conditions), and are difficult to predict *in silica* prior to experimentation. (Jonasson, 2002)

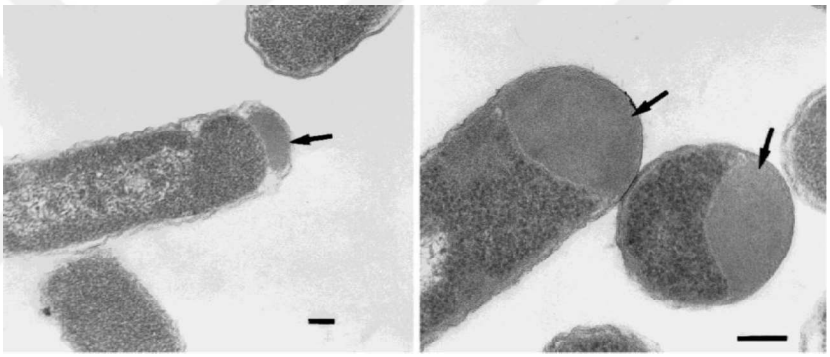


Figure 2.5: Formation of leptin IBs in *E. coli* (Jeong and Lee, 1999) Bars represent 100 nm.

2.5 Genetic Engineering Methods

2.5.1 Codon Optimization

Genetic code, the function that relates triplets of codons to amino acids, is degenerate. 61 codons code for 20 amino acids that make up most of the proteins. Although genetic code is almost universal across species (Crick, 1968), the codon frequency varies greatly. Therefore during heterologous protein expression in *E. coli*, codons commonly found in human gene sequences may create bottlenecks for efficient protein expression.

UUU	F	0.61	28.9	UCU	S	0.15	8.5	UAU	Y	0.69	18.6	UGU	C	0.42	4.2
UUC	F	0.39	18.8	UCC	S	0.14	8.0	UAC	Y	0.31	8.5	UGC	C	0.58	5.8
UUA	L	0.14	17.5	UCA	S	0.11	6.1	UAA	*	0.64	1.9	UGA	*	0.27	0.8
UUG	L	0.15	18.6	UCG	S	0.20	11.4	UAG	*	0.09	0.3	UGG	W	1.00	12.7
CUU	L	0.11	12.7	CCU	P	0.14	5.8	CAU	H	0.56	9.3	CGU	R	0.35	16.4
CUC	L	0.12	14.1	CCC	P	0.06	2.4	CAC	H	0.44	7.2	CGC	R	0.40	18.8
CUA	L	0.03	3.4	CCA	P	0.18	7.4	CAA	Q	0.35	13.5	CGA	R	0.05	2.4
CUG	L	0.45	54.9	CCG	P	0.61	24.9	CAG	Q	0.65	24.7	CGG	R	0.11	5.0
AUU	I	0.48	33.9	ACU	T	0.14	7.7	AAU	N	0.57	21.2	AGU	S	0.16	9.0
AUC	I	0.44	31.0	ACC	T	0.47	25.2	AAC	N	0.43	15.9	AGC	S	0.25	14.3
AUA	I	0.07	5.0	ACA	T	0.11	6.1	AAA	K	0.77	29.2	AGA	R	0.05	2.4
AUG	M	1.00	37.4	ACG	T	0.27	14.6	AAG	K	0.23	8.8	AGG	R	0.04	2.1
GUU	V	0.25	19.6	GCU	A	0.15	13.8	GAU	D	0.66	30.0	GGU	G	0.30	24.4
GUC	V	0.18	14.3	GCC	A	0.28	25.5	GAC	D	0.34	15.1	GGC	G	0.41	33.1
GUA	V	0.14	10.6	GCA	A	0.21	19.6	GAA	E	0.62	29.4	GGA	G	0.10	8.2
GUG	V	0.43	33.9	GCG	A	0.36	32.6	GAG	E	0.38	18.0	GGG	G	0.18	14.3

Figure 2.6: Codon bias for *E. coli* B strain.* (Codon) (Amino Acid) (Fraction) (Codon Frequency per Thousand) *

Codon optimization is a strategy of switching low or sub-optimal frequency codons of native gene to the host organism's codon usage. This does not alter the amino acid sequence of the protein product, but it can greatly enhance its expression. For G-CSF expression in *E. coli*, codon optimized strains and human codon sequence strains produced less than 1% and up to 55% G-CSF to total protein, respectively. (Karimi et al., 2015)

Codon adaptation index (CAI) is a common measure of how well a gene sequence is optimized for heterologous production. However the algorithm for CAI calculation does not value the usage of different codons over using only the most frequent codon for a given amino acid. Therefore, various results may stem from different codon optimization strategies. (Menzella, 2011)

---

*Retrieved from http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi ?species=413997, last accessed at 14.8.2017

2.5.2 Recombinant DNA Technology

For the purpose of heterologous protein production, recombinant DNA methods are employed for strain development. A suitable host microorganism is transformed to overexpress a gene of interest. The following steps are commonly employed for this purpose:

- cDNA for gene of interest is either amplified with PCR from tissue samples, or produced synthetically.
- DNA for gene and vector are digested with endonucleases to obtain sticky ends on DNA fragments.
- DNA and vector are joined in a ligation reaction.
- A suitable host strain is transformed with the ligation mixture.

Selection of true transformants after transformation is usually employed with an antibiotic resistance gene present in vector sequence.

2.6 Bioreactor Operation Conditions

2.6.1 Carbon Source for Fermentation

Fermentation of *E. coli* on glucose may commonly lead to acetate formation. This by-product may hinder growth and affect protein production undesirably during recombinant protein production processes. While other solutions to this problem exist such as process and/or genetic modifications, one of the most convenient solution is switching the substrate to glycerol. (Eiteman and Altman, 2006)

Glycerol is a by-product of biodiesel and saponification industries. The prices of glycerol has dramatically decreased during the expansion of the biodiesel production. In the U.S., the price for crude glycerol was 4 to 11 cents per kilogram in 2011 (Quiespe et al., 2011).

Semi-refined glycerol was selected as the carbon source for *E. coli* growth in bioreactor experiments in this study for its industrial relevance and relatively straightforward cell cultivation properties.

2.6.2 Temperature, pH and Dissolved Oxygen

Previously investigated optimum conditions were attempted to be chosen for this study. Although some researchers have found it plausible to explore different temperatures for fed-batch bioreactors of G-CSF production with *E. coli* (Jevsevar et al., 2005), the optimum temperature for *E. coli* growth is 37°C. This temperature may be too high for retaining the activity of some proteins. However, decreasing this temperature would also increase growth time. Çalık et al. reported the optimum growth pH for *E. coli* growth as 7.2 (2006). Dissolved oxygen is another important parameter for higher density *E. coli* cultivations. While *E. coli* is a facultative anaerobe, its optimum growth is achieved under dissolved oxygen concentrations of 40% that of equilbirium concentration of oxygen between an aqueous solution and atmospheric air.

2.6.3 Bioprocess Kinetics in Fed-Batch Bioreactor

2.6.3.1 Mass balance for biomass

Rate of biomass formation can be expressed in terms of specific growth rate and total biomass:

$$r_x = \mu C_x V \tag{2.1}$$

Neglecting cell loss during sampling and assuming no input or output streams, biomass formation is equal to the change in total biomass:

$$r_x V = \frac{d(C_x V)}{dt} \tag{2.2}$$

Combining equations 2.1 and 2.2 yields:

$$\mu C_x V = \frac{d(C_x V)}{dt} \tag{2.3}$$

Right hand side of equation 2.3 can be differentiated. Rewriting equation 2.3,

$$\mu C_x V = V \frac{d(C_x)}{dt} + C_x \frac{dV}{dt} \tag{2.4}$$

In fed-batch operation, there is no outflow of volume. This follows from the assumption of neglected sampling. Volumetric flow rate into the system can be defined as:

$$Q = \frac{dV}{dt} \tag{2.5}$$

Inserting equation 2.5 into equation 2.4,

$$\mu C_x V = V \frac{d(C_x)}{dt} + C_x Q \tag{2.6}$$

Further simplifying equation 2.6:

$$\frac{d(C_x)}{dt} = (\mu - \frac{Q}{V})C_x \tag{2.7}$$

2.6.3.2 Mass balance for substrate

For fed-batch bioreactor operation, where there is no outflowing stream and only volumetric input is feeding solution of Q_s at a constant concentration of $C_{s,0}$ and rate of substrate utilization r_s, mass balance for substrate is constructed as:

$$Q_s C_{s,0} + r_s V = \frac{d(C_s V)}{dt} \tag{2.8}$$

Rate of substrate utilization can be related to rate of biomass formation by introducing constants for biomass yield on substrate and maintenance coefficient.

$$-r_s = \frac{r_x}{\left(Y_{\frac{X}{S}}\right) + m} \tag{2.9}$$

To relate cell and substrate concentrations, equation 2.8 is rewritten with inserting equation 2.9 into it, and differentiating right hand side:

$$Q_s C_{s,0} - \frac{r_x}{Y_{\frac{X}{S}} + m} V = V\frac{dC_s}{dt} + C_s\frac{dV}{dt} \tag{2.10}$$

In quasi-steady state approximation of fed-batch bioreactor operation, no accumulation of substrate is assumed ($dC_s/dt = 0$). With the additional assumption of substrate depletion at the start of feeding ($C_s=0$ at $t=0$), equation 2.10 is simplified as:

$$Q_s C_{s,0} = \frac{r_x}{\left(Y_{\frac{X}{S}}\right) + m} V \tag{2.11}$$

Inserting equation 2.1 into 2.11:

$$Q_s C_{s,0} = \frac{\mu C_x V}{\left(Y_{\frac{X}{S}}\right) + m} V \tag{2.12}$$

Differential equation 2.3 can be integrated for t=0, $C_x = C_{x,0}$, $V = V_0$ and inserted into equation 2.12:

$$Q_s = \frac{\mu C_{x,0} V_0}{\left(Y_{\frac{X}{S}}\right) + m} \frac{1}{C_{s,0}} e^{\mu t} \tag{2.13}$$

Equation 2.13 provides a profile of volumetric feeding rate for the substrate as a function of time.

2.6.3.3 Mass balance for product

Fed-batch production of recombinant protein can be represented relating formation and accumulation terms.

$$r_p V = \frac{d(C_p V)}{dt} \tag{2.14}$$

Relating rate of product formation to cell concentration with a first order kinetic equation:

$$r_p = q_p C_x \tag{2.15}$$

Combining equations 2.14 and 2.15, equation for specific product formation rate is attained:

$$q_p = \frac{1}{C_x}\left(\frac{C_p}{V} Q + \frac{dC_p}{dt}\right) \tag{2.16}$$

2.6.3.4 Yield Coefficients

Yield coefficients that relate the rate of formation of biomass, rate of formation of product and rate of substrate utilization to each other are given as instantenous yields below:

$$Y_{\frac{X}{S}} = \frac{r_x}{-r_s} = \frac{dC_x/dt}{-dC_s/dt} \tag{2.17}$$

$$Y_{\frac{P}{S}} = \frac{r_p}{-r_s} = \frac{dC_p/dt}{-dC_s/dt} \tag{2.18}$$

$$Y_{\frac{P}{X}} = \frac{r_p}{r_x} = \frac{dC_p/dt}{dC_x/dt} \tag{2.19}$$

When yields are to be calculated within a finite time interval, approximations are done in following manner:

$$Y_{\frac{X}{S}} = \frac{r_x}{-r_s} = \frac{\Delta C_x/\Delta t}{-\Delta C_s/\Delta t} \tag{2.20}$$

$$Y_{\frac{P}{S}} = \frac{r_p}{-r_s} = \frac{\Delta C_p/\Delta t}{-\Delta C_s/\Delta t} \tag{2.21}$$

$$Y_{\frac{P}{X}} = \frac{r_p}{r_x} = \frac{\Delta C_p/\Delta t}{\Delta C_x/\Delta t} \tag{2.22}$$

2.7 Separation

Inclusion bodies, large protein aggregates with high heterologous protein content, are typically solubilized in a mixture of chaotropes (e.g. urea, guanidium chloride) and/or detergents (e.g. sarkosyl, Triton). (Bowden, 1991) If chaotropes are used during solubilization, refolding steps are necessary to achieve active form of the protein. (Singh, 2015)

Immobilized metal affinity chromatography (IMAC) is the use of fixed transition metal ions to selectively purify proteins. (Porath et al., 1975) While histidine and cysteine have an affinity toward ions of copper, zinc, nickel and cobalt ions, selectivity and yield among these ions vary. Histidine tags are sequences of histidine repeats, incorporated usually to either termina of a protein to facilitate IMAC purification. Such tags are commonly employed to purify therapeutic proteins in laboratory settings. However concerns regarding effects on immunogenicity, biological activity and stability prevent His-tagged therapeutic proteins to be used in humans. (Gaberc-Porekar and Menart, 2001)

Jevsevar et al., 2005 report 2-3 fold bioreactor yields of native G-CSF compared to His-tagged version of the protein. This may be due to partial solubilization of the G-CSF molecule by putting charged residues i.e. histidines at the N-terminal. The resulting more soluble structure may have weakened the otherwise highly

hydrophobic interaction of G-CSF molecules that plays a role in formation of IBs, causing lower yield.

2.8 *In vitro* bioactivity

Mouse cell line NFS-60 cells require a growth factor such as interleukin-3, G-CSF, GM-CSF or M-CSF for growth or maintaining life (Weinstein et al., 1986). M-NFS-60 cells are a subline derived from NFS-60 that is adapted to grow in macrophage-colony stimulating factor, but they still respond to G-CSF as well. M-NFS-60 cells were used by some laboratories that contributed to first and second international standardization collaboration studies of G-CSF (Mire-Sluis et al., 1994) (Wadhwa et al., 2010). The profileration method for G-CSF activity was initially described by Shirafuji and coworkers (1988).

MTT is a tetrazolium salt that goes through a color change only in cells with active mitochondria, therefore measure only living cells. (Mosmann, 1983) For a profileration assay, cells are grown in complete medium, washed, concentrated and inoculated into dilutions of samples to be assayed for G-CSF activity. After the assay duration, MTT is added to samples to measure viable cells.. Cells are then lysed, and absorbance is measured.

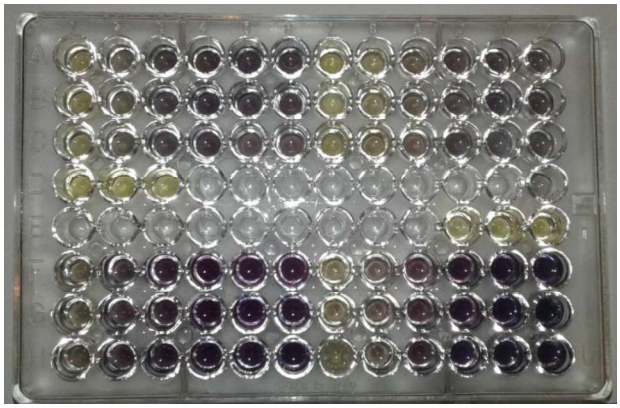


Figure 2.6: 96-well plate containing samples from M-NFS-60 cells grown in various G-CSF concentrations

2.9 PEGylation

In vivo half-life of native G-CSF is 3.5 to 3.8 hours (Molineux, 1999). There are two main pathways G-CSF is thought to be eliminated from circulatory system that does not result in therapeutic effect: renal clearance and proteolytic degradation. Protein PEGylation is covalent conjugation of a polyethylene glycol moiety with a side chain or termina of a protein (Figure 2.7). In the case of G-CSF, Amgen has marketed G-CSF conjugated with a 20 kDa PEG over the last 20 years under the tradename Neulasta® (Kinstler et al., 1998).

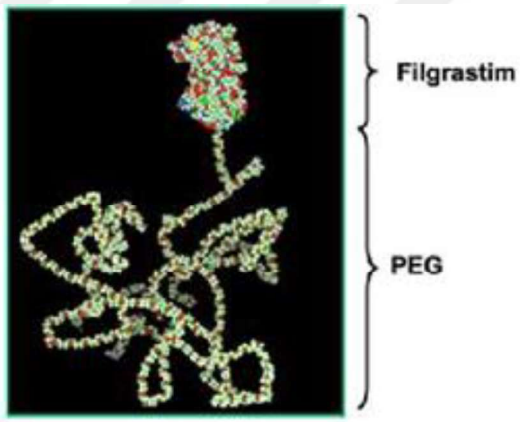


Figure 2.7: Space filling model of PEG-G-CSF. (Molineux, 2004)

Although the molecular weight of PEG is similar to that of G-CSF, it occupies a much larger hydrodynamic volume. Thus depending on the size of PEG molecule, the conjugated PEG-G-CSF displays longer *in vivo* half-lifes. Median half life for 20kDaPEG-G-CSF is 42 hours. A significant practical advantage of PEGylated G-CSF over G-CSF is reducing the need for repeated doses to the patient from 5 injections a week to a single injection for one week. (Molineux, 2004)

CHAPTER 3

MATERIALS AND METHODS

3.1 Chemicals

Chemicals used in this study were purchased from Sigma Aldrich, Merck and Fluka. Reagents were of analytical grade unless specified otherwise.

3.2 Buffers and Stock Solutions

Buffers and stock solutions in this study are listed in Appendix A. They were prepared with ultra pure water unless otherwise stated. Sterility was achieved by autoclaving solutions at 121°C for 20 minutes. Solutions with heat sensitive chemical content was sterilized by filtering through sterile, 0.20 µm cellulose acetate filters (Sartorius, Germany).

3.3 Strains and Plasmids

The vector construct was used to first transform *E. coli* DH5α in order to multiply the plasmid. Isolated plasmids from this strain was used to transform the final production microorganism *E. coli* BL21 (DE3).

Table 3.1: Strains and Plasmids Used in This Study

Microorganism	Strain	Genotype/Plasmid	Source
Escherichia coli	DH5α	wild type	Angardi, 2011
Escherichia coli	DH5α	pRSET A::*gcsf c17a*	This Study
Escherichia coli	BL21(DE3)	wild type	Angardi, 2011
Escherichia coli	BL21(DE3)	pRSET A::*gcsf c17a*	This Study
Escherichia coli	BL21(DE3)	pRSET A	This Study

M-NFS-60 (ATCC® CRL-1838™) was purchased from ATCC for *in vitro* bioactivity assays.

3.4 Genetic Engineering Methods

3.4.1 Restriction endonuclease digestion

Enzymes *Nde*I and *Eco*RI (New England Biolabs, UK) were used for digestion of the insert and vector plasmid. 10X O buffer (ThermoFisher, USA) was compatible to both enzymes. Digestion reactions are summarized in tables 3.2 and 3.3.

Table 3.2: Double digestion reaction of gene fragment.

Component	Amount
Nuclease free water	11 µL
10X O buffer	2 µL
DNA	5 µL (50 ng)

Table 3.2 (continued)

*Eco*RI	1 µL
*Nde*I	1 µL
Total volume	20 µL

Table 3.3: Double digestion reaction of pRSET A.

Component	Amount
Nuclease free water	36 µL
10X O buffer	10 µL
DNA	52 µL (3µg)
*Eco*RI	1 µL
*Nde*I	1 µL
Total volume	100 µL

The digestion reactions were carried out at 37°C for 3 hours with slight agitation in a water bath. Inactivation of enzymes were performed at 65°C for 20 minutes.

3.4.2 Agarose Gel Electrophoresis

Gels of various agarose concentrations (0.8-1.5 %) were used in this study depending on the theoretical length of the DNA molecules that are intended to be separated and/or imaged. Agarose was weighed and dissolved in TAE buffer in an Erlenmeyer flask, heated until boiling. The solution was left to cool for the addition of 1.75 uL of ethidium bromide. The mixture was poured into a gel tray with a suitable comb. After the gel solidifies, the samples and marker was loaded onto gel. The gel was run

at 90V for 30 to 120 minutes depending on the properties of the loaded samples and agarose content of gel. The gel was later photographed and discarded.

3.4.3 DNA Extraction From Agarose Gel

After agarose gel was run with DNA sample, the band of interest was excised with a razor blade. The sample was solubilized using GeneJET Gel Extraction Kit (Thermo Fisher, USA).

3.4.4 DNA Purification

DNA purification was done with GeneJET PCR Purification Kit (Thermo Fisher, USA). Elutions were done with ultrapure water.

3.4.5 Ligation Reaction

Molar ratio of gene to plasmid was taken as 5:1 in the ligation reaction. Ligation reactions were carried out at 25°C for 30 minutes. Deactivation was at 65°C for 10 minutes. The composition of ligation reaction is provided in table 3.4.

Table 3.4: Ligation mixture.

Component	Amount
Nuclease free water	Up to 20 µL
10X ligation buffer	2 µL
Plasmid DNA	50 ng
Gene fragment	42 ng
T4 DNA ligase	1 µL
Total volume	30 µL

3.4.6 Plasmid Isolation

Plasmid isolation was performed with GeneJET Plasmid Miniprep Kit (Thermo Fisher, USA). Elutions were done with ultrapure water.

3.4.7 Transformation

Transformations were performed by $CaCl_2$ method. (Sambrook and Russell, 2001) Briefly, chemically competent cells were prepared by growing cells in LSLB medium until early-log phase (OD_{600}= 0.35-0.50), Cells were centrifuged and resuspended in calcium chloride solution. DMSO was added to the solution, the competent cells were aliquoted and frozen in liquid nitrogen.

The cells were transformed by adding ligation mixture or isolated plasmids to chemically competent cells and incubating cells at 42°C for 90 seconds. After recovery of transformed cells in LSLB, the true transformants were selected on LSLB agar plates supplemented with ampicillin.

3.5 Growth Media

3.5.1 Solid Medium

LSLB Agar is dissolved in ultrapure water in a ratio of 3.5:100(w/v). The solution is autoclaved and upon cooling to ~50°C, poured into petridishes with or without ampicillin (100µg/mL). After gelation, the petridishes are used immediately, or stored at 4°C for later use up to two weeks.

3.5.2 Liquid Media

Low Salt Lysogeny Broth (LSLB)

LSLB is dissolved in ultrapure water at a ratio of 1:50(w/v). The solution is autoclaved and used upon cooling with or without ampicillin addition (100µg/mL).

Korz's medium

Shake-flask experiments and bioreactor runs were conducted with a chemically defined medium used in literature (Korz et al., 1995). This medium was named as Korz's medium for reference. In preparation of this medium, glycerol, monobasic potassium phosphate, dibasic potassium phosphate and magnesium sulfate were sterilized by autoclaving. Citric acid was sterilized by filtration and used freshly. Two trace mineral stock solutions were prepared by filtering corresponding to 250X versions of final concentrations of feeding and batch solutions. The trace mineral solutions were kept at 4°C until use. Thiamine solution was sterilized by filtering, and used freshly.

Table 3.5: Chemically defined batch and feeding media for bioreactor experiments.

Component	Batch (g/L)	Feeding (g/L)
Glycerol	30.0	750.
KH_2PO_4	13.3	
$(NH_4)_2HPO_4$	4.00	
$MgSO_4.7H_2O$	1.20	14.7
Citric acid.H_2O	1.86	
Trace mineral	**Batch (mg/L)**	**Feeding (mg/L)**
EDTA	14.1	9.56
$CoCl_2.6H_2O$	2.50	2.94
$MnCl_2.4H_2O$	15.0	17.3
$CuSO_4.5H_2O$	2.78	1.70
H_3BO_3	3.00	3.45

Table 3.5 (continued)

Trace mineral	Batch (mg/L)	Feeding (mg/L)
$Na_2MoO_4.2H_2O$	2.10	2.94
$ZnCl_2$	21.0	7.30
$FeCl_3.6H_2O$	111	32.4
Thiamine.HCl	4.50	3.3
Antifoam Y-30	500	
Ampicillin	50.0	50.0

For shake-flask experiments, a phosphate buffer was adopted to keep pH from getting too acidic for *E. coli* growth i.e. (pH < 5). In bioreactor experiments, pH was controlled by 25% NH_3 addition, which also served as a nitrogen source.

3.6 Cell Cultivation and Recombinant Protein Production

Fed-batch fermentations coded BR-1 and BR-2 were carried out in a 7.0 L bioreactor with 5.0 L maximum working volume. Starting volumes were 1.8 L with both experiments. An illustration of the bioreactor system is given in Figure 3.1. Batch, feeding and antifoam solutions were sterilized by autoclaving. Trace mineral stock solutions were sterilized by filtering and added to batch and feeding solutions. 25 % ammonia was not sterilized.

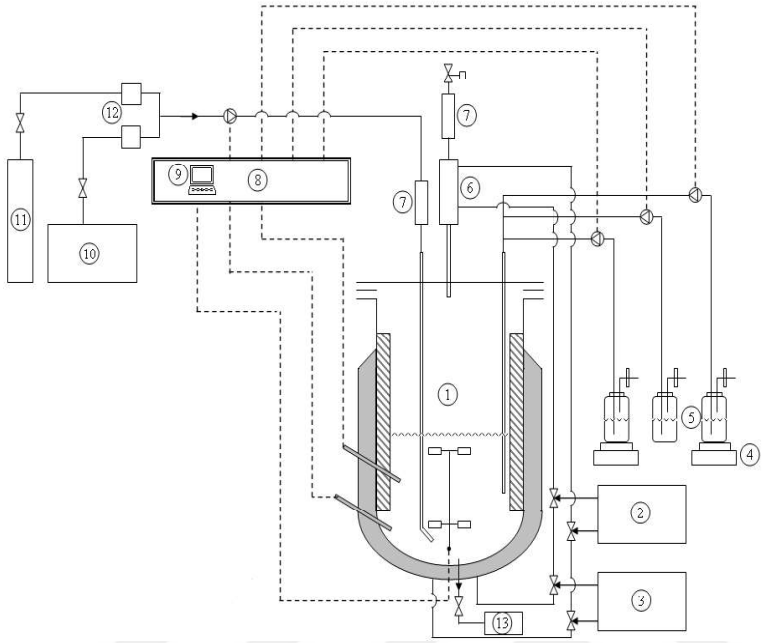


Figure 3.1: Illustration of employed 5.0-L bioreactor system: (1) Fermentor, Biostat Cplus 5L (2) Cooler (3) Steam generator (4) balances (5) substrate, antifoam and base reseviors (6) Exhaust cooler (7) Air filters (8) Regulator (9) Biostat Software (10) Air compressor (11) Pure O_2 tank (12) Digital mass flow meters (13) Sampling valve (Çelik, 2008).

3.7 Protein Purification

3.7.1 Cell Lysis

Bioreactor samples were lysed by homogenization, and shake-flask samples were lysed by lysozyme incubation.

Homogenization

Homogenization was performed through purchase of service from METU Central

Laboratory. Bioreactor samples were centrifuged at 4°C and washed with 0.5% NaCl solution twice. Then the washed pellet was resuspended in 0.5% NaCl to make up for 60% of original volume. Washed, resuspended cell culture was passed through the homogenizer three times at a maximum pressure of 2.54 MPa. The resulting cell paste was stored at 4°C until downstream purification.

Lysozyme incubation

Washed, resuspended cell culture was recentrifuged and constituted in 350 uL STET buffer per 1 mL of original culture. 25 uL lysozyme solution was added per 1 mL original culture. The mixture was vortexed briefly and incubated at 37°C for 30 minutes.

3.7.2 Inclusion Body Solubilization

Inclusion bodies were first washed 2.5M urea solution for 15 minutes and centrifuged at maximum speed for 30 minutes. This process was repeated twice. Then a 6 hour incubation in 7.5M urea solution was performed. Cells were centrifuged again at maximum speed for 30 minutes to pellet any undisturbed inclusion bodies and insoluble cell debris. The supernatant is removed and used for further purification.

3.7.3 Cobalt Affinity Chromatography

Solubilized inclusion bodies were purified with BD TALON™ Metal Affinity Resin under denaturing conditions (7.5M urea). Detailed protocol according to manufacturer's instructions is given at Appendix B.

3.7.4 Dialysis

Two-step dialysis was performed for correct renaturation of 6xHis-G-CSF C17A using 12000 MWCO (Sigma-Aldrich, Germany) Dialysis tubing was prepared according to detalied protocol given at Appendix C.

After preparation, sample was inserted into tubing and dialyzed against 4M urea solution for 20-28 hours at 4°C in a 50:1 volume ratio. Then the solution is dialyzed against, 10 mM sodium acetate at pH=5.0 for 20-28 hours.

3.7.5 Concentration

Vivaspin® 15R, 10,000 Da spin columns (Sartorius, Germany) are used to concentrate sample before Factor Xa digestion. After sample application into column reservoir, the tube was centrifuged at 6000 g at 4°C until desired approximate concentration volume is achieved.

3.7.6 Factor Xa Digestion

Factor Xa (Qiagen, PRC) and Xa removal resin (Qiagen, PRC) was used according to manufacturer's instructions. Detailed protocols from manufacturer is provided in Factor Xa digestion was carried out at 20-25°C for 1.5 hours in a water bath with slight agitation. Detailed protocol is provided in Appendix D.

3.8 *In vitro* Bioactivity Assay

M-NFS-60 cells were grown in RPMI-1640 medium supplemented with 10% FBS, 63 ng/ml human M-CSF (Peprotech, UK) and 0.05 mM 2-mercaptoethanol. Fourfold dilutions of G-CSF C17A and Neupogen in medium without M-CSF were prepared in 10 mL volumes in 6-well plates. Washed, centrifuged and concentrated cells were inoculated into dilutions of preparations at a final concentration of 50,000

cells/mL. The cells were incubated at 37°C, in a 5% CO_2 incubator. Triplicates of 100 µL samples from each dilution were transferred to 96-well plates at intervals up to 96 hours. 10 µL of MTT solution were added to the samples. The samples were incubated for a further 2 hours. The samples were titurated with 100 µL of 80% isopropanol, 10% Triton X-100 and 10% 1M HCl to disrupt cells. Absorbance of the samples were measured at 595 nm.

3.9 PEGylation

PEGylation experiment was performed with a 50 µL reaction volume in a 0.2mL tube. Sodium cyanoborohydride and PEG solutions were prepared freshly. Concentrations of reactants and catalysts are provided in Table 3.4. Factor Xa cleaved G-CSF C17A solution was reconcentrated further, and its buffer was exchanged to 10 mM sodium acetate, pH=5.0. Methoxy aldehyde PEG with molecular weight of 20,000 Da was purchased from Jenkem, China.

Table 3.6 Reaction mixture for PEGylation of GCSF C17A.

Chemical	Concentration
G-CSF C17A	130 ng/µL
PEG	940 ng/µL
$NaCNBH_3$	20mM
$NaCH_3COO$	10 mM

The reaction mixture was incubated at room temperature and at 250 rpm for 24h on a orbital shaker.

3.10 Analysis

3.10.1 Cell concentration

Cell concentration calibration curve was performed by washing pelleted cell solutions of known volume and absorbance in pre-weighed microfuge tubes with 0.5% NaCl. Then the cells were pelleted again and left to dry at 65°C until constant weight. The microfuge tubes were then put in a dessicator to reach ambient temperature, and weighed.

A dilution factor (DF) was adopted so that the read-out from the spectrophotometer corresponds to a value between 0.1-0.9 to ensure sensitivity of the read-out.

$$C_x(g/L) = 0.248\ x\ OD_{600} x\ DF$$

$$C_x(g/L) = 0.208\ x\ OD_{550} x\ DF$$

Calibration curves for wavelengths of 550 nm and 600 nm are given in Appendix E.

3.10.2 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed as described by Laemmli, 1970. Reduced samples were prepared by 100 mM final DTT concentration in 1X Laemmli buffer. Gels with 1.0 mm thickness were loaded with samples with either whole cells or purified protein solutions.

Staining

Coomassie blue staining was performed by the protocol provided by Turku Centre for Biotechnology.*

Imaging

Imaging of stained gels was performed in a UVP imaging cabinet equipped with a Hamamatsu camera, model C8484-51-03G.

Concentration Determination

Concentration was determined by a 2-3 point calibration curve using diluted Neupogen® standard samples on gel. The band intensities of unknown samples were then compared to that of calibration curves. Whole cell samples were prepared as described by (Sambrook and Russell, 2001).

3.10.3 Western Blotting

Transfer to methanol activated PVDF membrane was done at 50V for 3 hours. Western Blotting was performed by blocking in 5% milk in TBST buffer, followed by incubation with primary and secondary antibody in TBST-milk solutions at room temperature for 1-2 hours. The experiments were based on Biorad's protocol.**

---

*https://www.btk.fi/wp-content/uploads/sites/3/2017/06/ Coomassie_Blue_staining_version_W.01.pdf, last accessed on 25.8.2017

** http://www.bio-rad.com/webroot/web/pdf/lsr/literature /Bulletin_6376.pdf, last accessed on 25.7.2017

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Strain Development

A protein containing human G-CSF C17A amino acid sequence, Factor Xa site and 6xHistag was designed. A synthetic gene corresponding to this protein was created by optimization of its codons for *E. coli* B cells.

The expression vector was prepared by double digestion and ligation of vector and 6xHisTag-GCSF C17A gene. The vector was amplified by transforming *E. coli* DH5α cells with the ligation mixture. The isolated plasmid was used to transform *E. coli* BL21 (DE3) cells for the construction of the expression strain.

Verification steps comprised of PCR, restriction digestion and nucloetide sequencing between transformations. Microbank stocks of new strains were generated and stored at -80˚C until further use.

Steps for strain developments are shown as a flowchart in Figure 4.1.

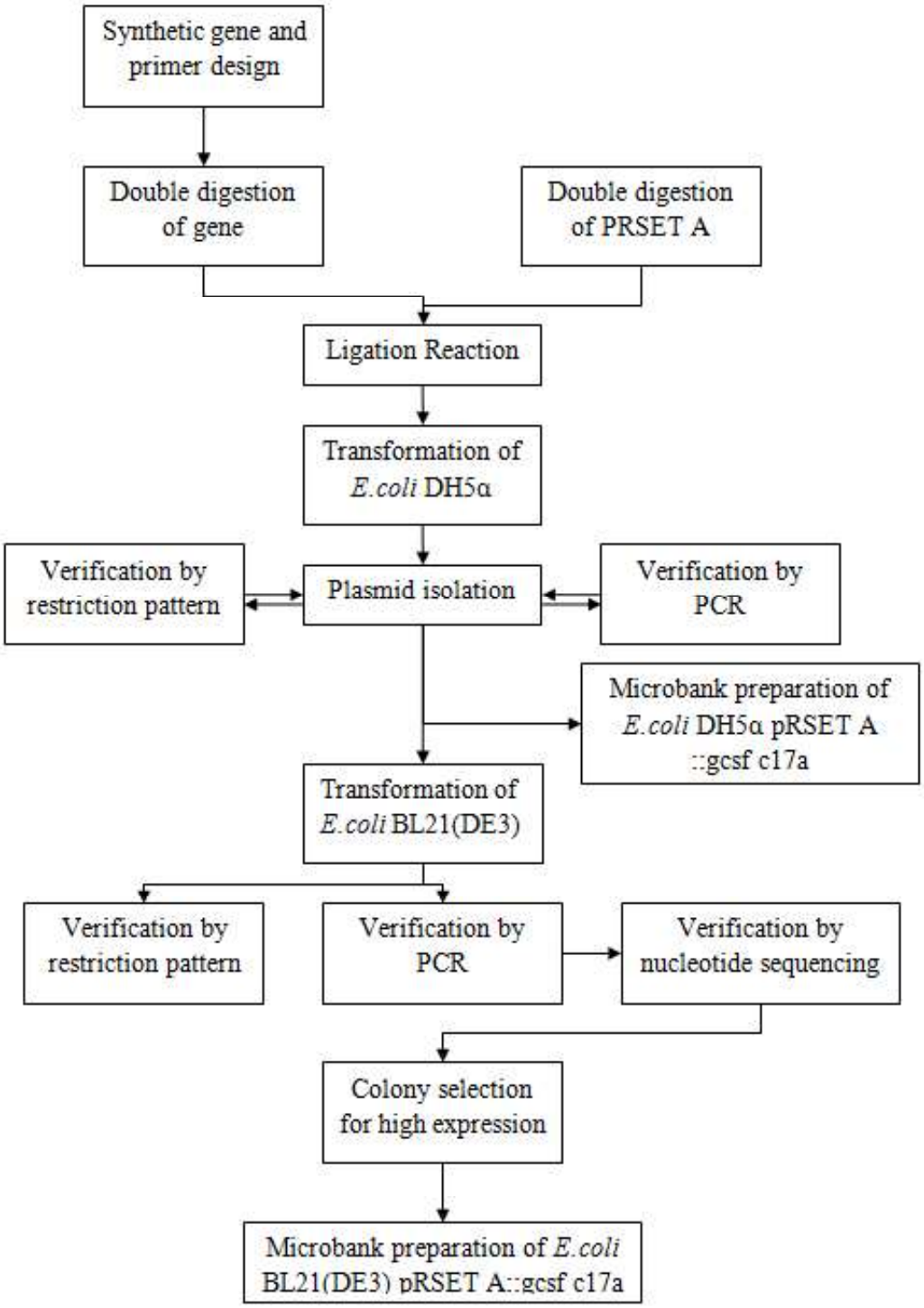


Figure 4.1: Flowchart for strain development.

4.1.1 Synthetic Gene Construct

Amino acid sequence for methionyl 6xHistagged human G-CSF C17A including a Factor Xa cleavage site after Histidine tag is as follows:

MHHHHHHIEGRTPLGPASSLPQSFLLKALEQVRKIQGDGAALQEKLVSECAT
YKLCHPEELVLLGHSLGIPWAPLSSCPSQALQLAGCLSQLHSGLFLYQGLLQ
ALEGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFAS
AFQRRAGGVLVASHLQSFLEVSYRVLRHLAQP

M : Methionine required for bacterial expression.

HHHHHH : 6x Histidine tag required for IMAC.

IEGR : Recognition sequence for removal of 6xHistidine tag with Factor Xa.

A :Alanine residue to be remained in mature peptide, replacing cysteine.

Factor Xa is a specific protease that cleaves right after C-terminus of the recognition sequence. The 6xHistag was placed at the N-terminus to avoid any remaining amino acids such as methionine and Factor Xa recognition site in the cleaved protein.

Synthetic gene sequence design was performed manually. Theoretical factors such as the ramp hypothesis, efficient folding during translation and GC content ratio of expressed fusion gene to host cell genome were considered. Codon bias of *E. coli* B strain was taken into acount while constructing one corresponding gene sequence out of 9.97×10^{100} possible nucleotide sequences for the desired amino acid sequence.

The synthetic G-CSF C17A gene was purchased from IDT, Belgium in the form of gblock® i.e. linear, double stranded DNA. The gene sequence is provided in Appendix F.

Primer Design

Primers were designed to amplify gene of interest that is cloned into pRSET A vector. Primers consisted of short regions from vector sequence designed to amplify any inserted sequence. Picking close melting points and avoiding any stable secondary structure within and between primers were considered while designing the primer set. Primers were purchased from IDT, Belgium.

Forward 5' - TAGGGAGACCACAACGGTTTCC - 3'

Reverse 5' - GCTTTGTTAGCAGCCGGATCAA - 3'

4.1.2 Expression vector construction

PRSET A plasmid was double digested with *Nde*I and *Eco*RI restriction endonucleases and was run on an agarose gel. The double digested vector was extracted from gel. (Figure 4.2)

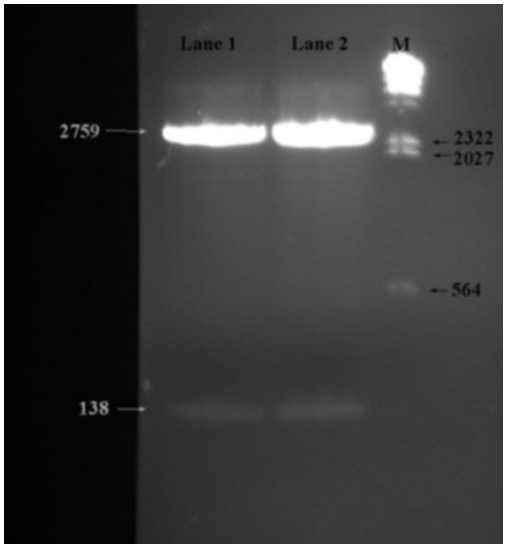


Figure 4.2: Double digested pRSET A plasmid against Lambda DNA/*Hind*III Marker on 1.5% agarose gel.

G-CSF C17A gene was also double digested with NdeI and EcoRI enzymes. The sample was purified with PCR Purification Kit. After that, a ligation mixture was prepared and used for transformation of cloning strain.

4.1.3 Transformation of cloning strain

The gene and pRSET A were double digested separately, and mixed into ligation mixture. The mixture was used to transform *E. coli* DH5α cells. The isolated plasmids from transformants were run on agarose gel for checking whether they give correct size. (Figures 4.3 and 4.4)

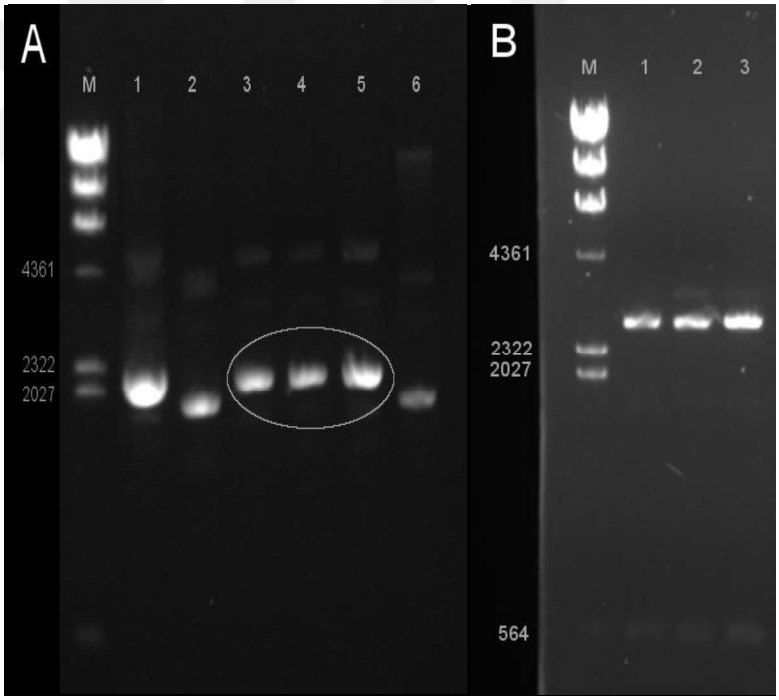


Figure 4.3: A) Undigested plasmid isolations from transformants of *E. coli* DH5α PRSETA::gcsf c17a. M: Lambda DNA/*Hind*III marker, Lane 1: colony 5, Lane 2: colony 7, Lane 3: colony 13, Lane 4: colony 14, Lane 5: colony 16, Lane 6: pRSET A plasmid as control. B) Double digestions of plasmids from selected samples. M: Lambda DNA/HindIII marker, Lane 1: colony 13, Lane 2: colony 14, Lane 3: colony 16.

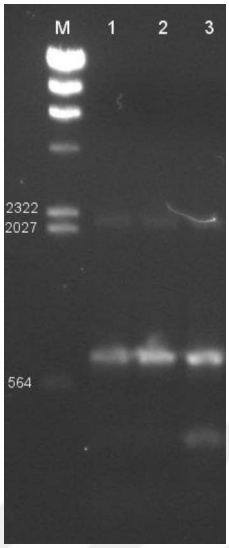


Figure 4.4: Amplified PCR products of selected colonies. M: Lambda DNA/HindIII marker, Lane 1: colony 13, Lane 2: colony 14, Lane 3: colony 16.

4.1.4 Transformation of expression strain

After the vector is readily available as plasmids in *E. coli* DH5α cells, the isolated pRSET A::gcsf c17a plasmid from colony 13 was used to transform *E. coli* BL21 (DE3) cells. Another transformation with wt pRSET A was also performed for use of negative control in future experiments.

As evident in Figure 4.5 A, circular plasmid isolates from colonies of *E. coli* BL21(DE3) pRSET A:: gcsf c17a show variance in apparent size. (2000-9000 bp) This may be explained by the fact that BL21(DE3) is not a cloning strain, and readily expresses endonucleases and recA. Thus the larger plasmids are thought to be concatomers of the intended plasmid with apparent size of 2000 bp. The double digestion of these samples confirm this hypothesis by showing the two correct theoretical linear DNA fragments of 564 bp and 2758 bp (Lanes 5-8).

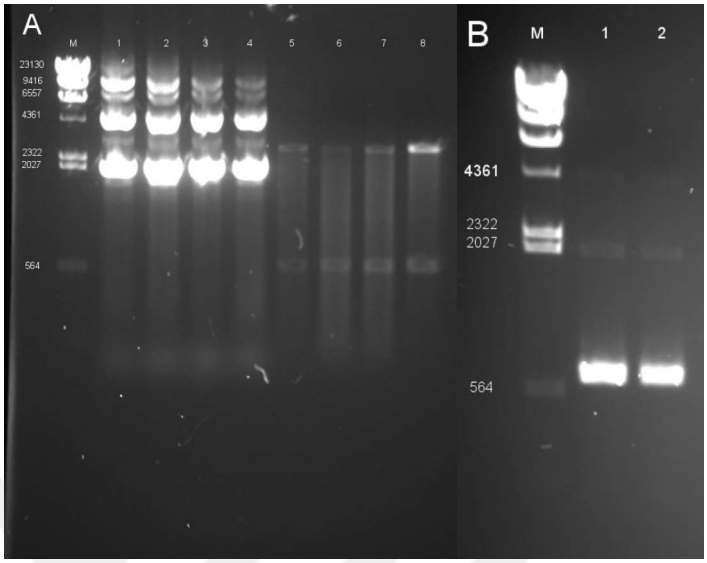


Figure 4.5: A) Undigested and digested plasmid isolations from transformants of *E. coli* BL21 (DE3) pRSETA::gcsf c17a. M: Lambda DNA/HindIII marker, Lane 1: colony 9, Lane 2: colony 10, Lane 3: colony 11, Lane 4: colony 12, Lane 5: colony 10, Lane 6: colony 11, Lane 7: colony 12, Lane 8: colony 9. Lanes 1-4 are undigested circular plasmids and Lanes 5-8 are their double digested plasmids. B) PCR product of gene amplicate from colony 9. M: Lambda DNA/HindIII marker, Lane 1: colony 9 amplicate, Lane 2: colony 9 amplicate at lower concentration.

Although there appears to be some nonspecificity, since the PCR product size is consistent with theoretical value (659 bp) on figure 4.4 B, the PCR product was sent to nucleotide sequencing.

4.1.5 Verification Steps

Nucleotide Sequencing

The cloned gene was verified by nucleotide sequencing. Alignment of this data with designed sequence gave 100% identity, indicating not mutations and base shifts. Sequencing data and BLASTn alignment is available in Appendix F.

Western Blot

A Western blot with polyclonal Anti-GCSF antibody was performed. G-CSF specific detection was observed with expression strain relative to the negative control strain. An image of the blot is shown in Figure 4.6.

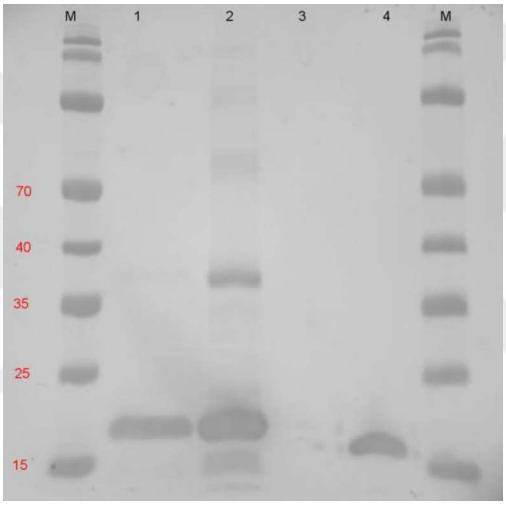


Figure 4.6: Western Blot image using Anti-G-CSF antibody. M: PageRuler™ Prest. Prot. Lane 1: purified histagged G-CSF C17A, Lane 2: Whole cell *E. coli* pRSET A::6xHis-G-CSF C17A, Lane 3: Whole cell *E. coli* pRSET A as negative control, Lane 4: G-CSF standard as positive control, M: Prot. Prest. Marker.

4.2 Recombinant Protein Production

4.2.1 Shake-flask Experiments

After developing the expression strain and confirming G-CSF C17A production with Western Blot, a shake flask experiment was conducted to test the defined medium. Precultivation was done in LSLB, and inoculation ratio to defined medium was 1:100. Firstly, glucose as a carbon source is tested as a duplicate shake-flask experiment. (Figure 4.7)

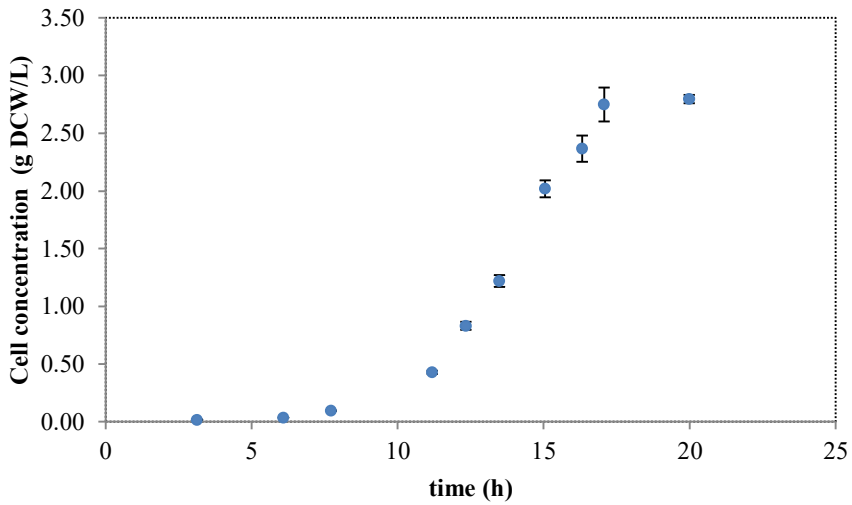


Figure 4.7: Growth curve for glucose shaker experiment. N=250 rpm, V= 50 mL, T= 37°C. Error bars represent standard deviation between duplicates.

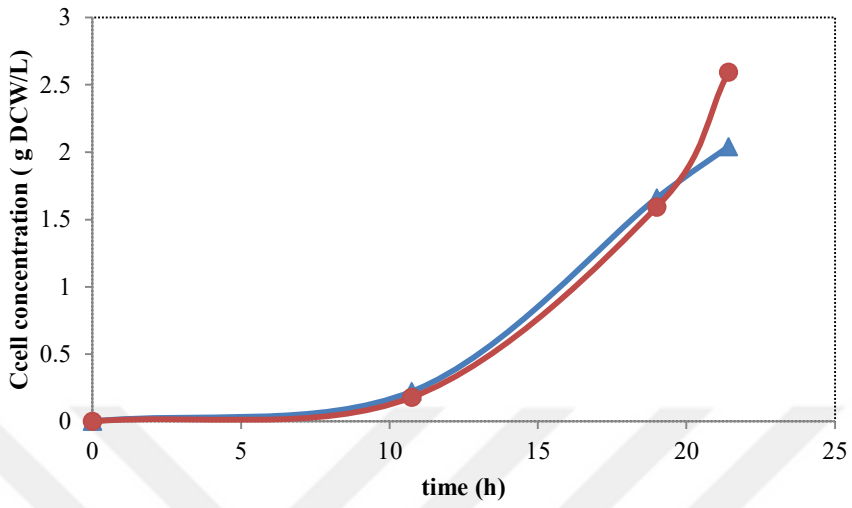


Figure 4.8: Growth curves for glucose vs glycerol shake-flask experiment. Glycerol (●), Glucose (▲).N=250 rpm, V= 50 mL, T= 37°C.

During *E. coli* growth on glucose, acetic acid is formed as a by product. Therefore, less growth in glucose as opposed to in glycerol in Figure 4.8 may be explained by acetic acid formation in medium. While it is stated in the literature that growth on glucose is fastest, *E. coli* is a facultative anaerobic bacterium. It follows that in any case growth would be limited in shake flask experiments by oxygen transfer through shaking. A maximum cell concentration of 2.8 g/L was attained in the shake flask experiments. (Figure 4.7)

4.2.2 Bioreactor Experiments

Two bioreactor experiments were conducted and coded BR-1 and BR-2. The culture grown in first experiment (BR-1) was uninduced to explore whether the cell concentration would or would not reach medium density values (~50 g/L). In the

second bioreactor experiment (BR-2), the culture was induced with 2mM IPTG.

A comparison between bioreactor medium compositions between this study and medium used in Korz and coworkers' study (1995) are tabulated in Table 4.1. The concentration departures mainly arose because of the use of different anion or cation salts of the same trace mineral. Furthermore, the concentrations of salts in feeding solution was normalized to the lower concentration of glycerol set at this study.

Table 4.1: Media modifications done while adapting Korz's medium to this study.

Korz's Medium	Batch (g/L)	Feeding (g/L)	This study	Batch (g/L)	Feeding (g/L)
Glycerol	30.0	1021.0	Glycerol	30.0	750.
KH_2PO_4	13.3		KH_2PO_4	13.3	
$(NH_4)_2HPO_4$	4.0		$(NH_4)_2HPO_4$	4.00	
$MgSO_4.7H_2O$	1.2	20.0	$MgSO_4.7H_2O$	1.20	14.7
Citric acid	1.7		Citric acid.H_2O	1.86	
Korz's Medium	**Batch (mg/L)**	**Feeding (mg/L)**	**This study**	**Batch (mg/L)**	**Feeding (mg/L)**
EDTA	14.1	13.0	EDTA	14.1	9.56
$CoCl_2.6H_2O$	2.5	4.0	$CoCl_2.6H_2O$	2.50	2.94
$MnCl_2.4H_2O$	15.0	23.5	$MnCl_2.4H_2O$	15.0	17.3
$CuCl_2.2H_2O$	1.5	2.5	$CuSO_4.5H_2O$	2.78	1.70
H_3BO_3	3.0	5.0	H_3BO_3	3.00	3.45
$Na_2MoO_4.2H_2O$	2.5	4.0	$Na_2MoO_4.2H_2O$	2.10	2.94
$Zn(CH3COO)_2.2H_2O$	13.0	16.0	$ZnCl_2$	21.0	7.30
Fe(III)citrate	100.0	40.0	$FeCl_3.6H_2O$	111	32.4
Thiamine.HCl	4.5		Thiamine.HCl	4.50	3.3
Antifoam SP1	100		Antifoam Y-30	500	
Ampicillin	50	44.07	Ampicillin	50.0	50.0

It can be inferred from equation 2.13 that the higher substrate concentration of feeding solution is, the less volumetric flow of substrate is needed. Limiting volumetric flow rate is important for keeping an acceptable working volume within the bioreactor. 84-88%(w/w) glycerol (Sigma, USA) was used in this study. Due to separately sterilized magnesium sulfate solution and trace element solution addition, glycerol concentration in feeding solution was set at 750 g/L.

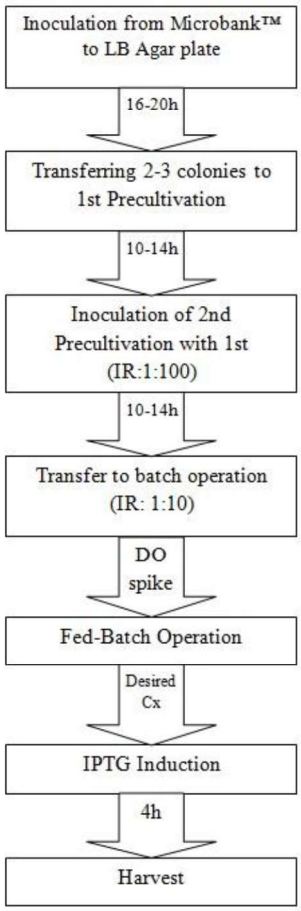


Figure 4.9: Process flowchart for BR-2. BR-1 has the same process except for IPTG induction step.

For this study, a medium cell density in bioreactor culture was targeted. The set values adopted from *E. coli* fed-batch bioreactor research articles (Riesenberg et al., 1991; Korz et al., 1995; Seeger et al., 1995) are given in Table 4.2.

Table 4.2: Adopted constant values for exponential fed-batch bioreactor experiments.

Symbol	Explanation	Value
m	Maintenance coefficient	0.025 (g g^{-1})
$Y_{\frac{X}{S}}$	Yield coefficient of biomass/glycerol	0.45 (g g^{-1})
μ_{set}	Specific growth rate	0.12 h^{-1}

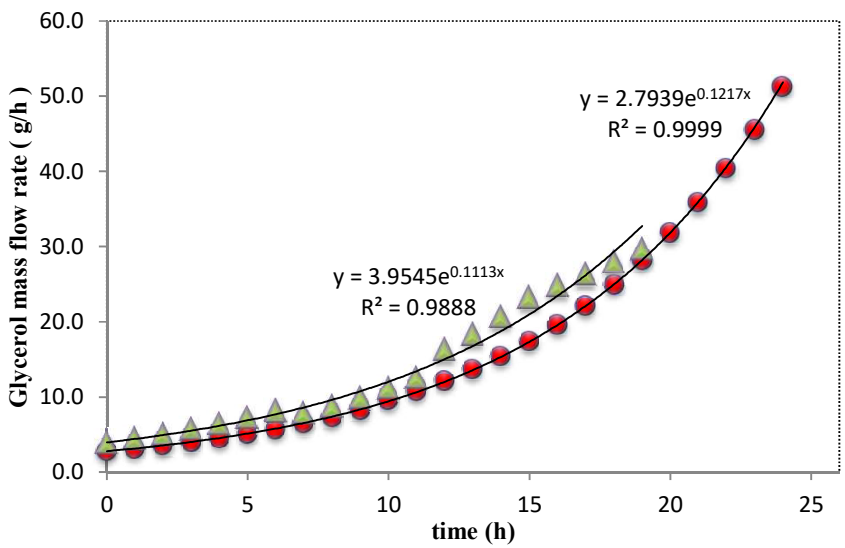


Figure 4.10: Feeding flow rate for bioreactor experiments; BR-1 (●), BR-2(▲).

As evident in figure 4.10, the equations for mass flow rate of substrate are different from each other. These equations are dependent on initial cell concentration (equation 2.13), and experimental $C_{x,0}$ of BR-1 was 20% lower than that of BR-2. Also feeding equation does not take account of additional volumetric perturbations of base, antifoam and sampling. Feeding profile for BR-2 was corrected for these perturbations by correcting volume values in equation 2.13 with experimental data every 6 hours.

Figure 4.11 shows the cell density concentration in the two bioreactor experiments. Experimental difficulties led to a shift in cell growth profile in BR-1 around t=12h. This shift results in an awkward conclusion of induced culture growth being faster than that of uninduced culture. However a careful cross examination of timewise specific growth rate data given in Table 4.3 and Table 4.4, gives insight on how the growth experimentally slowed down during induction with respect to uninduced cell culture.

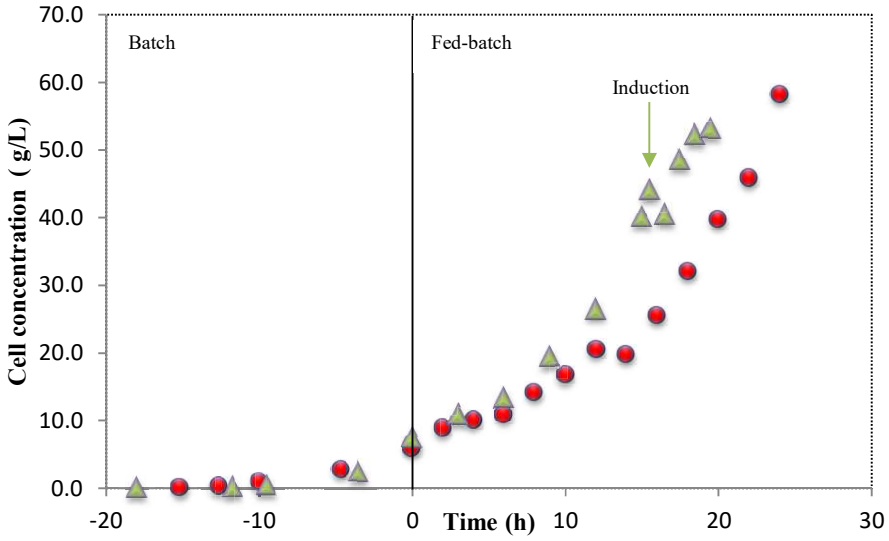


Figure 4.11: Cell concentration profiles in bioreactor experiments.

BR-1 (●), BR-2(▲). t = 0 signifies the start of fed-batch.

Product concentration was determined with SDS-PAGE, by comparing intensities of standard and sample bands. Gel images for samples taken from bioreactor experiments are provided in Appendix G.

Product concentration data of BR-1 and BR-2 during different time points is summarized in figure 4.12. Significant product formation during BR-1 was observed. There is no IPTG induction in BR-1, or lactose content in adopted defined medium. However, T7 is a very strong promoter, and that in the case of BR-1, caused leaky expression.

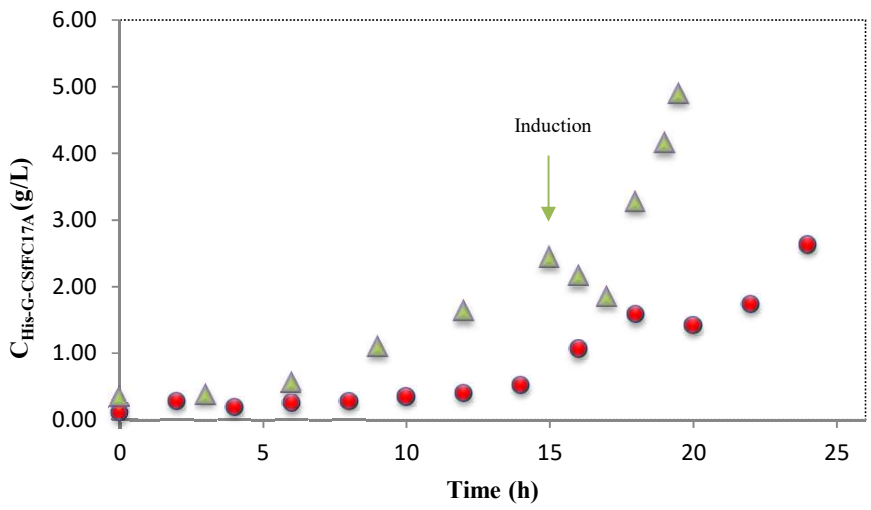


Figure 4.12: Variaton of 6xHistag-G-CSF C17A concentration with fed-batch cultivation time; BR-1 (●), BR-2(▲).

Biomass and product formation rates throughout the bioreactor experiments are shown in figures 4.13 and 4.14, respectively. While biomass formation data is hard to differentiate among BR-1 and BR-2, product formation appears to be more pronounced in BR-2.

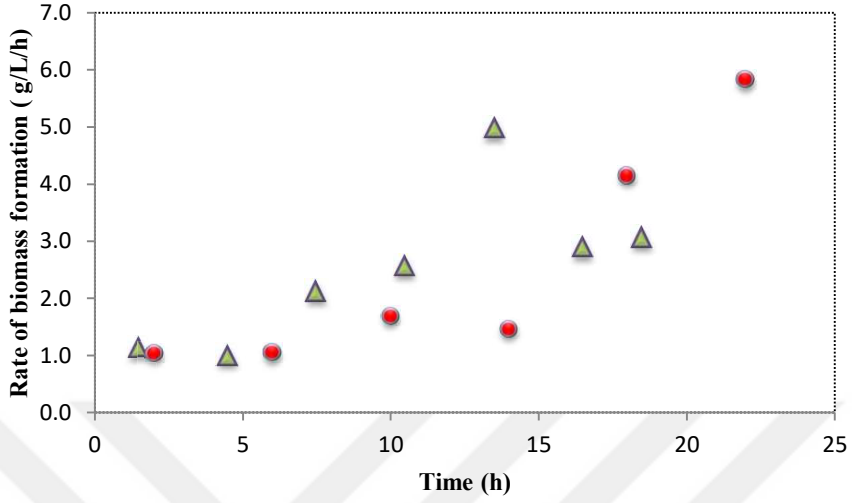


Figure 4.13: Variation of biomass formation rates in bioreactor experiments.

BR-1 (●), BR-2(▲).

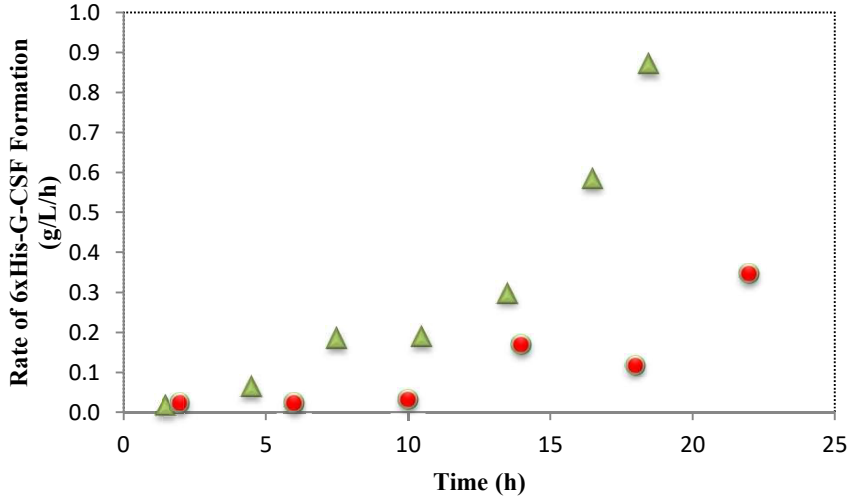


Figure 4.14: Variation of product formation rates in bioreactor experiments;

BR-1 (●), BR-2(▲).

Tables 4.3 and 4.4 show calculated experimental Best product yield per biomass per time values were obtained during the induction phase (15.5h-19.5h) of BR-2. Negative values of yields are due to process difficulties related to insufficient oxygen and partially due to semi-quantitative limitation of SDS-PAGE as a concentration measurement assay.

Experimental specific growth rate during fed-batch operation was 88% and 93% of set value of 0.12 h^{-1} during BR-1 and BR-2, respectively. Similarly the biomass yield on substrate was taken as 0.45 g g^{-1} for the bioreactor experiments. Fed-batch experimental yields are calculated as 0.34 g g^{-1} and 0.42 g g^{-1} for BR-1 and BR-2, respectively. Source of error for these values may be a result of error from total volume calculations.

Table 4.3: Experimental specific growth rate, specific biomass formation rate, specific product formation rate and yield values for BR-1.

time interval	μ_{exp} (h^{-1})	q_x (g $g^{-1}h^{-1}$)	q_p (g $g^{-1}h^{-1}$)	$Y_{X/S}$ (g g^{-1})	$Y_{P/S}$ (g g^{-1})	$Y_{P/X}$ (g g^{-1})
Batch	0.240	0.023	0.001	0.352	0.006	0.018
0-2	0.200	0.452	0.030	0.903	0.055	0.061
2-4	0.060	0.143	-0.038	0.285	-0.022	-0.076
4-6	0.036	0.076	0.035	0.151	0.010	0.069
6-8	0.132	0.255	0.005	0.510	0.006	0.011
8-10	0.085	0.160	0.013	0.320	0.008	0.027
10-12	0.107	0.188	0.007	0.376	0.005	0.015
12-14	-0.017	-0.015	-0.150	-0.031	0.009	-0.300
14-16	0.132	0.187	0.044	0.374	0.033	0.088
16-18	0.122	0.182	0.038	0.364	0.028	0.076
18-20	0.117	0.184	-0.005	0.367	-0.004	-0.010
20-22	0.085	0.139	0.024	0.278	0.013	0.048
22-24	0.130	0.209	0.033	0.417	0.027	0.065
Fed-batch	0.106	0.014	0.002	0.340	0.016	0.047
Overall	0.158	0.009	0.001	0.341	0.015	0.045

Table 4.4: Experimental specific growth rate, specific biomass formation rate, specific product formation rate and yield values for BR-2.

time interval	μ_{exp} (h^{-1})	q_x ($g\ g^{-1}h^{-1}$)	q_p ($g\ g^{-1}h^{-1}$)	$Y_{X/S}$ ($g\ g^{-1}$)	$Y_{P/S}$ ($g\ g^{-1}$)	$Y_{P/X}$ ($g\ g^{-1}$)
Batch	0.240	0.025	0.003	0.454	0.021	0.045
0-3	0.121	0.155	0.005	0.465	0.007	0.014
3-6	0.085	0.100	0.021	0.299	0.019	0.064
6-9	0.133	0.170	0.029	0.510	0.044	0.086
9-12	0.114	0.155	0.025	0.464	0.034	0.074
12-15	0.155	0.192	0.020	0.577	0.034	0.060
15-15.5	0.204	1.582	-0.123	0.791	-0.049	-0.061
15.5-16.5	-0.066	-0.252	0.099	-0.252	-0.025	0.099
16.5-17.5	0.194	0.743	0.169	0.743	0.125	0.169
17.5-18.5	0.088	0.364	0.212	0.364	0.077	0.212
18.5-19.5	0.033	0.139	0.466	0.139	0.065	0.466
Fed-batch	0.112	0.021	0.005	0.410	0.040	0.098
Overall	0.173	0.011	0.002	0.415	0.038	0.092

4.3 Separation of G-CSF C17A

After harvesting biomass at the end of bioreactor experiments, a separation scheme is applied to the samples to isolate purified and bioactive G-CSF C17A. The steps for this scheme is summarized in figure 4.15.

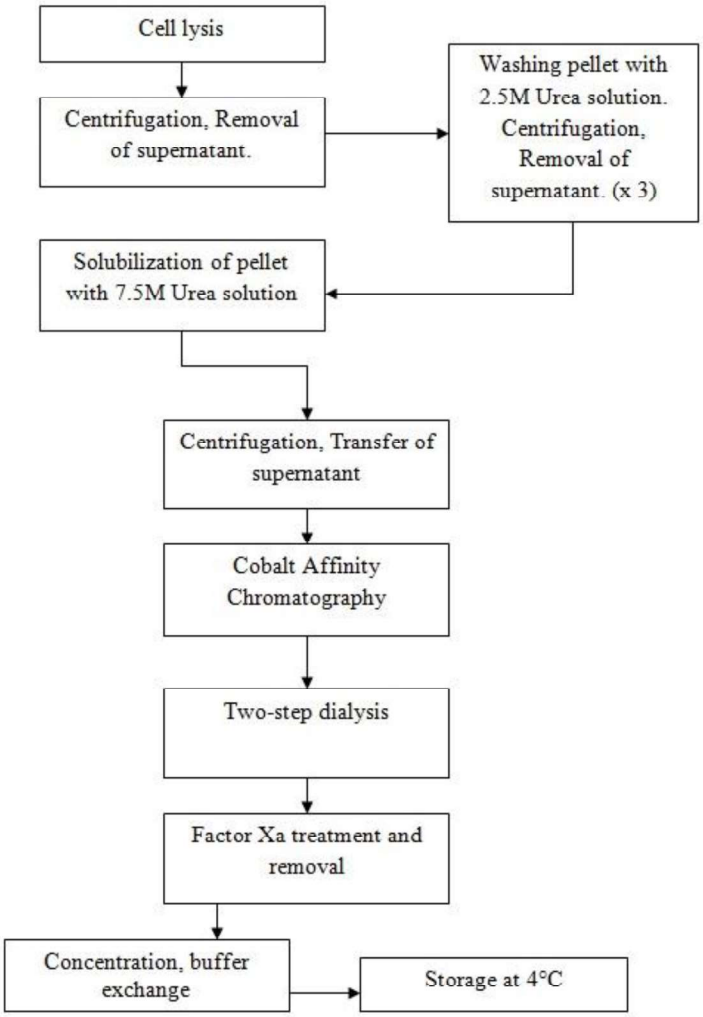


Figure 4.15: Flowchart for Separation of G-CSF C17A.

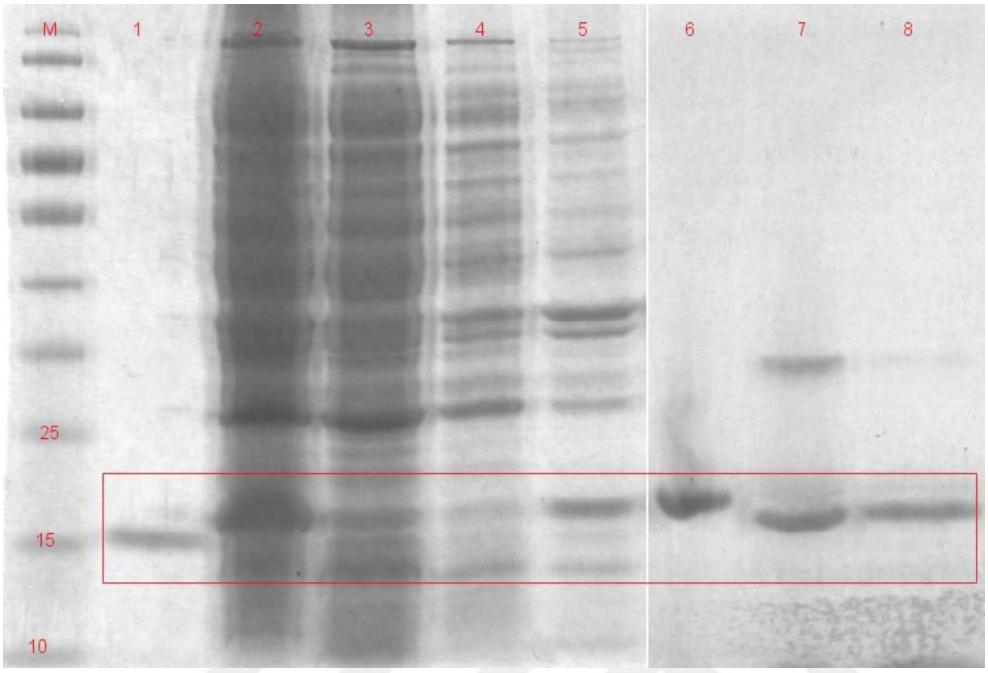


Figure 4.16: SDS-PAGE for separation. M: PageRuler™ Prest. Prot. Ladder Lane 1: Neupogen standard, Lane 2: Homogenized BR-2 sample, Lane 3: 2.5M urea wash 1, Lane 4: 2.5M urea wash 2, Lane 5: 7.5M urea extraction, Lane 6: Purified 6xHistag-GCSF C17A, Lane 7: Factor Xa reaction, Lane 8: Factor Xa removal.

Figure 4.16 visualizes the protein samples through some of the steps involved in separation. While inclusion body solubilization resulted in a experimental yield of 1.0%, experimental column purification yield was 8.2%. While low yields are not uncommon during purification of intracellularly produced proteins, experimental yields were particularly low. Separation parameters such as temperature, pH and salt concentration could be optimized for better yield. In any case GCSF C17A with a protein purity of 81% was purified for further application within this study.

After purification of G-CSF C17A, correct formation of disulfide bridges were checked by SDS-PAGE. Non-reduced samples were expected to display a smaller apparent size on the gel due to retaining intramolecular bonds. Reduced and non-reduced samples of Neupogen and purified G-CSF C17A were run on gel to compare difference in apparent size (Figure 4.17). The difference between reduced and non-reduced samples of G-CSF C17A and Neupogen were similar. While this is not an exhaustive analysis of correct structure of purified G-CSF C17A, there is evidence that disulfide bonds were correctly formed in purified product.

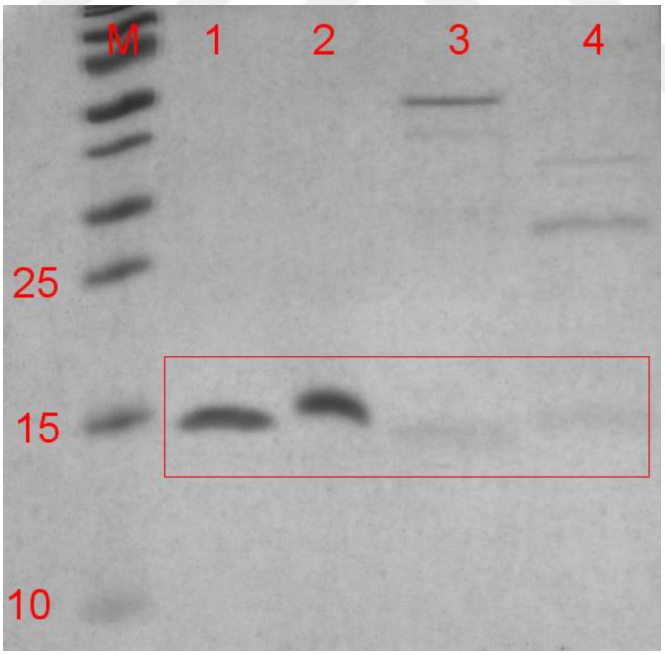


Figure 4.17: 15% SDS-PAGE gel for checking disulfide bridges. M: PageRuler™ Prest. Prot. Ladder Lanes 1-2: GCSF standard. Lanes 3-4: Purified G-CSF C17A. Odd lane samples were non-reduced while even lane samples were reduced with 100 mM DTT.

4.4 *In vitro* bioactivity assay of G-CSF C17A

A concentration and time dependent assay was designed and conducted. Six 4-fold dilutions of 500 pM G-CSF C17A and Neupogen were assayed triplicately in a time range of 96 hours.

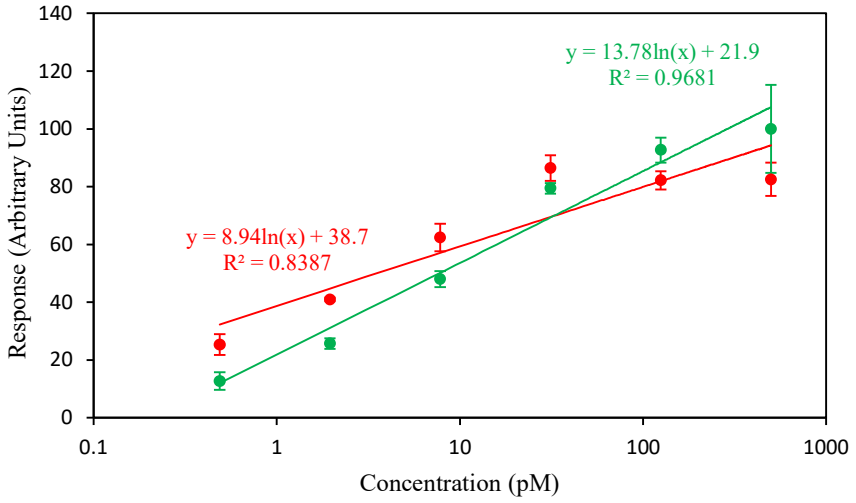


Figure 4.18: Activity data at the end of the assay (t=96h). Error bars represent sample standard deviation between triplicates.

Figure 4.18 shows the response vs concentration data points at the end of the experiment. Time course data for the activity assay is presented in Appendix H. Activity of G-CSF C17A samples were higher than that of Neupogen for the 2 highest concentrations tested. However, the remaining lower dosages yielded in poorer profileration with G-CSF C17A. EC_{50} values for G-CSF C17A and Neupogen were calculated to be 7.7 pM and 1.7 pM, respectively. By definition of EC_{50}, these concentration values are half the concentrations to give out the maximum response of their respective proteins. The Neupogen concentration needed to match the response of EC_{50} of G-CSF C17A was found to be 4.9 pM.

4.5 PEGylation of G-CSF C17A

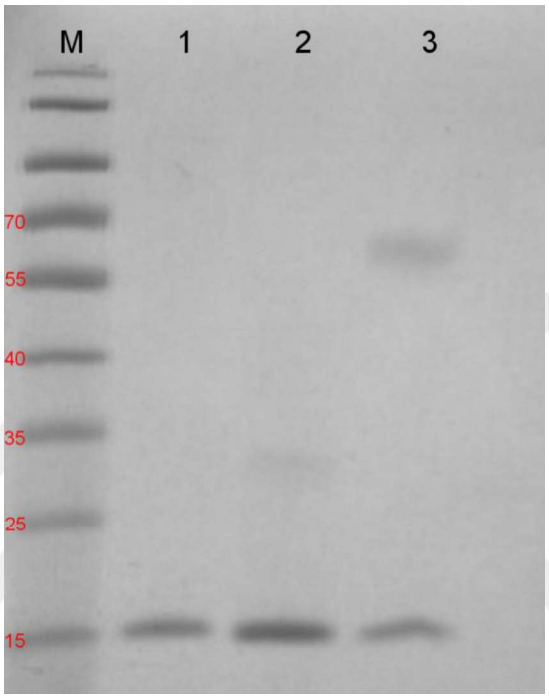


Figure 4.19: PEGylation reaction mixture on SDS-PAGE gel. M:Pageruler™ Prest. Prot. Ladder, Lane 1: G-CSF standard, Lane 2: Non-PEGylated protein as negative control, Lane 3: PEGylation reaction mixture.

Comparing band intensities between PEGylated and non-PEGylated samples show a 37% conversion ratio of G-CSF C17A to PEG20kDa-G-CSF C17A (Figure 4.19). Molecular size of PEGylated G-CSF C17A is between 55kDa and 70 kDa. This is consistent with Zhai et al., 2009's result, where monoPEGylated G-CSF shows an apparent molecular weight between 43.0 and 66.2 kDa. The higher apparent molecular weight as opposed to theoretical (38.6 kDa) may be explained by molecular interactions between SDS and PEG moieties. The variance in molecular weight observed by the diffuse appearance of PEG-protein band in Lane 4 of figure

4.19 may also be related to this phenomenon. Furthermore, the variance of size of purchased PEG may also have contributed to this issue.

CHAPTER 5

CONCLUSION

An *E. coli* strain for the intracellular production of a G-CSF analog was developed. The correctness of expression was tested with genetic expression methods and Western Blot with G-CSF antibodies.

A chemically defined medium was selected from literature and replicated with minor differences in laboratory. A shaker experiment was conducted with this medium, which resulted in 2.8 g/L cell concentration. Glycerol was compared to glucose in a shake flask experiment, and was found to be yielding in similar cell growth properties. To avoid acetic acid formation associated with *E. coli* growth on glucose and due to inexpensiveness of glycerol on industrial scale, glycerol was chosen as carbon source.

Two bioreactor experiments were conducted without and with IPTG induction. Maximum cell concentration at the end of bioreactor trials reached 58 g/L and 53 g/L in these experiments, respectively. Final His-tagged G-CSF C17A concentrations in bioreactor experiments were 2.6 g/L in uninduced bioreactor culture and 4.9 g/L in IPTG induced bioreactor culture.

Produced Histag G-CSF C17A inclusion bodies were solubilized in high urea, purified with Co affinity chromatography and renatured in a two-step dialysis scheme. Renatured proteins were treated with Factor Xa for the removal of Histag. GCSF C17A with 81% protein purity was attained and tested for correct disulfide bridge formation.

Purified proteins were tested for bioactivity *in vitro* on M-NFS-60 cells. While EC_{50} of the G-CSF C17A was found to be 2-3 times that of originator product, the maximum activity of GCSF C17A at saturated concentrations was slightly higher.

G-CSF C17A was PEGylated in a 50 µL reaction mixture to establish proof of concept. A 37% reaction conversion of GCSF C17A to PEG20kDa-GCSF C17A is achieved in PEGylation reaction.

RECOMMENDATIONS

PEGylated G-CSF C17A was not tested for bioactivity within the body of this work. The purpose for PEGylation is to increase *in vivo* half-life. A suitable mouse model of neutropenia can be used to test *in vivo* bioactivity.

IMAC separation is very useful in the sense that it provides a single step separation process. However the inclusion of histidine tag is a burden on cell metabolism. A combination of ion-exchange, size exclusion and/or hydrophobic interaction chromatography can be adopted to explore more industrially viable G-CSF separation processes.

REFERENCES

Angardi, V. (2011). Bioprocess development for thermostable glucose isomerase production. PhD Thesis, Middle East Technical University, Turkey.

Bensinger, W. I., Weaver, C. H., Appelbaum, F. R., Rowley, S., Demirer, T., Sanders, J., … Buckner, C. D. (1995). Transplantation of allogeneic peripheral blood stem cells mobilized by recombinant human granulocyte colony-stimulating factor [see comments]. *Blood*, *85*(6), 1655–1658.

Bowden, G. A., Paredes, A. M., & Georgiou, G. (1991). Structure and morphology of protein inclusion bodies in Escherichia coli. *Bio/technology (Nature Publishing Company)*, *9*(8), 725–730. https://doi.org/10.1038/nbt0891-725

Cox, G. N., Smith, D. J., Carlson, S. J., Bendele, A. M., Chlipala, E. A., & Doherty, D. H. (2004). Enhanced circulating half-life and hematopoietic properties of a human granulocyte colony-stimulating factor/immunoglobulin fusion protein. *Experimental Hematology*, *32*(5), 441–449. https://doi.org/10.1016/j.exphem.2004.01.012

Crick, F. H. C. (1968). The origin of the genetic code. *Journal of Molecular Biology*, *38*(3), 367–379. https://doi.org/10.1016/0022-2836(68)90392-6

Çalik, P., Yilgör, P., & Demir, A. S. (2006). Influence of controlled-pH and uncontrolled-pH operations on recombinant benzaldehyde lyase production by Escherichia coli. *Enzyme and Microbial Technology*, *38*(5), 617–627. https://doi.org/10.1016/j.enzmictec.2005.07.029

Çelik, E. (2008). Bioprocess development for therapeutical protein production. PhD Thesis, Middle East Technical University, Turkey.

Deotare, U., Al-Dawsari, G., Couban, S., & Lipton, J. H. (2015). G-CSF-primed bone marrow as a source of stem cells for allografting: revisiting the concept. *Bone Marrow Transplantation*, *50*(9), 1150–1156. https://doi.org/10.1038/bmt.2015.80

Eiteman, M. A., & Altman, E. (2006). Overcoming acetate in Escherichia coli recombinant protein fermentations. *Trends in Biotechnology*, *24*(11), 530–536. https://doi.org/10.1016/j.tibtech.2006.09.001

Gaberc-Porekar, V., & Menart, V. (2001). Perspectives of immobilized-metal affinity chromatography. *Journal of Biochemical and Biophysical Methods*, *49*(1–3), 335–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11694288

Ishikawa, M., Iijima, H., Satake-Ishikawa, R., Tsumura, H., Iwamatsu, A., Kadoya, T., … et al. (1992). The substitution of cysteine 17 of recombinant human G-CSF with alanine greatly enhanced its stability. *Cell Struct. Funct.*, *17*(1), 61–65. https://doi.org/10.1247/csf.17.61

Jeong, K. J., & Lee, S. Y. (2001). Secretory production of human granulocyte colony-stimulating factor in Escherichia coli. *Protein Expression and Purification*, *23*(2), 311–318. https://doi.org/10.1006/prep.2001.1508

Jevševar, S., Gaberc-Porekar, V., Fonda, I., Podobnik, B., Grdadolnik, J., & Menart, V. (2005). Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. *Biotechnology Progress*, *21*(2), 632–639. https://doi.org/10.1021/bp0497839

Jiang, Y., Jiang, W., Qiu, Y., & Dai, W. (2011). Effect of a structurally modified human granulocyte colony stimulating factor, G-CSFa, on leukopenia in mice and monkeys. *Journal of Hematology & Oncology*, *4*(1), 28. https://doi.org/10.1186/1756-8722-4-28

Jonasson, P., Liljeqvist, S., Nygren, P.-A. A., Stahl, S., & Ståhl, S. (2002). Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli. *Biotechnol Appl Biochem*, *35*(Pt 2), 91–105. https://doi.org/10.1042/BA20010099

Karimi, Z., Nezafat, N., Negahdaripour, M., Berenjian, A., Hemmati, S., & Ghasemi, Y. (2015). The effect of rare codons following the ATG start codon on expression of human granulocyte-colony stimulating factor in Escherichia coli. *Protein Expression and Purification*, *114*, 108–114. https://doi.org/10.1016/j.pep.2015.05.017

Kinstler O.B., Gabriel E. G., Farrar C.E., DePrince R.B. (1998). N-terminally chemically modified protein compositions and methods. *U.S. Patent No: U.S. 5824784A*. U.S. Patent Office.

Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A., & Deckwer, W. D. (1995). Simple fed-batch technique for high cell density cultivation of Escherichia coli. *Journal of Biotechnology*, *39*(1), 59–65. https://doi.org/10.1016/0168-1656(94)00143-Z

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680–685. http://doi.org/10.1038/227680a0

Layton, J. E., Morstyn, G., Fabri, L. J., Reid, G. E., Burgess, A. W., Simpson, R. J., & Nice, E. C. (1991). Identification of a functional domain of human granulocyte colony-stimulating factor using neutralizing monoclonal antibodies. *Journal of Biological Chemistry*, *266*(35), 23815–23823.

Menzella, H. G. (2011). Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli. *Microbial Cell Factories*, *10*(1), 15. https://doi.org/10.1186/1475-2859-10-15

Mire-sluis, A. R., Gaines, R., & Thorpe, R. (1995). The international standard for granulocyte-macrophage colony stimulating factor (GM-CSF) Evaluation in an international collaborative study, *179*(94), 127–135.

Molineux, G., Kinstler, O., Briddell, B., Hartley, C., McElroy, P., Kerzic, P., … Schwab, G. (1999). A new form of Filgrastim with sustained duration in vivo and enhanced ability to mobilize PBPC in both mice and humans. *Experimental Hematology*, *27*(12), 1724–1734. https://doi.org/10.1016/S0301-472X(99)00112-5

Nicola, N. A. (1987) Granulocyte Colony-Stimulating Factor and Differentiation-Induction in Myeloid Leukemic Cells. *Journal of Cell Cloning*. 5:1-15.

Oheda, M., Hase, S., Ono, M., Ikenaka, T. (1988) Structures of the Sugar Chains of Recombinant Human Granulocyte-Colony-Stimulating Factor Produced by Chinese Hamster Ovary Cells. *Journal of Biochemistry, 103(3),* 544-546.

Oheda, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T., Tomonou, K., … Ochi, N. (1990). O-Linked Sugar Chain of Human Granulocyte Colony-stimulating Factor Protects It against Polymerization and Denaturation Allowing It to Retain Its Biological Activity. *Journal of Biological Chemistry*, *265*(*20)*, 11432–11435.

Porath, J., Carlsson, J., Olsson, I., & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, *258*(5536), 598–599. https://doi.org/10.1038/258598a0

Quispe, C. A. G., Coronado, C. J. R., & Carvalho, J. A. (2013). Glycerol: Production, consumption, prices, characterization and new trends in combustion. *Renewable and Sustainable Energy Reviews*, *27*, 475–493. https://doi.org/10.1016/j.rser.2013.06.017

Raso, S. W., Abel, J., Barnes, J. M., Maloney, K. M., Pipes, G., Treuheit, M. J., … Brems, D. N. (2005). Aggregation of granulocyte-colony stimulating factor in vitro involves a conformationally altered monomeric state. *Protein Science : A Publication of the Protein Society*, *14*(9), 2246–57. https://doi.org/10.1110/ps.051489405

Riesenberg, D., Schulz, V., Knorre, W. I., Pohl, H., Korz, D., Sanders, E., … Gesellscho, G. (1991). High cell density cultivation of Escherichia coli at controlled specific growth rate. *Journal of Biotechnology*, *20*, 17–28. https://doi.org/10.1016/0168-1656(91)90032-Q

Sambrook, J., Russell, D.W. (2001). Molecular Cloning, A laboratory manual Vol. 1, 2, 3, 3rd Edition, Cold Spring Harbor, USA.

Seeger, A., Schneppe, B., McCarthy, J. E. G., Deckwer, W. D., & Rinas, U. (1995). Comparison of temperature- and isopropyl-beta-d-thiogalacto-pyranoside-induced synthesis of basic fibroblast growth factor in high-cell-density cultures of recombinant Escherichia coli. *Enzyme and Microbial Technology*, *17*(10), 947–953. https://doi.org/10.1016/0141-0229(94)00123-9

Shirafuji N, Asano 5, Matsuda 5, Watari K,Takaku F,Nagata 5: A new bioassay for human granulocyte colony-stimulatingfactor (hg-CSF) using murine mycloblastic NFS-60 cells astargets and estimation of its levels in sara from normal healthypersons and patients with infections and hematological disorders.Exp Hematol 17:1 16, 1988

Souza, L. M., Boone, T. C., Gabrilove, J., Lai, P. H., Zsebo, K. M., Murdock, D. C., … Chen, K. K. (1986). Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells. *Science (New York, N.Y.)*, *232*(4746), 61–65. https://doi.org/10.1126/science.2420009

Souza, L.M. (1991). Production of human pluripotent granulocyte colony-stimulating factor. *U.S. Patent No: US 4999291A.* U.S. Patent Office.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113–130. https://doi.org/10.1016/0022-2836(86)90385-2

Taylor R. G., Walker D. C., McInnes, R. R. (1993). *E.coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research.* 21(7) 1677-1678.

Toksöz, A., Yenice, i., Üzgün, M., Özer F. (2011). A novel process for preparing G-CSF (granulocyte colony stimulating factor). *European Patent No: EP 2341061 A1.* European Patent Office.

Vanz, A. L., Renard, G., Palma, M. S., Chies, J. M., Dalmora, S. L., Basso, L. A., … Tydeman, M. (2008). Human granulocyte colony stimulating factor (hG-CSF): cloning, overexpression, purification and characterization. *Microbial Cell Factories*, *7*(1), 13. https://doi.org/10.1186/1475-2859-7-13

Wadhwa, M., Bird, C., Hamill, M., Heath, A. B., Matejtschuk, P., & Thorpe, R. (2011). The 2nd International Standard for human granulocyte colony stimulating factor. *Journal of Immunological Methods*, *367*(1–2), 63–69. https://doi.org/10.1016/j.jim.2011.02.005

Weinstein, Y., Ihle, J. N., Lavu, S., & Reddy, E. P. (1986). Truncation of the c-myb gene by a retroviral integration in an interleukin 3-dependent myeloid leukemia cell line. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(July), 5010–5014. https://doi.org/10.1073/pnas.83.14.5010

Welte, K. (2014). G-CSF: filgrastim, lenograstim and biosimilars. *Expert Opin. Biol. Ther.*, *14*(7), 983–993. https://doi.org/10.1517/14712598.2014.905537

Wood, W. B. (1966). Host specificity of DNA produced by Escherichia coli: Bacterial mutations affecting the restriction and modification of DNA. *Journal of Molecular Biology*, *16*(1), 118–IN3. https://doi.org/10.1016/S0022-2836(66)80267-X

Zhai, Y., Zhao, Y., Lei, J., Su, Z., & Ma, G. (2009). Enhanced circulation half-life of site-specific PEGylated rhG-CSF: Optimization of PEG molecular weight. *Journal of Biotechnology*, *142*(3–4), 259–266. https://doi.org/10.1016/j.jbiotec.2009.05.012

Zink, T., Ross, a, Lüers, K., Cieslar, C., Rudolph, R., & Holak, T. a. (1994). Structure and dynamics of the human granulocyte colony-stimulating factor determined by NMR spectroscopy. Loop mobility in a four-helix-bundle protein. *Biochemistry*, *33*(28), 8453–8463. https://doi.org/10.1021/bi00194a009

Zsebo, K. M., Cohen, A. M., Murdock, D. C., Boone, T. C., Inoue, H., Chazin, V. R., … Souza, L. M. (1986). Recombinant Human Granulocyte Colony Stimulating Factor: Molecular and Biological Characterization. *Immunobiol.*, *172*(3–5), 175–184. https://doi.org/10.1016/S0171-2985(86)80097-3

APPENDICES

APPENDIX A

BUFFERS AND STOCK SOLUTIONS

All buffers were prepared with UltraPure water.

10X Factor Xa Buffer	200 mM Tris-HCl, 100 mM NaCl, 2 mM $CaCl_2$, pH=8.0. Solution was divided into aliquots and stored at -20°C.
100X Batch Trace Minerals Solution	70.5 mg of EDTA, 12.5 mg of CoCl2.6H2O, 75 mg of MnCl2.4H2O, 13.9 mg of CuSO4.5H2O, 15 mg of H3BO3, 10.5 mg of Na2MoO4.2H2O, 105 mg of ZnCl2 and 556 mg of FeCl3.6H2O were solubilized in 50 mL water. Solution was filter-sterilized and stored at 4°C.

100X Feeding Trace Minerals Solution	47.8 mg of EDTA, 14.7 mg of CoCl2.6H2O, 86 mg of MnCl2.4H2O, 8.5 mg of CuSO4.5H2O, 17.2 mg of H3BO3, 14.7 mg of Na2MoO4.2H2O, 36.5 mg of ZnCl2 and 162 mg of FeCl3.6H2O were solubilized in 50 mL water. Solution was filter-sterilized and stored at 4°C.
Denaturing Equilibrium/Wash Buffer	50 mM Na_3PO_4, 300 mM NaCl, 7.5M urea, pH = 7.4, prepared freshly
Denaturing Elution Buffer	50 mM Na_3PO_4, 300 mM NaCl, 7.5M urea, 150 mM imidazole, pH = 7.4, prepared freshly
Inclusion Body Wash Buffer	2.5M Urea, prepared freshly
Inclusion Body Solubilization Buffer	7.5M Urea, prepared freshly
STET Buffer	100 mM NaCl , 10 mM Tris-HCl, 1 mM EDTA, 5% (w/v) Triton X-100, pH=8.0
50X TAE Buffer	Dissolve 242 g of Tris base in 850 mL water. Add 57.1 mL of glacial acetic acid and add 14.6 g of EDTA. pH=8.3

APPENDIX B

COLUMN PURIFICATION PROTOCOL

1. Resin was thoroughly resuspended and transeferred to a centrifuge of at least 10 times larger volume of the resin.

2. Resin was pelleted at 700xg at 4°C for 2 minutes. Supernatant was discarded.

3. 10 bed volumes of denaturing equilibration/wash buffer was added to the pellet, and the solution was mixed until suspension.

4. The mixture was centrifuged at 700xg at 4°C for 2 minutes. Supernatant was discarded.

5. Steps 3 and 4 was repeated.

6. The solubilized sample was added to the resin, paying attention to not exceeding resin capacity. Resulting mixture was incubated on an end-to-end platform at 2-8°C for 1 hour.

7. The resin (and bound proteins) were pelleted by centrifuging at 700 x g at 4°C for 5 minutes.

8. Supernatant was removed. Resin was resuspended in 10 bed volumes of denaturing equilibrium/wash buffer.

9. Resulting mixture was incubated on an end-to-end platform at 4°C for 10 minutes. The resin is pelleted at 700 x g for 5 minutes at 4°C.

10. Steps 8,9 were repeated.

11. Supernatant was removed. One bed volume of denaturing equilibrium/wash buffer was added to the pellet.

12. The resin suspension was transeffered to a column and was allowed to settle out of suspension. End cap of column was removed and excess buffer was allowed to drain.

13. Column was washed with 5 bed volumes of denaturing equilibrium/wash buffer.

14. His-tagged G-CSF C17A was eluted by adding 5 bed volumes of denaturing elution buffer. Eluates were collected in 0.5 bed volume microfuge tubes.

15. Eluates were stored at 4°C until confirmation of purity and concentration with SDS-PAGE.

APPENDIX C

DIALYSIS TUBING PREPARATION PROTOCOL

1. Required amount of tubing was cut and clamped on both ends.

2. Glycerol was removed from tubing by washing under tap water for 3-4 hours. Tubing was declamped.

3. The tubing was treated in a 0.3%(w/v) sodium sulfide solution at 80°C for 1 minute.

4. Tubing was washed with hot water at 60°C for 2 minutes.

5. Tubing was treated wih 0.2%(v/v) sulfuric acid in water.

6. Tubing was rinsed with hot water to remove any acid.

7. Tubing was clamped, used for dialysis, and discarded.

APPENDIX D

PROTOCOLS FOR FACTOR XA DIGESTION AND FACTOR XA REMOVAL

Factor Xa Digestion

1. 10X Factor Xa buffer and protein sample is mixed in a 100 to 400 µL reaction mixture in a 1:9 volumetric ratio.

2. Factor Xa (2U/µL) is added in a concentration of 25 µL/mL.

3. Reaction mixture is incubated at 15-25°C in a water bath

Factor Xa Removal

1. Resin is resuspended by inversion. 50 µL bed volume (100 µL slurry) per 4 Units of Factor Xa is transferred to a centrifuge tube.

2. The resin is centrifuged at 1000 x g for 5 minutes. Supernatant is discarded.

3. Resin is resuspended in 10 bed volumes 1X Factor Xa buffer, recentrifuged at 1000 x g for 5 minutes, and its supernatant is discarded.

4. Factor Xa reaction mixture is applied to the resin. Mixture is incubated for 10-15 minutes on an end-to-end incubator.

5. The reaction is centrifuged at 1000 x g for 4 minutes. The supernatant is removed with a 1 mL syringe.

APPENDIX E

OPTICAL DENSITY CALIBRATION CURVES

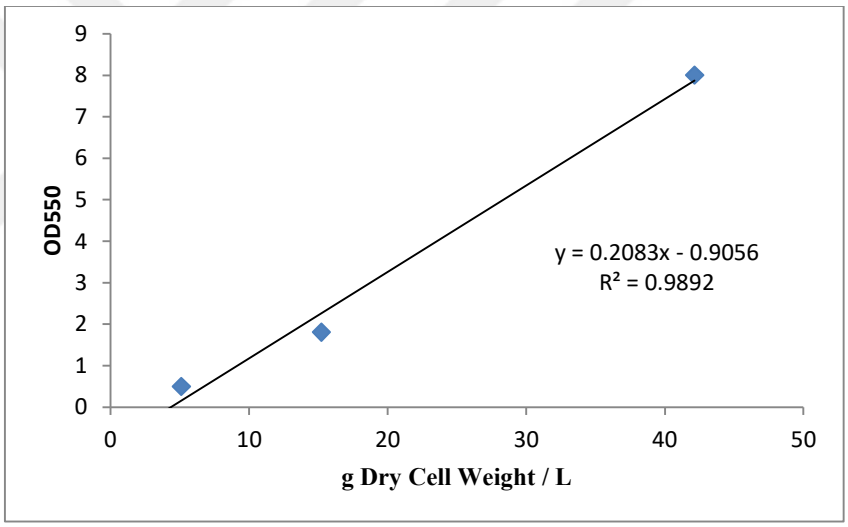


Figure E.1: Calibration curve for dry cell weight *E. coli* BL21 (DE3) pRSET A:: GCSF C17A cells at 550 nm.

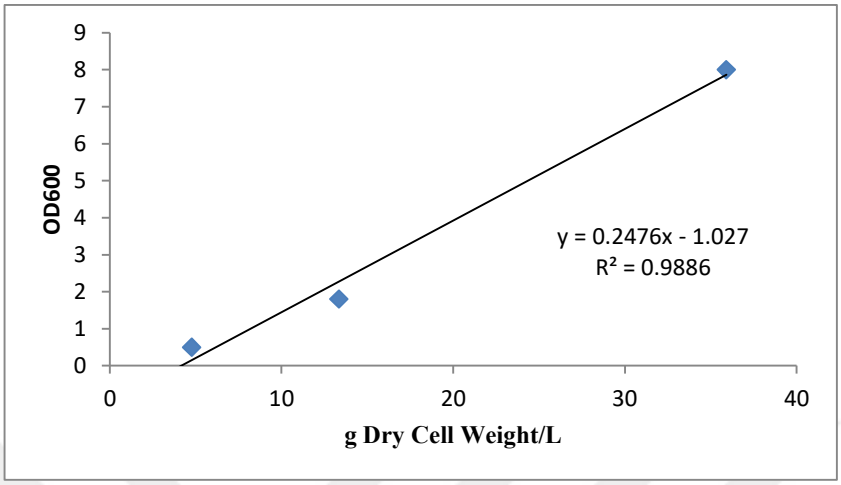


Figure E.2: Calibration curve for dry cell weight *E. coli* BL21 (DE3) pRSET A:: GCSF C17A cells at 600 nm.

APPENDIX F

NUCLEOTIDE SEQUENCES

Nucleotide sequence for pRSET A vector, 2897 base pairs:

GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCC
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCGGGGTTCTCAT
CATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGAT
CTGTACGACGATGACGATAAGGATCGATGGGGATCCGAGCTCGAGATCTGCAGCTGG
TACCATGGAATTCGAAGCTTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGT
TGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG
TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATCTGGCGTAATAGCGA
AGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGAC
CGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCT
CGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTT
CCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTC
ACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCAC
GTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGT
CTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGA
GCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTA
GGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATA
CATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATAT
TGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGAT
GCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGT
AAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAA
GTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGT
CGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAG
CATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGT
GATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACC
GCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAG

CTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCA
ACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAA
TTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTT
CCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGT
ATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACG
ACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCC
TCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATT
GATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAAT
CTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTA
GAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG
CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTT
CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATAC
CTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT
ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAC
CTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG
TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGA
AACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCC
TTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTA
TCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC
CGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCA
ATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAG

Nucleotide sequence for pRSET A:: 6xHistag-GCSF C17A, 3322 base pairs:

GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCC
TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATA**CATATG**CATCACCATCAC
CATCAC**ATTGAAGGCCGTACTCCTCTGGGTCCAGCTTCCTCTCTTCCGCAGTCGTTC**
CTGTTGAAAGCTCTTGAGCAGGTTCGCAAAATCCAAGGTGACGGCGCTGCACTGCAA
GAAAAACTGTGTGCGACCTATAAACTGTGCCATCCGGAAGAACTGGTGTTACTCGGT
CACAGTCTGGGTATCCCGTGGGCTCCGCTGAGCAGTTGCCCGTCCCAAGCTCTTCAG
CTGGCTGGTTGCTTATCGCAACTCCATTCAGGTTTGTTTTTATATCAGGGTTTACTG
CAAGCACTGGAAGGCATCAGCCCGGAACTGGGTCCTACCTTGGACACGCTCCAGCTG
GACGTTGCTGACTTCGCCACGACCATCTGGCAGCAAATGGAGGAATTAGGCATGGCT

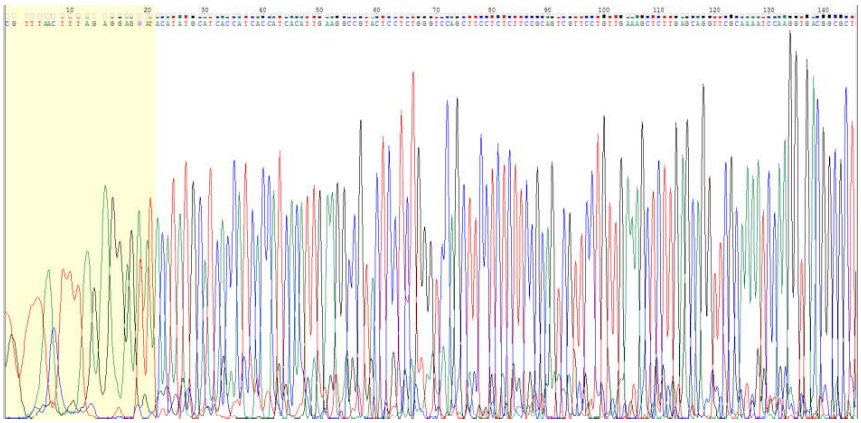
CCGGCGCTGCAACCGACCCAGGGCGCCATGCCAGCATTTGCGAGCGCTTTTCAACGC
CGTGCGGGTGGCGTTTTAGTCGCGAGCCACTTACAATCGTTCCTGGAAGTGTCTTAT
CGCGTCCTTCGCCATCTTGCACAGCCGTAATAAGAATTCGAAGCTTGATCCGGCTGC
TAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGC
ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAAC
TATATCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAG
TTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCG
GGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCT
CCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCT
CTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTT
TTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACT
GGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCG
ATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTT
AACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAA
CCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAAT
AACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATT
TCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACC
CAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAAC
GTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTA
TTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGG
TTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAA
CGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA
CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG
ACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAC
TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTG
CAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTG
GAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGC
CCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAA
ATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACC
AAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGA
TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTT
CGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTT
TTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG
TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA
GAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCA
AGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGG

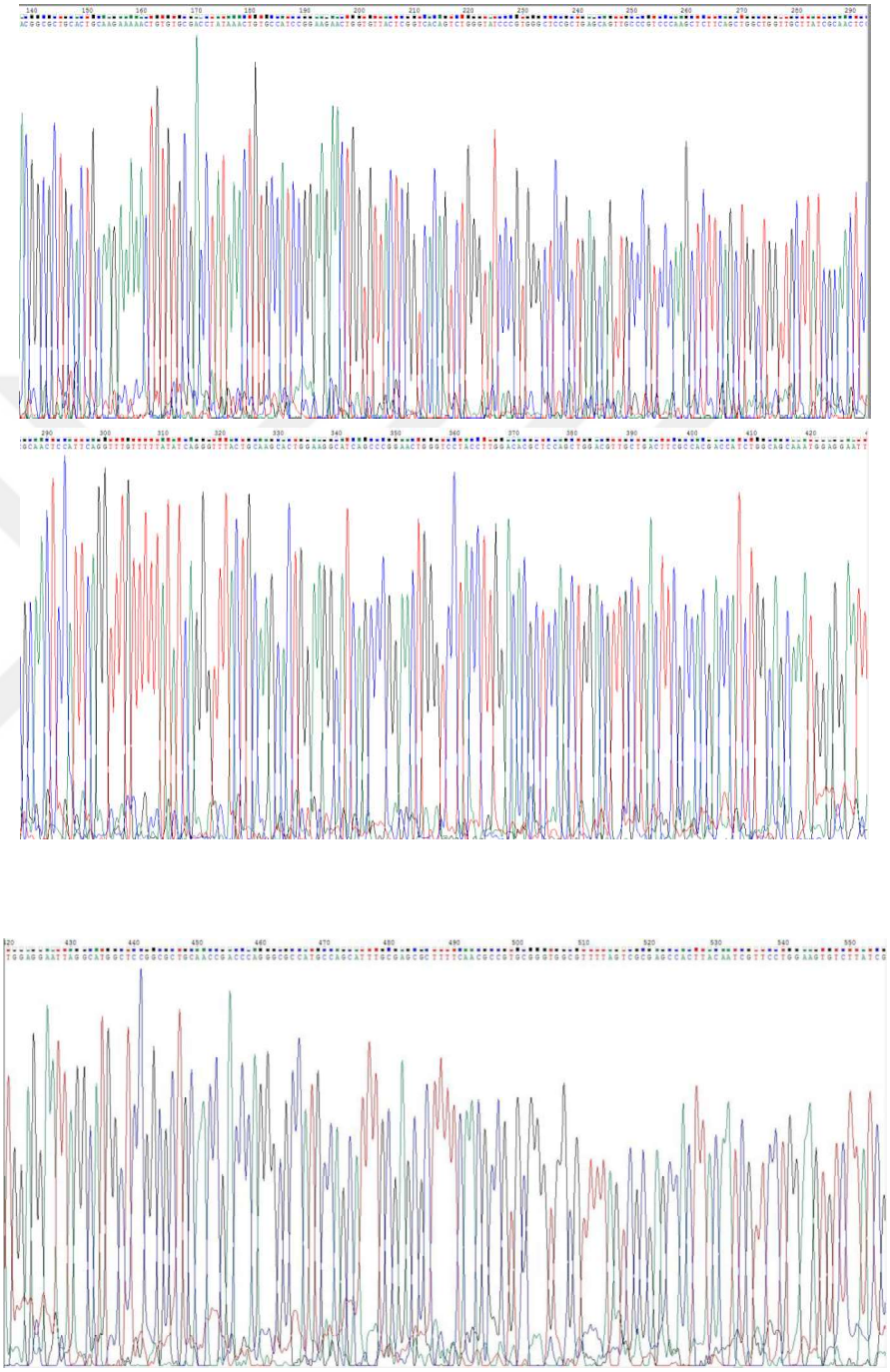
ATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGC
GAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAG
AGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT
TTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCC
TATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCG
CCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAG
TGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGC
CGATTCATTAATGCAG

His tag	**Factor Xa Processed Protein**
Factor Xa Site	NdeI site
pRSET A	EcoRI site

Figure F.1: Legend for nucleotide sequence.

Sequencing data visualized in Chromas:





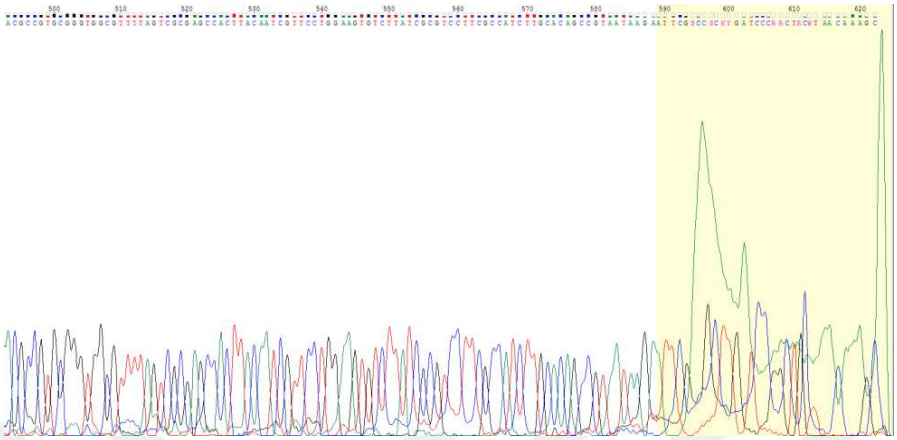


Figure F.2: Sequencing Data for G-CSF C17A gene cloned into pRSET A vector.

```
Amplicate  1    ACATATGCATCACCATCACCATCACATTGAAGGCCGTACTCCTCTGGGTCCAGCTTCCTC  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       58   ACATATGCATCACCATCACCATCACATTGAAGGCCGTACTCCTCTGGGTCCAGCTTCCTC  117

Amplicate  61   TCTTCCGCAGTCGTTCCTGTTGAAAGCTCTTGAGCAGGTTCGCAAAATCCAAGGTGACGG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       118  TCTTCCGCAGTCGTTCCTGTTGAAAGCTCTTGAGCAGGTTCGCAAAATCCAAGGTGACGG  177

Amplicate  121  CGCTGCACTGCAAGAAAAACTGTGTGCGACCTATAAACTGTGCCATCCGGAAGAACTGGT  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       178  CGCTGCACTGCAAGAAAAACTGTGTGCGACCTATAAACTGTGCCATCCGGAAGAACTGGT  237

Amplicate  181  GTTACTCGGTCACAGTCTGGGTATCCCGTGGGCTCCGCTGAGCAGTTGCCCGTCCCAAGC  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       238  GTTACTCGGTCACAGTCTGGGTATCCCGTGGGCTCCGCTGAGCAGTTGCCCGTCCCAAGC  297

Amplicate  241  TCTTCAGCTGGCTGGTTGCTTATCGCAACTCCATTCAGGTTTGTTTTTATATCAGGGTTT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       298  TCTTCAGCTGGCTGGTTGCTTATCGCAACTCCATTCAGGTTTGTTTTTATATCAGGGTTT  357

Amplicate  301  ACTGCAAGCACTGGAAGGCATCAGCCCGGAACTGGGTCCTACCTTGGACACGCTCCAGCT  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       358  ACTGCAAGCACTGGAAGGCATCAGCCCGGAACTGGGTCCTACCTTGGACACGCTCCAGCT  417

Amplicate  361  GGACGTTGCTGACTTCGCCACGACCATCTGGCAGCAAATGGAGGAATTAGGCATGGCTCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       418  GGACGTTGCTGACTTCGCCACGACCATCTGGCAGCAAATGGAGGAATTAGGCATGGCTCC  477

Amplicate  421  GGCGCTGCAACCGACCCAGGGCGCCATGCCAGCATTTGCGAGCGCTTTTCAACGCCGTGC  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       478  GGCGCTGCAACCGACCCAGGGCGCCATGCCAGCATTTGCGAGCGCTTTTCAACGCCGTGC  537

Amplicate  481  GGGTGGCGTTTTAGTCGCGAGCCACTTACAATCGTTCCTGGAAGTGTCTTATCGCGTCCT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Gene       538  GGGTGGCGTTTTAGTCGCGAGCCACTTACAATCGTTCCTGGAAGTGTCTTATCGCGTCCT  597

Amplicate  541  TCGCCATCTTGCACAGCCGTAATAAGA  567
               ||||||||||||||||||||||||||
Gene       598  TCGCCATCTTGCACAGCCGTAATAAGA  624
```

Figure F.3: BLASTn result, aligning nucleotide sequencing data to intended sequence.

APPENDIX G

SDS-PAGE GELS

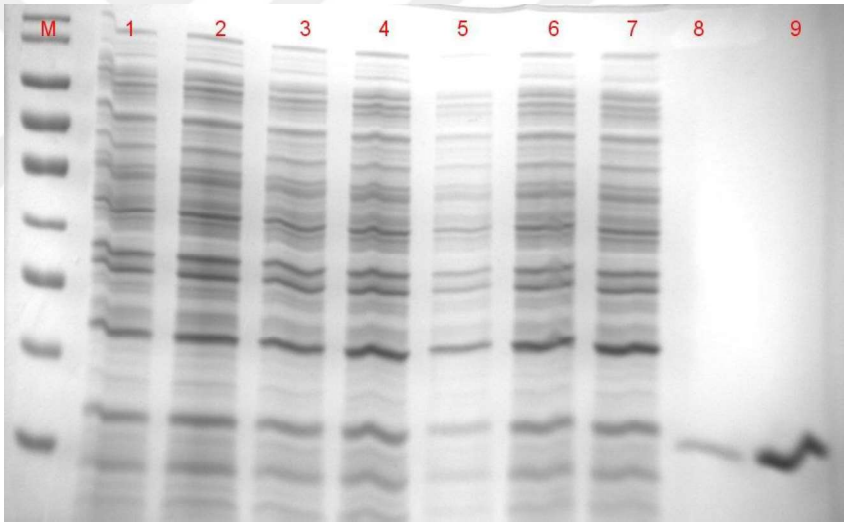


Figure G.1: SDS-PAGE gel for BR-1. M: PageRuler™ Prest. Prot. Ladder, Lane 1: 0h, Lane 2h, Lane 3: 4h, Lane 4: 6h, Lane 5: 10h, Lane 6: 8h, Lane 7: 12h, Lane 8: Neupogen 25 mg/L, Lane 9: Neupogen 150 mg/L.

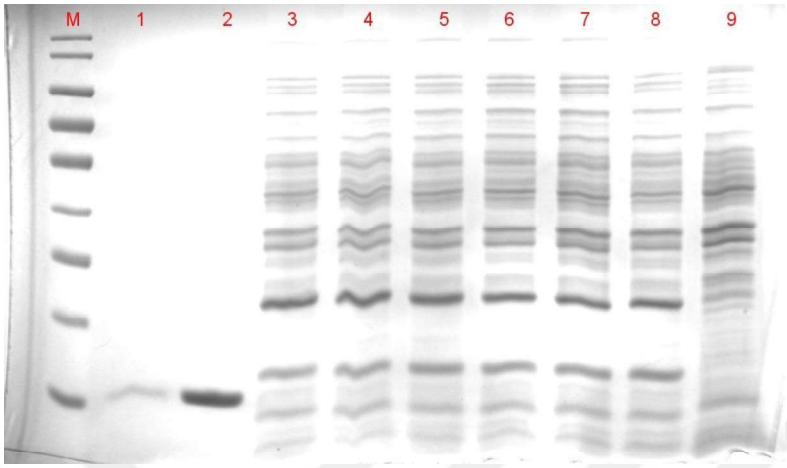


Figure G.2: SDS-PAGE gel for BR-1. M: PageRuler™ Prest. Prot. Ladder, Lane 1: Neupogen 25 mg/L, Lane 2: Neupogen 150 mg/L, Lane 3: 14h, Lane 4: 16h, Lane 5: 18h, Lane 6: 20h, Lane 7: 22h, Lane 8: 24h, Lane 9: Negative control.

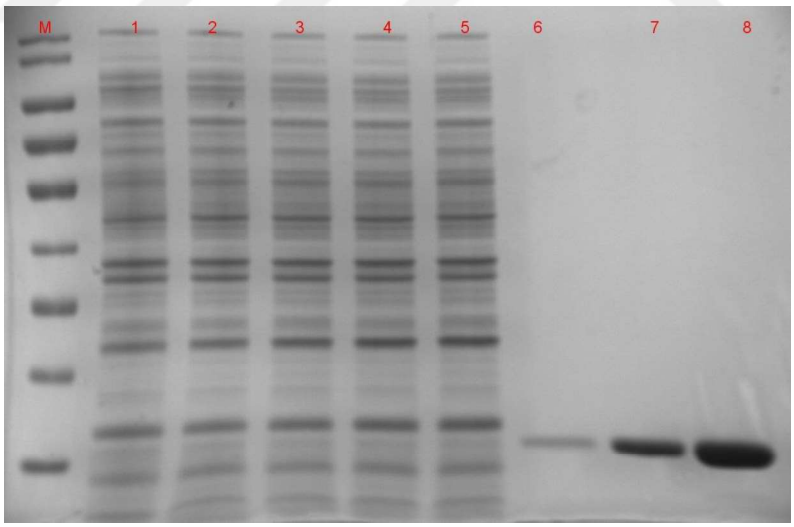


Figure G.3: SDS-PAGE gel for BR-2. M: PageRuler™ Prest. Prot. Ladder, Lane 1: 0h, Lane 2: 3h, Lane 3: 6h, Lane 4: 9h, Lane 5: 12h, Lane 6: Neupogen 25 mg/L, Lane 7: Neupogen 150 mg/L., Lane 8: Neupogen 600 mg/L.

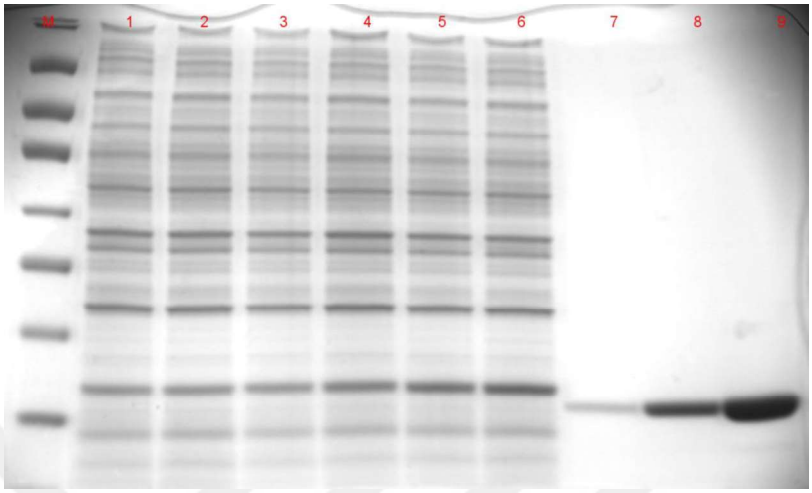


Figure G.4: SDS-PAGE gel for BR-2. M: PageRuler™ Prest. Prot. Ladder, Lane 1: 15h, Lane 2: 15.5h, Lane 3: 16.5h, Lane 4: 17.5h, Lane 5: 18.5h, Lane 6: 19.5h, Lane 6: Neupogen 25 mg/L, Lane 79: Neupogen 150 mg/L., Lane 8: Neupogen 600 mg/L.

APPENDIX H

BIOACTIVITY DATA

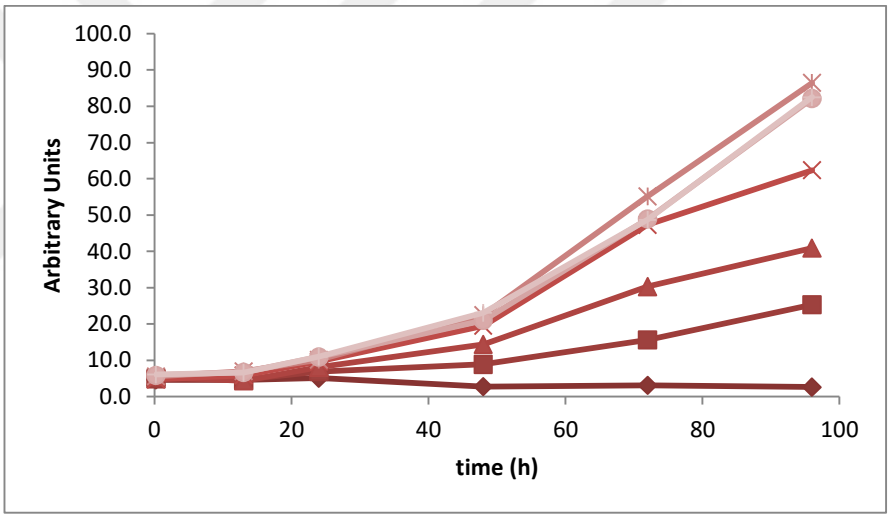


Figure H.1: Dose and time response of Neupogen®. Data points signify average of triplicates.

0 pM 0.49 pM 1.95 pM 7.81 pM 31.2 pM 125 pM 500 pM

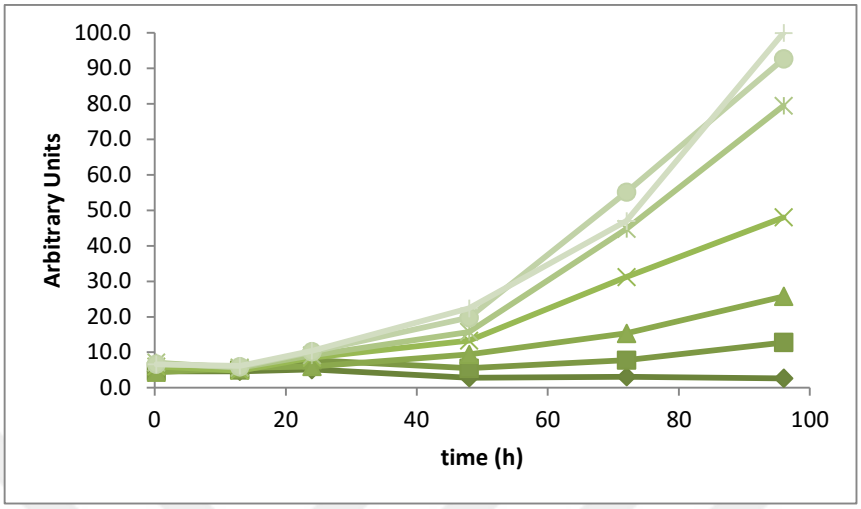


Figure H.2: Dose and time response of G-CSF C17A. Data points signify average of triplicates.

0 pM 0.49 pM 1.95 pM 7.81 pM 31.2 pM 125 pM 500 pM