CLONING AND EXPRESSION OF SYNTHETIC HUMAN ERYTHROPOIETIN IN METHYLOTROPHIC YEAST, *Pichia pastoris*

by

YAZMIN BUSTAMI

Thesis submitted in fulfillment of the
requirement for the degree
Master of Science

November 2010

ACKNOWLEDGEMENTS

Alhamdulillah, firstly I would like to express my gratitude to my supervisor, Dr. Yahya Mat Arip and my co-supervisors, Dr. Ahmad Ramli Mohd Yahya and Assoc. Prof. Dr. Amirul Al-Ashraf Abdullah for their invaluable guidance and encouragement on my research work. I would also like to thank to Assoc. Prof. Dr. Tengku Sifzizul Tengku Muhammad and Assoc. Prof. Dr. Alexander Chong Shu Chien for their advice and timely help.

My sincere thanks to lab members from Lab 207, 218, 318 for their helps, advices and generosity in allowing me to use of various pieces of equipment vital for the completion of my studies. Also to all staff of School of Biological Sciences office for their kind help on purchasing and delivery of scientific items. My special thanks to my friends for their endless support and making my research work in USM a wonderful and enjoyable moment.

I would like to express my gratitude to USM Academic Staff Training Scheme (ASTS) and Ministry of Higher Education for the financial assistance during my master study.

Last but not least, a heartfelt appreciation to my beloved husband, Mohd Salihin and daughter, Siti Khadijah Amani for their warmth love and blessing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A base	Adenine base
AOX	Alcohol hidrogenase
BHK	Baby hamster kidney
BLAST	Basic Local Alignment Search Tool
bp	base pair
C base	cystine base
$CaCl_2$	calcium chloride
CBB	Coomasie brilliant blue
cDNA	Complimentary DNA
CH_2	methylene
CHO	chinese hamster ovary
cm	centimetre
dH_2O	distilled water
DMSO	Dimethyl sulfoxide
dNTPs	Deoxyribonuclotide triphosphate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent ass**ay**
g	gram
g	Relative centrifugal force (centrifugation)
G base	Guanine base
Glc	glucose
GlcNAc	N-acetyl glucosamine
H_3PO_4	Phosphoric acid
Hct	Hematocrit
hr	hour
huEPO	human erythropoietin
IPTG	Isopropyl-ß-D-thiogalactopyranosid
IV	intravenous
K_2HPO_4	dipotassium hydrogen phosphate
kb	kilobases
kDa	kilo dalton
KH_2PO_4	Monopotassium phosphate
KOH	potassium hydroxide
L	litre
LB	Luria Bertani
M	Molar
Man	Mannose
mg	miligram
mg/mL	miligram per mililitre
$MgCl_2$	magnesium chloride

$MgSO_4$	magnesium sulphate
min	minute
mL	mililitre
mM	mili molar
mRNA	messenger RNA
mU/mL	mili unit per mililitre
Mut^+	Methanol utilization plus
Mut^s	Methanol utilization slow
$(NH_4)_2SO_4$	Ammonium Sulphate
OD	Optical density
oligos	oligonucletides
P. pastoris	*Pichia pastoris*
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
rhuEPO	Recombinant human erythropoietin
rpm	revolutions per minute
RT-PCR	Real time - Polymerase chain reaction
RVC	Reticulated vitreous carbon
S. cerevisiae	*Saccharomyces cerevisiae*
SC	subcutaneous
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Second
T base	Thymine base
T_m	melting temperature
Taq	*Thermus aquaticus*
TEMED	N,N,N',N'- tetramethylethylenediamine
Tris	Tris[hydroxymethyl]aminomethane
Tween® 20	polyoxyethylene-sorbitan mono-laurate
U	Unit of enzyme activity
U//kg	Unit per kilogram
U/mL	Unit per militre
U/wk	Unit per week
w/v	weight per volume
α	alfa
β	beta
µM	micro Molar
X-Gal	5-bromo-4-chloro-3 indolyl- β -D-galactopyranoside

PENGKLONAN DAN PENGEKSPRESAN GEN SINTETIK ERYTHROPOIETIN MANUSIA DALAM YIS METALOTROFIK, *Pichia pastoris*

ABSTRAK

Erythropoietin manusia adalah hormon atau glikoprotein yang dihasilkan di buah pinggang dan tulang sumsum dengan fungsi fisiologikal yang penting seperti penghasilan sel darah merah, pertumbuhan sel darah baru dan penyembuh luka. Sumber asal EPO manusia diekstrak dari urin manusia (uhuEPO). Alternatifnya, rekombinan erythropoietin manusia dapat menampung bekalan uhuEPO yang terbatas. Oleh itu, fokus dalam kajian ini adalah untuk menghasilkan sintetik huEPO dan diekspres di dalam *Pichia pastoris*. Kaedah pemasangan tindak balas rantaian polymerase (PCR) digunakan untuk menghasilkan gen sintetik menggunakan dua puluh set oligonukleotida yang saling bertindih, yang terdiri daripada jujukan gen huEPO A dan enzim penyekat. Keputusan dari analisis jujukan DNA menunjukan gen sintetik huEPO dipadankan dengan tepat walaupun sedikit mutasi terhasil. Gen sintetik huEPO A yang tiada mutasi telah berjaya diperolehi menggunakan sistem mutagenesis terarah tapak. Kemudian, ia disubklonkan ke dalam pPIC9k dan ditransfomasi ke dalam strain *Pichia*, GS115. Transformasi plasmid rekombinan yang telah diluruskan telah berjaya dintegrasi di lokus *HIS4* antara plasmid rekombinan dan genom *P. pastoris* seterusnya menghasilkan tranforman *His*$^+$ Mut$^+$. Kajian pengekspresan skala kecil dijalankan dan kultur-kultur diinduksi menggunakan metanol selama tiga hari (0 hingga 72 jam) dengan selang masa 24 jam. Protein rhuEPO yang dirembeskan terhasil pada jalur 30 kDa dan disahkan dengan analisis ELISA. Berdasarkan analisis ELISA, jumlah EPO yang tertinggi dihasilkan pada 72 jam.

CLONING AND EXPRESSION OF SYNTHETIC HUMAN ERYTHROPOIETIN IN METHYLOTROPHIC YEAST, *Pichia pastoris*

ABSTRACT

Human erythropoietin is a hormone or glycoprotein produced by the kidney and bone marrow with important physiological functions, such as, erythropoiesis, angiogenesis, and wound healing. The natural source of huEpo is purified from human urine (uhuEPO). Alternatively, recombinant human erythropoietin could overcome the limited supply of uhuEPO. Thus, the focus of this study is to generate a synthetic huEPO gene and express in *Pichia pastoris*. PCR assembly method utilized to construct a synthetic gene using twenty sets of overlapping oligonucleotides, covering the huEPO A gene sequence and two newly introduced restriction enzyme sites. Results from DNA sequence analysis showed accurate assembled synthetic huEPO gene albeit minor base mutations detected. A free-mutation synthetic huEPO A gene successfully obtained using site directed mutagenesis system. Then it was sub-cloned into pPIC9k and was transformed into *Pichia* strain, GS115. Transformation of the linearized recombinant construct successfully integrated at the *HIS4* locus between the recombinant plasmid and the *P. pastoris* genome, hence, generating *His*$^+$ Mut$^+$ transformants. The small scale expression studies were carried out and the cultures were induced with methanol for three days (0 to 72 hr) with 24 hr interval time. Secreted rhuEPO protein was observed at approximately 30 kDa band and further confirmed using ELISA analysis. Based on ELISA analysis, the highest amount of Epo concentration was observed at 72 hr.

1.0 INTRODUCTION

1.1 Human Erythropoietin (huEPO)

Anemia is a life-threatening disease and one of the world's leading clinical problems. Its incident will further increasing in ageing population (Weiss and Goodnough, 2005). World Health Organization estimates that anemic patient in worldwide population will be staggering two billion and that approximately 50% of all anemia can be attributed to iron deficiency (World Health Organization, 2001). Development of anemia treatment had been carried out with many strategies and become an active research area. Realized there is a close relationship between impaired renal and anemia had prompted the researcher to study erythropoietin **(**EPO) as the predominant hematopoietic growth factor found in the renal that controls red blood cells formation. The existence postulated as early as 1906 (as cited by Carnot *et al.,* 1906 in Sytkowski, 2004) but very little information about the structure of EPO had been published because of its very limited availability.

EPO derived from anemic sheep was purified in 1971 (Goldwasser and Kung, 1971) but the amount obtained was too small. The need to produce EPO in high quantity had encouraged the pioneering work to isolate 10 mg EPO from 2550 liters of human urine (Miyake *et al.*, 1977; Jelkmann, 2000). The preparation allowed the identification of the amino acid sequence and synthesized human EPO DNA probes for the isolation and cloning of the human EPO gene from mRNA in kidney and liver which are the major site of EPO production (Jacobs *et al.,* 1985; Dame *et al.,* 1998; Jelkmann, 2000). Attempt to increase EPO production using recombinant DNA technique became a big success as in 1985, the first recombinant human EPO (rhuEPO) became available for clinical trials. It was introduced into clinical practical practice for correction of anemia of renal failure in 1989 and become available for renal patients throughout the world (Eschbach, 1994) since the use of recombinant human EPO had received United States Food and Drug Administration (US FDA) approval. RhuEPO had been produced in various cell lines; in peculiar CHO and BHK cells (Inoue *et al.,* 1995). Currently, recombinant human EPO in CHO cell line that exhibit as similar as the natural EPO is extensively used in the therapy to cure severe anemia (Didier *et al.,* 2000).

EPO becomes a breakthrough in replacement therapy area (Eschbach *et al.,* 1987) for correction of anemia and now it is not solely known for its therapeutic used in hematopoietic cells as the importance extended to various tissues and systems.

Knowing EPO multifunctions and become the top list product in biopharmaceutical market (Pavlou and Reichert, 2004) encourage researchers seek an alternative strategy to replace EPO in mammalian cells with other low cost expression system that produced higher expression level. Thus, EPO has been produced in insect cells (Quelle *et al.,* 1989), bacteria (Lee, 1984), plants (Matsumoto *et al.,* 1995; Cheon *et al.,* 2004; Weise *et al.,* 2007) and yeasts (Elliott *et al.,* 1989) that showed variation in expression level, glycosylation pattern and biological activity.

1.1.1 EPO gene and structure

EPO belongs to hematopoietic growth factor group involves in red blood cells maintenance. Hematopoietic growth factor is a family of cytokines that interact with specific receptor on hematopoietic cells (Clark and Kamen, 1987; Nicola, 1989). It is required for the survival, proliferation and differentiation of hematopietic progenitors by regulating the specific cells which they interact and further stimulate the formation of red blood cells (Clark and Kamen, 1987).

The human EPO gene present as a single copy and located on the long arm chromosome 7 in the region q11-q22 (Powell *et al.,* 1986; Law *et al.,* 1986). The restriction fragment length polymorphisms, inherited in a Mendelian fashion, have been identified (Lin, 1987). It spans approximately 2.2 kb from the ATG codon to the stop codon. EPO consists of five exons (582 base pairs) and four introns (1562 base pairs). No promoter-like sequences were originally identified and the promoter was more like that of housekeeping gene than an inducible one (Shoemaker and Mitsock, 1986). It proposed that EPO is constitutively produced, since it is never absent from plasma (Spivak and Hogans, 1987), even when there is extreme erythrocytosis (Moccia *et al.,* 1980).

It is a heavily glycosylated protein consisting 165 amino acid with calculated molecular weight is 18,398 dalton (Jacobs *et al.,* 1985; Lin *et al.,* 1985) and it migrates to 30,000 dalton (Celik *et al.,* 2007) in fully glycosylated form. The crystal structure of human EPO is shown in Figure 1.1. This hydrophobic polypeptide contain one *O*- and three *N*- linked sites, the oligosaccharide side chains or glycosylation sites appears to be required for the processing and transportation of EPO from its cell of origin (Dube *et al.,* 1988). It is also required for prolongation of its biological half life, however it is not required for receptor binding (Narhi *et al.,* 1991). The glycosylation sites for three complex type N-glycans located at

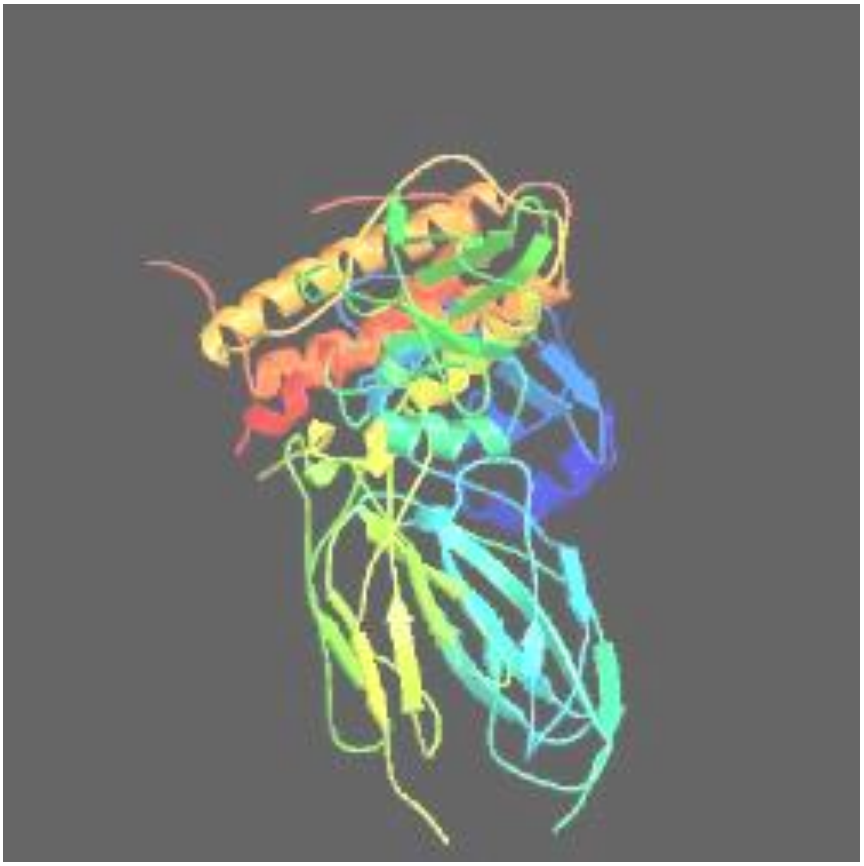

Figure 1.1 The crystal **s**tructure of EPO complexed to the extracellular ligand- binding domains of the EPO-R shown in ribbons illustration (Adapted from Syed *et al.*, 1998).

asparagine (Asn) residues at positions 24, 38 and 83 and a mucin- type- O- glycan located at ser-126 and comprise about 40% of the total mass of EPO (Celik *et al.*, 2007). The molecules contain two internal disulfide bonds, and their conformational equivalent appears to be necessary for biological activity (MacDonald *et al.*, 1986).

1.1.2 Erythropoietin (EPO) production in the cells

EPO production in the cell was linked with an early erythropoiesis developmental biology and EPO was found in the yolk sac. It is an organ comprising of two layers, extra-embryonic mesodermal cells and visceral endodermal cells. The endodermal layer is in direct contact with macromolecules derived from the maternal plasma, and these molecules along with others synthesized by the endoderm itself influence development in the mesodermal layer within these so called "blood islands" (Jollie, 1990). Lee *et al.,* (2001) examined mouse embryos for EPO and EPO-R expression in situ hybridization. They discovered EPO-R expression in the yolk sac at stage E8.0 and in the yolk sac vasculature at E9.0. However, their results were varied with Yasuda *et al.,* (2002) who carried out a very similar study on gestational day 10 embryos using in situ hybridization and immunohistochemistry. They detected EPO mRNA in 57.6% of endodermal epithelial cells by *in situ* hybridization and EPO protein in 52.8% by immunohistochemistry.

Kidney shows importance to the regulation of red blood cells thus it is known as the site production of circulating EPO. EPO is produced in the kidney in response to the oxygen tension of the blood and work as a positive feedback to increase the production of red blood cells. When it alarmed with the signal such as hypoxia or oxygen stress from cells, EPO production significantly increased as it circulates from kidney to the bone marrow where it stimulates the proliferation and differentiation of red blood cell progenitors (Chen *et al.,* 2004). This mechanism helps in increasing oxygen carrying capacity. An early immunohistochemistry studies by Fisher and coworker localized EPO to the glomerulus (Fisher *et al.,* 1965). After continuous studies carried out, the investigator detected EPO mRNA in the peritubular interstitial cells from anemic rats using a combination of *in situ* hybridization and immunohistochemistry (Bachmann *et al.,* 1993).

There is evidence indicating that EPO produced in the bone marrow, the main producer of red blood cells. Hermine *et al.,* (1991) used RT-PCR and demonstrated

an EPO PCR product in both human and murine bone marrow. They used antisense oligodeoxynucleotides and showed that downregulating either EPO or its receptor caused decreased in mixed erythroid/ non-erythroid colony formation *in vitro*. Stopka and coworkers showed that differentiating CD34+ cells expressed both EPO and EPO-R detected by RT-PCR (Stopka *et al.*, 1998). This finding suggests a potential autocrine or paracrine action of EPO within the bone marrow.

In the mammalian fetus, EPO is expressed primarily in the liver. As ontogeny continues, EPO production in liver is suppressed and kidney production becomes predominant. Two types of cells in the liver had been identified to produce EPO using *in situ* hybridization. Koury *et al.*, (1991) first studied adult mice using *in situ* hybridization and demonstrated EPO mRNA in isolated hepatocytes. Additionally, a small population of nonepithelial appearing cells in or near the sinusoids also contained EPO mRNA. Wintour *et al.*, (1996) examined the 41 day ovine fetus and identified EPO mRNA in both mesonephros and the metanephros as well as in the liver by RT-PCR. *In situ* hybridization of the mesonephros demonstrated that the EPO mRNA was present in interstitial cells lying between the proximal but not the distal tubules, similar to those cells identified in the anemic adult mouse kidney (Koury *et al.*, 1991; Lacombe *et al.*, 1988)

In order to determine the presence of EPO in the plasma, several assays have been developed. Qualitative assay was the earliest assay used to determine EPO; consist of measuring increasing number of red blood cell, reticulocyte numbers or hematocrit (Krantz *et al.*, 1970; Fisher, 1998). Followed by a method that measured EPO with the incorporation of ^{59}Fe into the circulating red blood cells of normal rats (White *et al.*, 1960). As many assays being developed, it becomes clear that there was a need for standardization. Initially the "Cobalt unit" that was defined as the amount of EPO stimulated in the fasted rat assay equal to 5 µmol of cobalt was developed (White *et al.*, 1960). Then Standard A was introduced which recognized as lyophilized Step IV sheep EPO (23 U/mg) as the crude preparation of EPO partially purified from the plasma of anemic sheep was relatively stable (Goldwasser *et al.*, 1962). Later, Standard B was developed consist of lyophilized crude human urinary EPO. It defined 1 unit of EPO that contain in 1.48 mg of the Standard protein 10 International Units (10 IU) per ampoule. This is also known as the Second International Reference Preparation (Annable *et al.*, 1972). Now, it is being replaced

by the Second International Standard 2003, which also is composed of pure recombinant EPO (120 IU/ampoule) (World Health Organization, 2003).

1.1.2.1 Regulation of EPO production

In normal condition, EPO present in a low concentration but under anemic or anoxic stress, high amount of EPO were found in the plasma and also excreted in the urine. An early study showed that the starting material to extract EPO was obtained from urine of anemic patient (Miyake *et al.*, 1977; Jelkman, 2000). There are several factors that regulate EPO production.

Hypoxia-inducible factor-1 (HIF-1) is a fundamental physiologic stimulus that causes a rapid increase in renal production of EPO through an exponential increase in the number of EPO-producing cells (Koury *et al.*, 1989). It acts as a transcription factor for a large number of hypoxia-inducible genes, including those coding for vascular endothelial growth factor, platelet-derived growth factor, glycolytic enzymes, as well as for EPO (Wang, 1993). Hypoxia is a condition when the tissues are deprived of oxygen. There are four types of hypoxia that include hypoxemic, anemic, stagnant, and histotoxic. Model for EPO production regulation is shown in Figure 1.2. It seems most likely that several transducer substances released during hypoxia (adenosine, eicosanoids, catecholamines) and reoxygenation (oxygen-derived metabolites such as hydrogen peroxide) activate adenylate cyclase to generate cAMP, which in turn activates protein kinase A. Finally it leads to the production of phosphoprotein which are involved in transcription and translation of the final 166-amino acid EPO molecule. Hydroxylation of HIF-1 by prolyl hydroxylase, which requires the presence of oxygen, predisposes HIF-1 to ubiquitination by the von Hippel-Lindau protein, followed by degradation in the proteosome. In contrast when oxygen is absent, HIF-1 is not susceptible to degradation and HIF-1 levels increase rapidly, followed by up-regulation of many hypoxia-responsive genes, including that for EPO (Prchal, 2003).

Factors other than tissue hypoxia might be involved in the regulation of EPO production or may influence serum concentration. Abnormally high EPO levels have been reported in patients with aplastic anemia (Gaines *et al.*, 1992), and dramatic changes in serum levels have been described after chemotherapy (Schapira *et al.*, 1990; Birgegard *et al.*, 1989) and during vitamin B12 or iron replacement therapy

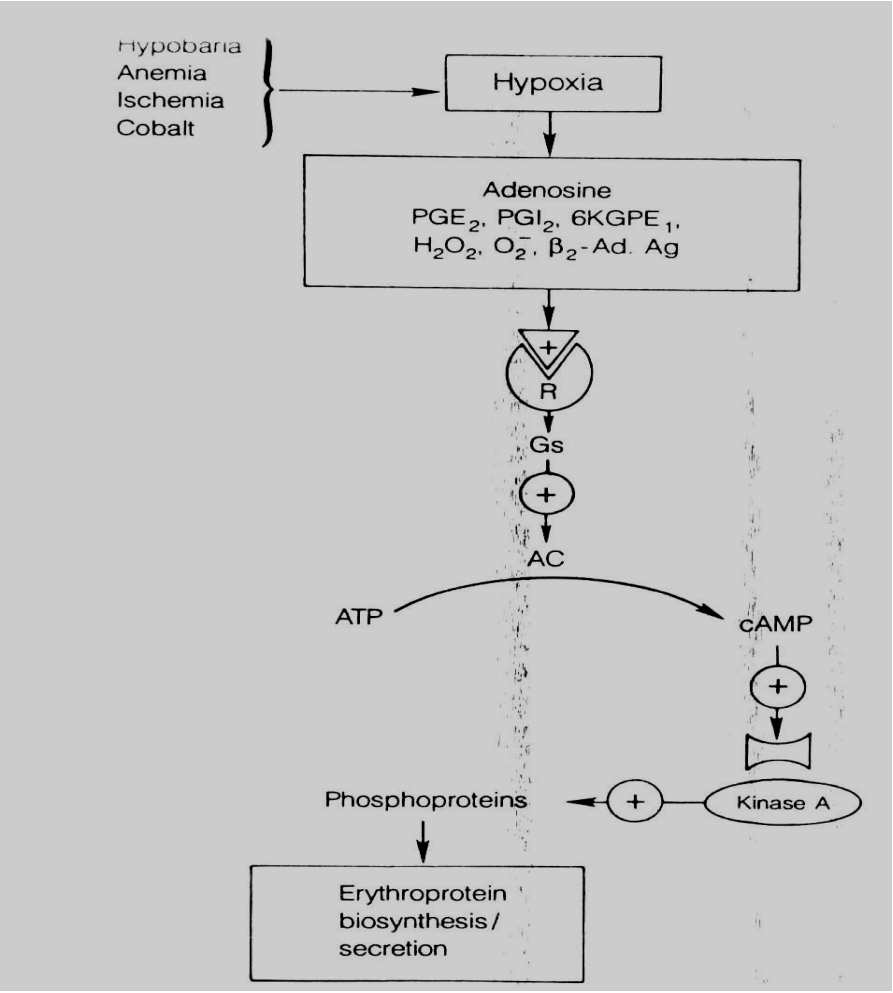

Figure 1.2 Regulation of hypoxia that stimulates EPO secretion. Several chemical agents released and activation of receptor in the cell membrane further initiates transcriptional and/or translational stages of EPO biosynthesis (Adapted from Sytkowski, 2004).

(Kendall *et al.*, 1992; Cazzola *et al.*, 1992). These finding shows that the higher number of RBC precursors the faster the EPO clearance (Cazzola *et al.*, 1996a).

Inflammatory cytokines may interfere with EPO gene expression. Interleukin-1 (IL-1) and transforming growth factor β (TGF β) have found to inhibit hypoxia-induced EPO production *in vitro* (Faquin *et al.*, 1992; Jelkmann *et al.*, 1992). These cytokines also inhibit erythroid progenitor cells proliferation, thus playing a major role in the pathogenesis of the anemia of chronic disease (Means *et al.*, 1992). At variance, IL-6 was shown to mimic hypoxia *in vitro* (Faquin *et al.*, 1992) and elevated levels in human are associated with adequate endogenous EPO production (Cazzola *et al.*, 1996b). Another factor is increase in plasma viscosity that inhibits EPO formation, thus contribute to anemia both inflammation and monoclonal gammophaties (Singh *et al.*, 1993).

1.1.3 EPO receptor

Cytokines activate its biological activity by binding to the specific receptor that found on the surface of the cells known as EPO receptor (EPO-R) (D'Andrea *et al.*, 1990). Unlike human insulin that produced in a specific islet, EPO is produced by numerous cell types and EPO-R is expressed on a wide variety of cells, both hematopoietic and non hematopoietic cells. It is suggesting that EPO expanding its therapeutic function beyond erythropoiesis owing to the wide distribution of EPO receptor.

The EPO-R is a member of the cytokine receptor superfamily (Latini *et al.*, 2008) that forms a dimer on the cell surface and for erythropoiesis, it exists in a preformed dimerized configuration in which a single ligand molecule engages two identical receptor extracellular domains (Frank, 2002). EPO functioned mechanistically as a cross-linker that brings the intracellular portion of its receptor (EPO-R) into close proximity, producing the signal transduction events that eventually allow the cells to mature into red blood cells. However the monomeric EPO-R also appears to be capable of interacting with other membrane receptor proteins forming heteromeric receptor complexes that modulates different signals unrelated to erythropoiesis (Latini *et al.*, 2008).

It is also play a beneficial role in cells protection, activating cytoprotection (e.g., in the brain, heart and kidney), reducing inflammatory responses, preserving vascular integrity, and mobilizing stem cells, including proliferation and

differentiation of endothelial progenitor cells (Latini *et al.,* 2008). Observations of EPO-R expression in cancer cells, coupled with identification of non-haematopoietic functions of EPO, have stimulated much preclinical research into the potential growth-modulating and hypoxia-related effects of EPO on cancer cells. (Ősterborg *et al.,* 2007). Thus, accumulated data and studies of EPO related signaling pathway give a promising treatment for ischemic stroke in brain (Noguchi *et al.,* 2007), heart failure (Latini *et al.,* 2008) and vascular injuries (Fliser *et al.,* 2008).

1.1.3.1 Activation of EPO signaling pathway

In response to the interaction between EPO and EPO-R had allowed the conformational changes of receptor dimers and further initiate tyrosine phosphorylation and activation of several interacellular signaling protein; including JAK-STAT system, phosphatidylinositol-3 kinase and protein kinase B, mitogen-activated protein-kinase kinase (MAP kinase), phospholipase C/ protein kinase C, nuclear factor *k*B and recently endothelial nitric oxide synthase (eNOS) (Latini *et al.,* 2008; Fliser *et al.,* 2008). Activation of EPO and EPO-R that initiates several signal pathways are shown in Figure 1.3.

Activation of janus tyrosine kinase 2 and signal transducer-activator of transcription 5 (JAK2-STAT5) system leads to homodimeric formation following the phosphorylation, translocates to the nucleus and activates target genes related to cell growth, survival and differentiation (Foley, 2008). This process also results in up-regulation of anti-apoptotic proteins such as Bcl-2 that inhibit programmed cell death of erythrocyte precursors (Wojchowski *et al.,* 1999). Therefore it gives benefit for cells protection against apoptosis and prolongs cells survival other than showed beneficial role for proliferation and differentiation red blood cells.

Another important signaling pathway induced by EPO is the phosphatidylinositol (PI)-3 kinase that activates the serine/threonine protein kinase B also known as Akt (Latini *et al.,* 2008). After Akt targeted at PI-rich membrane, and is activated by phosphorylation at Thr^{308} and Ser^{474} from two distinct PI-dependent kinases (Bao *et al.,* 1999) and subsequently, the proapoptic factor Bad phosphorylate, which in turn disassociates from a cell survival factor, Bcl-XL resulting in protection from apoptosis (Datta *et al.,* 1997). Prevention of Akt phosphorylation eliminates the tissues protection effect by EPO (Fliser *et al.,* 2008). As well as PI 3-kinase signal pathway, mitogen-activated protein kinase (MAP kinase) also contribute to glycogen

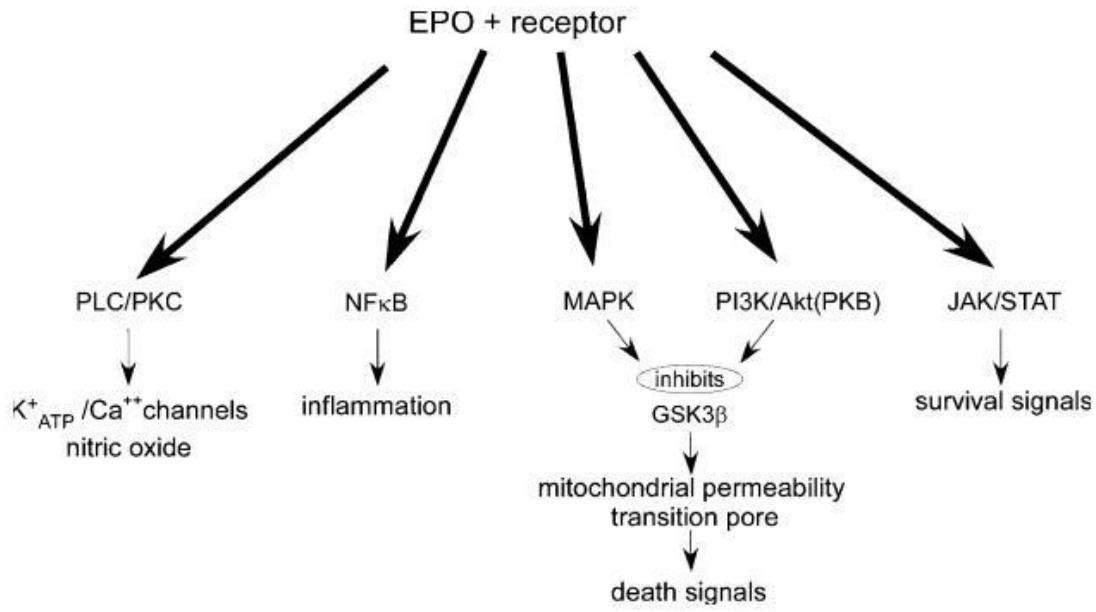

Figure 1.3 Activation of several intercellular signaling pathways by EPO-EPO-R interaction (Adapted from Latini *et al.,* 2008).

synthase kinase 3ß inhibition, a key regulator of the mitochondrial permeability transition pore and therefore reduced in synergistic manner, the probability of mitochondrial injury, leakage of cytochrome c, and subsequent activation of apoptosis (Juhaszova *et al.,* 2004)

Recently, activation of endothelial nitric oxide synthase (eNOS) by EPO has been identified to be of particular importance for its endothelial and vascular effect (Beleslin-Cokic *et al.,* 2004 and d'Uscio *et al.,* 2007). This interesting discovery showed that EPO capable to regulate proliferation and differentiation of endothelial progenitor cells (EPCs) that required for vascular restorative process and promote endothelial regeneration (Asahara *et al.,* 1999; Dzau *et al.,* 2005). Phoshorylation of eNOS in EPCs begin with PI3K/AKT activation then result in nitric oxide (NO) formation which is necessary for endothelial repair. EPO-EPO-R interaction also stimulates eNOS signaling pathway that promotes proliferation and differentiation of neural progenitor cells or neuron that exhibit as neuroprotective in brain (Noguchi *et al.,* 2007).

1.1.4 Expression of rhuEPO in various expression system

Introduction of EPO to recombinant DNA technology and the abilty to produce recombinant protein in a large scale manufacture has profoundly impacted the replacement therapy field as EPO has made important contributions to human health mainly for patients suffer from anemia. Since then, researchers have applied expression of rhuEPO in different protein expression systems derived from mammalian, plant, insect, bacteria and yeast cells.

RhuEPO was initially expressed in mammalian cells-culture system, CHO cells and was extensively used for biomedicine application. Lin *et al.,* (1985) first reported that human EPO has been isolated from genomic phage library using 20-mer and 17-mer oligonucleotide probes and expression of EPO was biologically active *in vitro* and *in vivo* in CHO cells. This pioneering work had contributed to active research works for production of EPO in CHO cells. Inoue *et al.,* (1995) reported that EPO gene amplification can be achieved in CHO cell clones deficient in the dihydrofolate reductase gene, which allows for co-selection in the presence of methotrexate. Thus, chinese hamster ovary (CHO) cell lines are preferably host for the production of therapeutic glycoproteins as it offers several advantages, including an established infrastructure in the biotechnology industry, processs capability, and

the capacity to produce proteins with N-glycans similar to those found on human proteins. However, CHO cell-based expression systems also have several disadvantages such as a relatively high cost of goods, the potential for propagating infectious agents, such as viruses and prions; along development time from gene to production cell line; and the inability to adequately control N-glycosylation (Sethuraman and Stadheim, 2006). The structure of the glycan chains of this rhuEPO slightly differ of those of the urinary human EPO (uhuEPO), considered as the natural EPO molecule. These later differ from those of urinary human EPO by (1) the presence of NeuGc (less than 3% of total sialic acid) which exhibits a strong immunogenicity in adult human, (2) the kind of sialyl linkages (only α2–3 sialyl linkage), and (3) a number of LacNAc repeat slightly increased (up to three per molecule for some glycans) (Didier *et al.,* 1999). In order to produce EPO similar with natural EPO glycan chain structures, EPO gene was expressed in a human lymphoblastoid cell line, named RPMI 1788 (Didier *et al.,* 1999). But the data obtained from this study showed that *in vivo* and *in vitro* biological activities were not impared when compared to EPO-CHO and uhuEPO.

Transgenic plants and animals are alternative expression system for the production of therapeutic glycoprotein. Studies for expression of rhuEPO in transgenic plants and animals were carried out. EPO had been expressed in tobacco and *Arabidopsis*, but encountered male sterility and retarded vegetative growth (Cheon e*t al.,* 2004). It was reported that the male sterility in EPO plants was caused by less pollen and shorter filaments compared to wild type controls, which affected self- pollination (Cheon e*t al.,* 2004). Also, a major limitation of plant-based expression systems is the production of non-human glycan structures that essentially lack galactose and sialic acid, and which contain the potentially immunogenic sugars xylose and α1,3-fucose (Gomord and Faye, 2004). The potential allergic responses to these sugars could limit the development of plants for the production of therapeutic glycoproteins (Sethuraman and Stadheim, 2006). A study was carried out to produce EPO in transgenic animals. Park *et al.,* (2005) have developed a line of transgenic swine harboring rhuEPO using microinjection into fertilized one cell pig zygotes. This study provides evidence that production of purified rhuEPO from transgenic pig milk is a potentially valuable technology, and can be used as a cost effective alternative in clinical applications as well as providing other clinical advantages (Park *et al.,* 2005). However the recombinant protein derive from transgenic animals

are unable to generate proper glycosylation as N-glycans of proteins produced in transgenic animals contain high-mannose and hybrid type glycans with low sialic acid content compared to human proteins (Van Berkel *et al.,* 2002). This difference contributes to reduce the proteins serum half-life.

1.1.5 Clinical application of rhuEPO

RhuEPO has been in widespread clinical use for over 15 years. Initially it used for the treatment of anemia associated with chronic renal failure, it has now been demonstrated to be effective in treating anemia in a variety of other clinical settings including HIV, cancer, surgery, and most recently, critical illness (Henry *et al.,* 2004). The rationale for rhuEPO therapy is that increase of erythropoiesis process will result in higher hemoglobin, a more rapid return to a normal hemoglobin level, and subsequently, a reduced need for red blood cells transfusions.

1.1.5.1 Treatment of Anemia

EPO was first used in end-stage renal disease to lower blood transfusion dependency, a major clinical challenge of this era. Two other unexpected clinical effects also became apparent: improved quality of life and exacerbation (Eschbach, 1987; 1989). Chronic kidney disease is common in patients with cardiac failure thus they share major risk factor and similar treatment. Hence it would benefits both conditions if major clinical trials in chronic kidney disease could inform the design of trials in cardiac failure and vice-versa. Anemia presence in both condition and gene therapy strategy with EPO served a potential benefits in reduction of hypoxic vasodilatation, lower venous return, regression of left ventricular dilation and increased capacity to buffer hypoxic stress conferred by higher hemoglobin concentrations (Foley, 2008). It has been controversial in defining optimum hemoglobin targets even several trials had been conducted. While most guideline suggests that target hemoglobin should be at least 11 g/dl (Locatelli, 2004; National Kidney Foundation, 2006). Evidence from recent studies suggesting hemoglobin target above 13 g/dl may increase the risk of cardiovascular events and death (Foley, 2008).

Impaired renal failure leads to anemia is more frequent among diabetic patient with nephropathy than non-diabetic patients with comparable renal failure caused by other factors (Dikow *et al.,* 2002; Thomas *et al.,* 2005). It is now clear that

advanced renal failure causes anemia because of EPO deficiency, as well as haematologic consequences of uraemia. Angiotensin converting enzyme inhibitor (ACE-I) medication and volume expansion can also exacerbate anemia in patients with renal failure (Bosman *et al.,* 2001; Silverberg *et al.,* 2003; Thomas *et al.,* 2004). EPO administration protects against myocardial cell apoptosis, decrease infarct size and enhances microvascular growth resulting in enhanced cardiac function, improved left ventricular function and improves exercise capacity (Mancini *et al.,* 2003; Meer *et al.,* 2004; Wright *et al.,* 2004; Maiese *et al.,* 2005). Clinical application of 10 days rhuEPO in nine stable subjects with diabetes on maintenance dialysis showed improvement regarding insulin resistance (Spaia *et al.,* 2000). It has also been reported that EPO improves mental abilities, exercise tolerance, well-being and life style of the recipients (Thomas *et al.,* 2004).

Anemia is a frequent complication in patients with cancer. In cancer, anemia may result either from the evolution of the disease itself or from applied treatments and particularly, chemotherapy and/or radiotherapy (Milano and Schneider, 2006). Several studies have shown that survival of anemic patients is reduced and this in various types of cancers (lung, cervix, head and neck, prostate, multiple myelomas, lymphomas); the risk of a fatal outcome increases from 19 to 75 % in anemic patients according to the disease localization (Caro *et al.,* 2001). It has been suggested that normalization of blood hemoglobin may improve quality of life and survival in cancer, particularly in the case of a disseminated disease (Manegold, 1998; Caro *et al.,* 2001). EPO α and EPO β isoforms are currently in clinical use and these EPO isoforms are classically administered by the subcutaneous routine three times per week at doses ranging from 150 to 300 U/kg. However, recent studies suggest that a once-a-week subcutaneous injection of 30,000–40,000 units leads to equivalent results for anemia correction (Gabrilove *et al.,* 2001).

Anemia is found in about two thirds of patients with acquired immunodefiency syndrome (AIDS) and generally worsens during the treatment with zidozudine (AZT) (Cazzola *et al.,* 1997). Patient suffers with AIDS normally has inappropriately low endogenous EPO levels. Treatment with zidozudine helps to increase EPO levels but still remain disproportionately depressed. Hence EPO therapy is a favorable approach instead of blood transfusion. Dosages of rhuEPO of 100 to 200 IU/kg administered IV or SC three times a week induced increased hematocrit (Hct) and reduction of transfusion requirement in patients with baseline

serum EPO levels less than 500 mU/mL (Fischl *et al.,* 1990; Henry *et al.,* 1992) A large study using weekly dosages of 24 000 to 48 000 U/wk in patients with baseline EPO less than 500 mU/mL and Hct less than 30 % showed an increase in Hct of at least 6 Hct points and a decrease in transfusion requirement in 44 % of the patients (Phair *et al.,* 1993).

1.1.5.2 Miscellaneous uses

As an alternative to blood transfusion in critically ill patients, rhuEPO therapy has been demonstrated to increase production of red blood cells that complimentary to other approaches to reduce loss in the Intensive Care Unit (ICU) and decrease the transfusion threshold in the management of the critically ill patients. Critically ill patients receive extraordinarily large number of red blood cells and blood transfusion practice is a favorable approach. Although advances in transfusion medicine have greatly decreased the risk of viral transmission during blood transfusion, other concerns regarding the risk and efficacy of transfusion practice and have led to a reexamination of the approach to blood transfusion (Corwin 2006). In patients with multiple organ failure, rhuEPO therapy (600 U/kg) has been shown to stimulate erythropoiesis (Gabriel *et al.,* 1998). In a small, randomized, placebo controlled trial of 160 patients (Corwin *et al.,* 1999) therapy with rhuEPO resulted in an almost 50 % reduction in the number of RBC units transfused as compared with placebo. In this trial, patients with hematocrit less than 38 % on Intensive Care Unit (ICU) day 3, rhuEPO was given at a dose of 300 U/kg daily for 5 days and then every other day until Intensive Care Unit (ICU) discharge. Despite receiving fewer RBC transfusions, patients in the rhuEPO group had a significantly greater increase in hemoglobin level (Corwin, 2006).

Patients undergoing organ transplantation are treated with cyclosporine A to avoid rejection. This drug reduces EPO production and therefore may be responsible for the development of hyporegenerative anemia. RhuEPO was shown to correct anemia and eliminate transfusion requirement in children undergoing cardiac transplantation (Locatelli, 1994).

1.2 Technique of synthetic gene construction

1.2.1 Background

Genetic constructs for the expression of proteins now frequently use synthetic DNA. This is because sequence information from genome and metagenome sequencing project has increased exponentially over the last decade (Venter *et al.,* 2004) but most of these sequences are not available as physical DNA. Thus, synthetic gene that appears as similar as native gene with respectively mimics the natural gene characteristic used widely to generate a gene encoding virtual proteins. Synthetic gene is composed of genetic material that provides an immediate and easy path from sequence databases and sequence manipulation to physical DNA, enabling research without relying on natural DNA sequences or existing DNA sequences.

Gene synthesis technique becomes a powerful gene tool in generating functional synthetic gene since it serves an easy, rapid and cheaper way without depending on mRNA isolation, genomic library (cDNA) as well as preparation of the natural material. Traditional technique that utilizes cDNA libraries or genomic DNA preps is costly, time-consuming, and error-prone and sometimes is impossible to obtain. It is a convenient method to obtain sequence-verified cloned DNA, especially when the sources of biological materials are rare, lost or dangerous specimen, when codons need to be optimized for expression in particular host or when the desired sequence is chimeric or designed *de novo* (Marsic *et al.,* 2008). Cloning of cDNA of interest from natural mRNA has been reported to have problems due to poor sample availability and the high level of nucleases present in the collected source (Smith *et al.,* 1982).

Expansion of gene synthesis technology coupled with the availability of gene sequence data raise much attention from researcher to fully construct DNA fragment and applied in *de nova* synthesis of novel biopolymers (Van Hest and Tirrell, 2001), codon optimization (Gustafsson *et al.,* 2004), construction of DNA vaccines (Yang *et al.,* 2004) or simple gaining access to known DNA sequences when the original templates are unavailable (Dong *et al.,* 2007). Several strategies of gene synthesis have been described including oligonucleotide ligation (Scarpulla *et al.,* 1982), the *FokI* method (Mandecki and Bolling, 1988), self-primer PCR (Dillon and Rosen, 1990; Chen *et al.,* 1994; Hayashi *et al.,* 1994) and PCR assembly (Stemmer *et al.,* 1995)

1.2.2 Gene assembly technique

The principal method for assembling DNA duplexes from synthetic oligos was the joining of complimentary overlapping complexes with the aid of T4 DNA ligase and first elaborated in the late 1960s by Har Gobind Khorana (Agarwal *et al.,* 1970). In the classic gene assembly method (Khorana, 1979), synthetic DNA oligos are 5' phosphorylated using T4 kinase, annealed to form overlapping duplexes, and then enzymatically joined using T4 DNA ligase. A variation of that method was used to synthesize a 5386 bp ΦX174 RFI DNA molecules in 2 weeks (Smith *et al.,* 2003) where 42- mer oligos were 5' phosphorylated, gel- purified, annealed and ligated using *Taq* ligase at 55^{o}C. This strategy was applied broadly in order to synthesis many genes prior to the expression in *E. coli*, for instance, human immune interferon (Tanaka *et al.,* 1983) and human interferon- gene (Edge *et al.,* 1983).

Although gene assembly technique offers a simple and rapid step, the cost in generating a synthetic gene is high especially when construct a large gene. The assembly efficiency of the host cells is low, thus multiple screening test needed to be carried out for positive results. Narang *et al.,* (1986) had reported that mixture of linearized plasmid containing six synthetic complimentary oligos was directly transformed into competent cells. They found out that 1 out of 100 transformants was positive in colony hybridization using one of the synthetic fragment probes.

1.2.3 PCR mediated gene assembly

The most appealing method had been introduced for gene synthesis strategy is PCR assembly described by Stemmer *et al.,* (1995) due to its inherent simplicity. This method uses oligos of 40 nucleotides long that overlap each other by 20 nucleotides. The oligos are designed to cover the complete sequence of both strands, and the full length gene is generated progressively in a single reaction by overlap extension PCR (OEPCR), followed by amplification in a separate tube by PCR with two outer primers (Dong *et al.,* 2007).

In assembly PCR strategy, short oligos are selected which cover the desired gene duplex, with overlaps between successive oligos on the complementary (so-called sense and anti-sense) strands of the duplex. The oligos are synthesized separately, and are pooled in solution. Assembly of the oligos is achieved via hybridization of overlapping oligos on the sense and anti-sense strands. PCR extension with the presence of DNA polymerase is used to fill in any gaps in the

assembly, and second PCR performed in order to amplify the product (Stemmer *et al.*, 1995). This strategy initially derived by the success of DNA shuffling that was introduced as a method for *in vitro* recombination by combinatorial assembly of genes from random fragments generated by partial DNase I digestion (Stemmer *et al.*, 1994 a, b), or from a mixture of oligos and random fragments (Crameri and Stemmer, 1995).

Emerging technologies aim to significantly improve the scale and reliability of gene synthesis, enabling synthesis of genes of length 10 kb and up to 100 kb, as well as multiplexed synthesis of sets of genes (Tian *et al.*, 2004). These technologies typically avoid the traditional high cost of individual oligo synthesis, by the use of parallel synthesis of primer-tagged oligos on photolithographic microarrays, followed by amplification of oligos and cleavage of primers. Further array-based hybridization techniques are used to identify and remove oligos with in correct base composition due to synthesis errors (Thachuk and Condon, 2007).

1.3 *Pichia pastoris* as an expression system

1.3.1 Background

Yeast and other fungal protein expression host are widely accepted as a cost effective expression systems, easy to manipulate, able to grow rapidly and capability to produce protein using a eukaryotic protein-synthesis pathway. Since the early 1980's, yeast have been used the large-scale production of intracellular and extracellular proteins of human, animal and plant origins (Romanos *et al.*, 1992; Romanos, 1995). One of the most successful and widely used yeast is *Pichia pastoris*, a methylotrophic yeast that able to yield higher recombinant protein. At present, the methylotrophic yeast *P. pastoris* has been used for the production of over 420 heterologous proteins (Macauley-Patrick *et al.*, 2005; Liu and Liu, 2008).

Discovery of *P. pastoris* started about 30 years ago when Koichi Ogata first described a new yeast species that can utilize methanol as its sole source of carbon and energy (Ogata *et al.*, 1969). The methanol utilized-yeast (methylotrophs) attracted immediate attention as potential sources of single-cell protein (SCP) to be marketed primarily as high-protein animal feed. During 1970s, Philip Petroleum Company of Bartiesville, USA, started to cultivate *P. pastoris* in media containing methanol as carbon source (Wegner, 1990). Unfortunately, the oil crisis in 1970s caused a dramatic increased in the cost of methane but the price of soybeans

decreased. As soybean was the major alternative source of animal feed, then the choice of SCP production from methanol become less popular. Then *P. pastoris* was developed as a single-cell heterologous gene expression system and become success in biotechnology area. Researcher at SIBIA (Salk Institute Biotechnology/Industrial associates. Inc) isolated the alcohol oxidase 1 (*AOX1*) gene and developed vectors, strains and methods for molecular genetic manipulation of *P. pastoris* (Cregg *et al.,* 1985; Ellis *et al.,* 1985; Cregg and Madden, 1987; Tschopp *et al.,* 1987b; Cregg *et al.,* 1989; Koutz *et al.,* 1989).

The genetic nomenclature adopted for *P. pastoris* mirrors that used for *Saccharomyces cerevisiae*, one of the most well characterized systems in modern biology (Higgins, 1995). *S. cerevisiae* was the first eukaryotic expression system to be used, and remains the most common due to the vast amount of information available on its genetics and physiology (Zhang *et al.,* 2000). However *P. pastoris* had raised much attention than *S. cerevisiae* due to several factors. It can be grown to higher densities up to 100 g/L [dry weight] which is difficult to reach with *S. cerevisiae* that is not always optimal for large-scale production due to problems such as loss of the plasmid during scale-up, hyperglycosylation, and low protein yield (Romanos *et al.,* 1992). The rapid acceptance of *P. pastoris* as an expression system is due to the strong regulated expression under the control of *AOX1* promoter, along with the strong preference of *P. pastoris* for respiratory growth that greatly facilitates *P. pastoris* culturing at high cell densities (Higgins and Cregg, 1998).

1.3.2 AOX1 promoter and methanol metabolism

Methyltrophic yeast is a yeast species that able to use methanol as their source of energy. There are four different genera that capable metabolizing methanol including *Pichia spp.*, *Candida spp*, *Hansenula spp.* and *Torulopsis spp.* (Higgins and Cregg, 1998). The methanol metabolic pathway appears to be the same in all yeasts and involves a unique set of pathway enzymes (Veenhuis *et al.,* 1983). The first step in the metabolism of methanol is the oxidation of methanol to formaldehyde, generating hydrogen peroxide in the process, by the enzyme alcohol oxidase (*AOX*). This reaction is accomplished by catalase (CAT), which converts the hydrogen peroxide into water and hydrogen (Veenhuis and Harder, 1991). In order to avoid hydrogen peroxide toxicity, this first step in methanol metabolism takes

place within a specialized organelle, called the peroxisome, which sequesters toxic hydrogen peroxide away from the rest of the cell (Higgins and Cregg, 1998).

Alcohol oxidase is a homooctomer with each subunit containing one non-covalently bound FAD (flavin adenine di-nucleotide) co-factor that act as a prosthetic group (Van der Klei *et al.*, 1991). It has a poor affinity for oxygen and methylotrophic yeasts appear to compensate for this deficiency by synthesizing large amounts of the enzyme (Higgins and Cregg, 1998). There are two genes in *P. pastoris* that encode alcohol oxidase, *AOX1* and *AOX2*. When induced with methanol, the AOX1 promoter is responsible for the vast majority of alcohol oxidase activity in the cell (Cregg *et al.*, 1989). Disruption of the *AOX1* gene or its promoter leads to a slow methanol utilization (Mut^s) phenotype. As the cells must rely on the weaker *AOX2* for methanol metabolism, and this gene yields 10-20 times less alcohol oxidase activity than the *AOX1* gene, a slower growing and slower methanol utilizing strain is produced. Because of its slower growth, it is desirable when a gene product is difficult to synthesize, slow to fold, or must undergo other complex posttranslational modifications (Daly and Hearn, 2005).

In methanol grown cultures, alcohol oxidase can constitute up to 30 % of the total cellular protein (Gellissen, 2000). This strong AOX1 promoter has therefore been utilized to drive the expression of recombinant proteins to high levels. In methanol-grown shake-flask cultures, this level is typically about 5 % of total soluble protein but can be ≥ 30 % in cells fed methanol at growth limiting rates in fermentor cultures (Couderc and Baratti, 1980).

1.3.3 Common expression strains

All *P. pastoris* expression strains are derived from NRRL-Y 11430 (Northern Regional Research Laboratories, USA). Some have a mutation in one or more auxotrophic genes to allow for selection of expression vectors containing the complementing biosynthetic gene (e.g. *HIS4*) upon transformation (Cregg *et al.*, 1985). All of these strains grow on complex media but require supplementation with histidine (or other appropriate nutrient) for growth on minimal media (Higgins and Cregg, 1998).

The most commonly used expression host is wild-type or GS115 (*his4*), which are wild type with regard to the *AOX1* and *AOX2* genes and grow on methanol at the wild-type rate (methanol utilization plus or Mut^+ phenotype). Strains with

deleted *AOX* genes sometimes are better producers of a foreign protein than wild-type strains (Tschopp *et al.*, 1987b; Cregg *et al.*, 1987; Chirulova *et al.*, 1997). These strains also require much less methanol to induce expression, which can be useful in large fermenter cultures where large amounts of methanol are sometimes considered a significant fire hazard. KM71 (*his4 arg4 aox1Δ::ARG4*) is a strain in which the chromosomal *AOX1* gene is largely deleted and replaced with the *S. cerevisiae ARG4* gene (Cregg and Madden, 1987) As a result, this strain must rely on the much weaker *AOX2* gene for *AOX* and grows on methanol at a slow rate (methanol utilization slow or Mut^S phenotype). The third host, MC100-3 (*his4 arg4 aox1Δ::SARG4 aox2Δ::Phis4*), is deleted for both *AOX* genes and is totally unable to grow on methanol (methanol utilization minus or Mut^- phenotype) (Cregg *et al.*, 1989; Chirulova *et al.*, 1997).

Some foreign protein secretes into the culture medium and unstable due to rapid degradation by proteases there. This condition can contribute to reduce the total amount of protein production and complicate the recovery process. Major vacuolar proteases appear to be a significant factor in degradation, particularly in fermenter cultures, due to the high cell density environment combination with the lysis of a small percentage of cells. The use of host strains that are defective in these proteases has proven to help reduce degradation in several instances. SMD1163 (*his4 pep4 prb1*), SMD1165 (*his4 prb1*), or SMD1168 (*his4 pep4*) are a series of protease-deficient strains that may provide a more suitable environment for expression of certain heterologous proteins.

The *PEP4* gene encodes proteinase A, a vacuolar aspartyl protease required for the activation of other vacuolar proteases such as carboxy peptidase Y and proteinase B. Proteinase B, prior to processing and activation by proteinase A, has about half the activity of the processed enzyme. The *PRB1* gene codes for proteinase B. Therefore, *pep4* mutants display a substantial decrease or elimination in proteinase A and carboxy peptidase Y activities, and partial reduction in proteinase B activity. In the *prb1* mutant, only proteinase B activity is eliminated, while pep4 prb1 double mutants show a substantial reduction or elimination in all three of these protease activities (Higgins and Cregg, 1998). Unfortunately, these protease deficient cells are not as vigorous as wild type strains due to lower viability, that possesses a slower growth rate and show difficulty to transform (Cereghino and Cregg, 2000).

So, this type of strains are recommended when proteolysis protection in foreign proteins are strictly required.

1.3.4 Expression vectors

Vectors containing the AOX1 promoter have been commercially available and have been used for the production of cytosolic or secreted proteins for approximately two decades (Romanos *et al.,* 1992; Cregg *et al.,* 1993). Even though autonomous replication sequences *PARS1* and *PARS2* have been described for generating stable episomal plasmids (Cregg *et al.,* 1985), the majority of available *P. pastoris* vectors are designed for integration into various yeast chromosomal loci (Romanos *et al.,* 1992; Cregg *et al.,* 1993; Scorer *et al.,* 1994). The most common features for *P. pastoris* expression vector components include a fragment of the promoter, a multiple cloning site (MCS), *AOX1* 5' for insertion of foreign DNA sequences and a 3' fragment of the *AOX1* gene required for transcription termination (Tschopp *et al.,* 1987b; Cregg *et al.,* 1987). Additional features that are present in certain *P. pastoris* expression vectors serve as tools for specialized functions.

Heterologous proteins produce in *P. pastoris* can either be expressed intracellularly or secreted into the medium. As yeast secrete only low levels of native protein, extracellular production of recombinant protein is most desirable as the secreted heterologous protein will constitute the vast majority of the protein in the medium. For secretion of foreign proteins, vectors have been constructed that contain a DNA sequence encoding a secretion signal. Signal sequences derived from the *P. pastoris* acid phosphatase pho1p or the *S. cerevisiae*-mating factors are introduced to generate in-frame gene fusions in the vectors (Higgins and Cregg, 1998). Thus, secretion serves an easy and speed purification process compared to expression protein intracellularly. However, the option of secretion is usually limited to foreign proteins that are normally secreted by their native hosts.

Two groups of marker genes involve are the auxotrophic marker group and the dominant antibiotic resistance group (Ilgen *et al.,* 2005; Cereghino and Cereghino, 2007). Several vectors are available which have auxotrophic markers for elements of the arginine, adenine, histidine, uracil and methionine biosynthetic pathways in *P. pastoris* (Cereghino *et al.,* 2001; Nett and Gerngross, 2003; Nett *et al.,* 2005; Thor *et al.,* 2005). The most common marker genes containing dominant antibiotic resistance markers are the *Sh ble* gene from *Streptoalloteichus hinustanus*

and the *BSD* gene from *Aspergillus terreus*, encoding zeocin and blasticid in resistance, respectively (Miles *et al.*, 1998). Another dominant selectable marker that has been described for use in *P. pastoris* is the *sor*^{R} system, based on the *Sorangium cellulosum* enzyme acetyl-CoA carboxylase, which confers resistance to the macrocyclic polyketide soraphen A (Wan *et al.*, 2004). The *P. pastoris* formaldehyde dehydrogenasegene (FLD1), which has been developed as a marker for transformant selection using *fld1* host cells (Sunga and Cregg, 2004; Resina *et al.*, 2005, 2007). The G-418 sulphate resistance gene, under the control of the bacterial *Tn903* kan^{R} promoter, has previously been utilized for work with *P. pastoris*, but only as a secondary marker after primary selection with an auxotrophic marker such as *HIS4* (Scorer *et al.*, 1994; Clare *et al.*, 1991a). Recently, substitution of the native bacterial promoterin *P. pastoris* with the constitutive GAP promoter resulted in improved expression of a modified *Tn903* kan^{R} gene that conferred resistance to G-418 sulphate and demonstrated transformants containing up to 4 copies of the *E. coli*-lactamase gene, as determined from real-time PCR analysis (Cereghino *et al.*, 2008).

1.3.5 Integration of expression vectors into the *P. pastoris* genome

In order to generate a stable transformants, *P. pastoris* is transformed by integration of the expression cassette into the chromosome at a specific locus (Sreekrishna *et al.*, 1997; Cregg and Higgins, 1995). Chromosomal integration is more desirable than the use of episomal plasmid expression systems as episomal plasmids tend to have low copy number, and this will affect the amount of product expressed (Daly and Hearn, 2005). The size of the plasmid may also affect the stability in the host since large episomal plasmids can be lost during repeated generations as they are mitotically unstable (Thiry and Cingolani, 2002; Romanos *et al.*, 1992). In addition, transformants containing episomal plasmids need to be cultured under continual 'selection-based' media conditions in order to maintain the transformed population of cells (Romanos *et al.*, 1992). This procedure may require the use of additives such as antibiotics, which in turn result in increased production costs (Thiry and Cingolani, 2002). Development of genetically stable expression strains is therefore highly desirable (Thiry and Cingolani, 2002) with a rate of vector loss less than 1 % per generation in the absence of selectable markers usually set as the target (Romanos, 1995). Moreover, such integration vectors usually contain selectable markers that enable detection of the transformants. Some vectors allow for

direct analysis of tandem multiple integration events, but other wise further analysis of the integration number is required (Daly and Hearn, 2005). Integration into the genome can occur via homologous recombination when the vector or expression cassette contains regions that are homologous to the *P. pastoris* genome and hence integration can occur via gene insertion or gene replacement shown in Figure 1.4. Integration by gene insertion can result in tandem multiple integration events due to repeated recombination events at a rate of 1–10 % of transformants (Clare *et al.,* 1991a). Transformations that target gene replacement generally result in single copy transformants; however, gene replacement transformants are usually more genetically stable (Romanos *et al.,* 1992; Clare *et al.,* 1991a). Gene replacement is achieved by digesting the 5'and 3' ends of the vector correspond to the 5' and 3' *AOX1* regions of the *AOX1* chromosomal locus. Transformation, therefore, results in site-specific eviction of the *AOX1* gene (Figure 1.4). This event occurs at a frequency of 5–25 % of the transformants (Sreekrishna *et al.,* 1997; Romanos, 1995). This event occurs at a frequency of 5–25 % of the transformants (Sreekrishna *et al.,* 1997; Romanos, 1995). The other transformants are either *His*$^+$ conversions or of the Mut$^+$ phenotype as a result of gene insertion events at either the *his4* or *AOX1* locus (Figure 1.4). (Romanos, 1995). The site of integration and the type (insertion or replacement) events can be confirmed by southern blot analysis hybridized with a probe generated from the AOX1 promoter region (Paifer *et al.,* 1994).

The introduction of the expression cassette into the yeast chromosome can be achieved in a variety of ways including spheroplast formation, electroporation and lithium chloride treatments. The spheroplast transformation method has been used to generate multi-copy transformants by using vectors such as pPIC9K and pPIC9 (Greenwald *et al.,* 1998; Berger *et al.,* 2002). This method requires several steps with the risk that contamination of the yeast may occur. Also, over digestion with the cell-lysing enzyme, zymolyase can reduce cell viability. Electroporation has become increasingly popular and can be used successfully with zeocin-resistant vectors. This method requires fewer steps and the risk of contamination is reduced. Experience in this laboratory has shown that very efficient expression systems can be constructed through application of this strategy, enabling adiverse range of mature and correctly folded proteins to be prepared and readily purified, particularly when they contain a peptide tag, such as hexahistidine, at the N- or C- terminal positions. For example,