Chapter 8.

CLONING & EXPRESSION OF hGM-CSF IN

PICHIA PASTORIS
The experience with both *E. coli* and *Pichia methanolica* exhibiting higher specific activities with secretory production indicated that rhGM-CSF is poorly expressed in the cytoplasm. Therefore for further studies it was more rational to try host-vector combinations for effective expression of rhGM-CSF as a secretory protein. However *Pichia methanolica* is a new system and not many options exist to this end. Besides, the specific activities we obtained were fairly low. It was therefore decided to shift to a more established system like *Pichia pastoris*. The following section describes the experience in this direction.

**8.1. Construction of the vectors pPicZα-GMCSF and pGAPZα-GMCSF:**

pPicZα A is a 3.6 kb vector used to express and secrete recombinant proteins in *Pichia pastoris*. Recombinant proteins are expressed as fusions to an N-terminal peptide encoding the *Saccharomyces cerevisiae* α-factor secretion signal. The vector allows high-level, methanol inducible expression of the gene of interest in *Pichia*, and is usually used in the strains X-33, and KM71H. The recombinant gene is under the control of the AOX1 promoter (described in chapter 1) for tightly regulated, methanol-induced expression. It has a zeocin resistance gene for selection in both *E. coli* and *Pichia* (Baron et al., 1992; Drocourt et al., 1990).

The pGAPZα A (3.1 kb) vector on the other hand uses the GAP promoter to constitutively express recombinant proteins in *Pichia pastoris*. The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) enzyme is constitutively expressed at high levels in many organisms, including *Pichia pastoris*. The promoter of the gene (GAP) encoding the GAPDH protein has recently been characterized and shown to express recombinant proteins to high levels in *Pichia pastoris*, depending on the carbon source used (Waterham et al., 1997). The level of expression seen with the GAP promoter (PGAP) can be slightly higher.
Fig. 8.1. Schematic representation of the construction of the recombinant plasmid pPICZα-GMCSF
Fig. 8.2. Schematic representation of the construction of the recombinant plasmid pGAPZα-GMCSF
than that obtained with the AOX1 promoter. Selection of this vector too is based on zeocin resistance.

The EcoRI site in the MCS of both these vectors was utilized for cloning the hGM-CSF gene as it made the gene in-frame with the α-factor signal peptide (Fig. 8.1 & 8.2). However for the C-terminal end a new primer was synthesized (having a NotI restriction) site with the sequence

5' - TCTCGAGCGCGCGCTCACTCCTGGACTGG - 3' (Primer number 5)

NotI

![Agarose gel showing restriction digestion of pPICZα-GMCSF and pGAPZα-GMCSF](image)

**Fig. 8.3.** Agarose gel showing restriction digestion of pPICZα-GMCSF and pGAPZα-GMCSF
A PCR reaction with this primer and the one bearing an EcoR1 site (Primer number 3) generated a fragment which was digested with EcoR1 and Not1. The vectors were also digested likewise, all the necessary fragments gel eluted, purified and ligated. The ligation mix was used to transform *E. coli* DH5α cells and spread on LB plates with zeocin (25 μg/ml). The clones were selected by restriction digestion after isolating plasmids from the transformed colonies (Fig. 8.3). The recombinant plasmids were named pPICZα-GMCSF and pGAPzα-GMCSF respectively.

### 8.2. Transformation of *Pichia* GS115 strain:

In pPICZα integration can only occur at the AOX1 locus. The linearized vector will integrate by gene insertion into the host 5’ AOX1 region. Therefore a *Pichia* strain containing a native AOX1 gene (e.g. X-33, GS115, SMD1168H transformed with pPICZα-GMCSF will be mut<sup>+</sup> while a strain containing a deletion in the AOX1 gene (e.g. KM71H) will be Mut<sup>S</sup>. 10 μg of plasmid DNA (pPICZα-GMCSF) was digested with restriction enzyme SacI which cuts once inside the 5’AOX1 site. Linearization was verified by observing on an agarose gel. Freshly prepared competent cells of a GS115 strain was transformed with this digested vector (LiCl<sub>2</sub> method) and the whole transformation mixture spread on an YPD plate with zeocin (100 μg/ml). GS115 host strains contain the his4 allele and are hence auxotrophic for Histidine. However pPICZα-GMCSF does not contain the *HIS4* gene. Therefore, even though complex media contains histidine, it was added extraneously to a final concentration of 0.004% to ensure growth of transformants (as per the recomendations of INVITROGEN).

In pGAPzα integration events occur at the GAP promoter and hence the phenotypic difference (mut<sup>+</sup> and mut<sup>S</sup>) does not arise. About 5 μg of the vector pGAPzα-GMCSF was digested with Avr II, which cuts once in the GAP promoter. After linearization it was cleaned and used for transforming a GS115 strain in the
same way as was done with pPICZα-GMCSF and spread on a YPD plate with zeocin (supplemented with Histidine).

Both plates were incubated at 30°C for 3 days after which colonies appeared. The colonies were further patched on YPD plates and used for expression studies.

8.3. Expression of rhGM-CSF in recombinant GS115 strain transformed with plasmid pPICZα-GMCSF in shake flasks:

Using a single colony, BMGY (supplemented with 0.004% Histidine) medium was inoculated in a 250 ml baffled flask and grown at 28-30°C in a shaking incubator (250-300 rpm) until the culture reached an OD₆₀₀ of 6. The cells were harvested by centrifuging at 3000 x g for 5 minutes at room temperature. The supernatant was decanted and the cell pellet resuspended so as to get an OD₆₀₀ of 1.0 in BMMY medium in order to induce expression. The culture was placed in a 500 ml flask and returned to the incubator to continue growth. 100% methanol was added to a final concentration of 0.5% methanol every 24 hours to maintain

![Image of SDS-PAGE gel showing rhGM-CSF production](image.png)

Fig. 8.4. Silver stained SDS-PAGE showing rhGM-CSF production in recombinant GS115 strain transformed with plasmid pPICZα-GMCSF. Lane M: protein molecular weight marker; lanes 1, 2, 3: 12hr, 24hr, 36hr samples (media supernates) after induction.
inducing conditions. Samples were taken every 12 hours, and the culture supernatant was used for analysis of rhGM-CSF by SDS-PAGE (Fig. 8.4) and ELISA.

From the SDS-PAGE (Fig 8.4) it was seen that rhGM-CSF migrated mainly as a diffuse band spanning a range between 28-35 kD and also as a minor hazy band between 18-22 kD. The amount of protein steadily increased till 36 hours to a value of 1.2 μg/ml (Fig. 8.6) as revealed through ELISA, which remained constant for the next 24 hours and declined beyond that. The O.D₆₀₀ of the culture at that point (36 hours after induction) was about 3.5 and did not rise beyond it. This corresponded to a specific activity of 340ng/ml/OD comparable to the values obtained in P. methanolicus.

8.4. Expression of rhGM-CSF in recombinant GS115 strain transformed with plasmid pGAPZα-GMCSF in shake flasks:

The GAP promoter is a constitutively expressing promoter, hence no induction is necessary for expression of recombinant genes cloned under it. Hence a single colony was inoculated in 50 ml of YPD (supplemented with 0.004% Histidine).

![Silver stained SDS-PAGE showing rhGM-CSF production in recombinant GS115 strain transformed with plasmid pGAPZα-GMCSF. Lanes 1, 2, 3, 4, 5: 12hr, 24hr, 36hr, 48hr, 60hr samples (media supernates) from the time of inoculation.](image-url)
medium and incubated at 30°C with constant shaking at 250 rpm. Samples were drawn every 24 hours and the culture supernatants analyzed on SDS-PAGE (Fig. 8.5). The cell density was also monitored. The maximum O.D_{600} reached by the culture was 17, which was achieved after 32 hours. This remained constant for the subsequent period.

The same hazy band was seen on SDS-PAGE (Western Blot, Fig. 8.7) migrating between 28-35 kD. The maximum product formation was 3.5 μg/ml, higher than that observed with methanol induction in the earlier experiment, though the specific activity was lower (200 ng/ml.O_D_{600} as against 340 ng/ml.O_D_{600} with methanol induction, Fig 8.6).

![Bar Chart](image)

**Fig. 8.6.** Comparison of secretory rhGM-CSF production between constitutive and inducible systems.
8.5. Generating multicopy inserts in vivo:

Before going into bioreactor studies it is important to increase specific activity by looking at different genetic strategies. One way of improving specific production formation rate is to generate clones carrying multiple copies of the gene. During transformation multiple gene insertion events at a single locus in a cell do occur spontaneously with a low frequency. Because of the low frequency of multiple gene insertion events, hundreds to thousands of transformants are usually needed for screening to locate these "jack-pot" clones. It is assumed that multiple integrants would confer greater resistance to zeocin and therefore they can be selected on plates with serially higher amounts of zeocin. However multiple gene
insertion with the vector pGAPZα has not been functionally tested by Invitrogen. Multimeric plasmids can also be generated with pPICZα vectors which contain unique Bgl II and BamHI sites to allow construction of plasmids containing multiple copies the gene of interest. Inspite of the seeming advantages, these are time consuming and cumbersome methods. For generating multiple copies the preferred vector is pPIC9K which contains the bacterial kanamycin gene (kan from Tn903) that confers resistance to geneticin (or G418 sulphate) in Pichia. The level of geneticin resistance roughly depends on the number of kanamycin genes integrated. A single copy of pPIC9K integrated into the Pichia genome confers resistance to Geneticin to a level of ~0.25 mg/ml. Multiple integrated copies of pPIC9K can increase the geneticin resistance level from 0.5 mg/ml (1-2 copies) up to 4 mg/ml (7-12 copies). Because of the genetic linkage between the kanamycin gene and the "expression cassette" (PAOX1 and the gene of interest), one can infer that geneticin resistant clones contain multiple copies of the gene of interest.

8.6. Construction of the plasmid pPIC9K-GMCSF:

The plasmid pPIC9K has the HIS4 gene, which allows for selection in the absence of Histidine. The gene of interest has to be cloned within a multiple cloning site under the methanol inducible AOX1 promoter. The amplified PCR fragment with the hGM-CSF gene digested with EcoRI and NotI (used earlier for cloning in pGAPZα and pPICZα) was used once again to set up a ligation with the vector pPIC9K, digested likewise (Fig. 8.8). The cloning was confirmed by transforming DH5α cells and digesting plasmids isolated from the colonies appearing after transformation (Fig 8.9).
Fig. 8.8. Schematic representation of the construction of the recombinant plasmid pPIC9K-GMCSF
8.7. Transformation of *Pichia* strains GS115 and KM71:

The strain KM71 has a mutated *AOX1* locus, which makes its phenotype mut^S^. Sometimes mut^S^ strains exhibit higher recombinant expression than mut^+^ strains. When the vector pPIC9K is linearized with the restriction enzyme Sac I, during transformation in *Pichia*, insertion occurs at *AOX1* in both KM71 and GS115 strains generating His^+^ Mut^+^ phenotype in GS115 and His^+^ Mut^S^ in KM71. Accordingly 10 μg of plasmid pPIC9K-GMCSF was digested with Sac1, linearization confirmed on an agarose gel and cleaned by phenol extraction. This was used to transform freshly prepared competent cells of both GS115 and KM71 strains and spread on MD plates (without Histidine). Colonies appeared after 3 days and were subsequently patched on fresh YPD plates for further use.

8.8. Selection of multi copy integrants:

The His^+^ colonies obtained after transformation (about 20 mut^+^ and 11 mut^S^) were patched on four YPD plates with increasing geneticin concentrations of 0.25, 0.5, 0.75 and 1 μg/ml respectively. The plates were incubated for 5 days at 30°C. All colonies patched on the plate with 0.25 μg/ml geneticin grew confirming that they all carried at least one copy of the *kan* gene. However only one mut^+^ colony grew on the plate with 0.5 μg/ml geneticin, while none grew on the plates with higher concentrations. It was concluded that only one mut^+^ clone was obtained with 2 copies of the gene while all the mut^S^ clones carried single copies.

8.9. Small-scale expression in shake flasks:

The mut^+^ (2 copy) recombinant strain was inoculated in 25 ml BMGY and grown at 30°C with constant shaking at 250 rpm till the culture reached an O.D_{600} of 6. The cells were harvested and resuspended in BMMY medium for induction so as to get a final O.D_{600} of 1. The culture was returned to the incubator with the same conditions of growth and samples were collected every 12 hours and 100% methanol added every 24 hours to maintain the final concentration of methanol at 199
The culture supernates were analyzed by ELISA. The culture grew continuously to an OD<sub>600</sub> of 8 after 48 hours after which it remained constant. The 48-hour sample showed a rhGM-CSF concentration of 4.3 µg/ml (by ELISA) which implied a specific activity ~ 540 ng/ml (Fig. 8.9) which was higher than that obtained with either of the pGAPZα-GMCSF or pPICZα-GMCSF transformants.

In another study the pPIC9K transformed KM71 strain (mut<sup>S</sup>, single copy) was inoculated in 100 ml BMGY medium and allowed to grow with the same conditions (30°C, 250 rpm) till O.D<sub>600</sub> reached a value of 6. Once again, the cells were harvested and resuspended in 25 ml of BMMY to a final O.D<sub>600</sub> of 24.

![Comparison of rhGM-CSF production between mut<sup>+</sup> (2 copy) and mut<sup>S</sup> strains.](image)

Fig. 8.9 Comparison of rhGM-CSF production between mut<sup>+</sup> (2 copy) and mut<sup>S</sup> strains.
Samples were drawn every 12 hours and the amount of rhGM-CSF formed was estimated by ELISA. There was a very marginal increase of cell density over the entire induction period and the O.D600 of the culture after 48 hours was 26. The maximum rhGM-CSF concentration (after 48 hours of induction) was 11 µg/ml corresponding to a specific activity of 420 ng/ml (Fig. 9.9).

The samples from the above cultures were also analyzed by SDS-PAGE (Fig. 8.10) where the heterogeneous band of rhGM-CSF was seen corresponding to a molecular weight range between 28-35 kD.

![Silver stained SDS-PAGE showing rhGM-CSF production in recombinant GS115 (mut+, 2 copy) and KM71 (mutS,single copy) strains (O.D600 equivalent amounts loaded). Lane M: protein molecular weight marker; Lanes 1,2,3: 12hr, 24hr, 36hr after induction with recombinant GS115 strain; lanes 4,5,6: 12hr,24hr, 36hr after induction in recombinant KM71 strain; Lane C: GS115 sample, 10 fold concentrated by spin filtration with a 3kD membrane.](image)

We thus observe that while the maximum volumetric activity was obtained with the mutS KM71 strain primarily because it was resuspended in a small culture volume. Thus both in terms of growth as well as specific activities the best results were obtained with the two-copy mut+ GS115 strain. It was therefore decided to study this strain in a fermenter where a combination of high specific activity and high cell density would help in maximizing the volumetric activity of the culture.
8.10. High cell density culture of mut+ (2 copy) recombinant GS115 strain:

The fermentation strategy for the production of rhGM-CSF from the recombinant GS115 strain included three distinct stages. First, the cells were grown in a batch mode using excess glycerol as the carbon source. The excess glycerol allowed a rapid accumulation of cell mass, although no rhGM-CSF was produced owing to the repression of the AOX1 promoter by excess glycerol. Following exhaustion of glycerol a limited glycerol feed was initiated in a fed-batch mode, taking care that glycerol did not accumulate. During this second stage, cell mass continued to accumulate rapidly, but since glycerol was limited in the medium the AOX1 promoter became derepressed, causing small amounts of the methanol utilization pathway enzymes to be produced, which prepared the culture for growth on methanol. The third stage of the fermentation was initiated by replacing the glycerol with methanol as the carbon source feed into the fermentor. Cell mass accumulated more slowly in this phase, and rhGM-CSF was synthesized. The principal features of the fermentation was as follows.

Media Composition was the same as described in section 7.13, except in place of 20 gm of Dextrose, 40 ml of Glycerol was used.

Initially, during the batch phase, the OD$_{600}$ reached a maximum of 70 (after 20 hours), which was equivalent to 96 g/l of wet biomass. The glycerol fed batch phase was initiated after this with a feed rate of 18 ml/h at a constant rate. This feed rate was continued for the next 12 hours, when O.D$_{600}$ reached a value of 210 (212 g/l of wet biomass). After this the glycerol feed was stopped for the next 6 hours to starve the cells as this prepares them for efficient utilization of methanol. The starvation can be seen by the decrease in the value of rpm and decrease in OUR (Fig. 8.11). After that the methanol fed-batch phase was initiated to induce the cells as well as to sustain cell growth which helps efficient expression of the recombinant protein. The specific activity (as measured by ELISA of the media supernatant) as well as the volumetric activity increased
Fig. 8.11. Fermentation profiles for secretory production of rhGM-CSF using (two copy) mut+ GS115 strain.
Fig. 8.11. SDS-PAGE showing rhGM-CSF production in the fermentor with a mut+ GS115 strain. Lane M: protein molecular weight marker; Lanes 1, 2, 3, 4, 5, 6, 7, 8: media supernates 6hr, 9hr, 12hr...27 hr after induction. The maximum specific activity reached a value of 0.41 mg/l OD_{600} after 31 hours of induction which was close to the maximum value of 0.540 mg/l obtained in shake flask studies. The cell OD increased to a fairly high value of 320 thus giving a volumetric activity equal to 131 mg/l. At this time point both specific as well as volumetric activity showed signs of reaching a plateau.

Fig. 8.12. Western Blot showing rhGM-CSF production in the fermentor with a mut+ GS115 strain. Lane M: protein molecular weight marker; Lanes 1, 2, 3, 4, 5: 6hr, 9hr, 15hr, 24hr, 33hr after induction (media supernates).
Added to it, the fermentor could not sustain the high oxygen uptake rate (inspite of supplying pure oxygen through the mass flow controller) associated with a high O.D600 of 320 (wet biomass= 275 g/l). Post induction the continuous increase in both volumetric as well as specific activity showed that efficient induction was achieved. Post induction samples were also analyzed by SDS-PAGE (Fig. 8.12) and Western Blot (Fig. 8.13). In both cases a diffuse rhGM-CSF band could be seen migrating in the range of 28-35 kD.

8.11. Purification of rhGMCSF:

One of the advantages of getting the protein in the supernatant is that purification becomes comparatively easy. Simple and easy to scale up techniques can be used to purify the protein which drastically reduces the downstream processing costs. As a demonstration a sample purification was attempted though no attempts were made to optimize the same. The culture media was separated...
from the cells by centrifugation at 10,000 g for 30 minutes. The media was collected and diluted to twice its volume using 50 mM acetate buffer pH 5.0 (to reduce pH and help binding to the ion exchange column. It was filtered through 0.22μ membrane filter and loaded on a cation exchanger. This was followed by washing. The bound protein was eluted using a linear gradient of 0-1M NaCl in 50 mM acetate buffer, as shown in figure 8.13.

A fairly clean elution profile was observed (Fig. 8.13). The eluted protein was further purified by size exclusion chromatography with Sephadex G-25. The purified protein was visualized on a 12.5% SDS-PAGE by silver staining (Fig. 8.14). The final recovery as estimated by ELISA was about 20% (estimated by ELISA).

![Fig. 8.14. Silver stained SDS-PAGE showing successive purification levels of rhGM-CSF after fermentation. Lane M: protein molecular weight marker. Lane 1: crude fermentor media supernate; Lane 2: FPLC purified rhGM-CSF (the lower molecular counterpart is also visible); lane 3: First elution after size exclusion with Sephadex G-25; Lane E: 2nd elution after G-25 exclusion.](image-url)
8.12. Discussions:

The results obtained in this section (Chapter 2) clearly demonstrate the advantages of using the more established *Pichia pastoris*, as the host for expression rather than *P. methanolica*. In the former we were able to obtain a 2 copy integrant of the GMCSF gene which gave us significantly higher specific activities. It is possible that further intensive sequencing could lead to higher multicopy integrants leading to further increases in specific activity. Even though the specific activity in the 2 copy integrant was fairly low compared to *E. coli* the fermentation strategy adopted allowed us to grow the cells to densities which are difficult to obtain in *E. coli*. Thus at an OD of 320 we got reasonable volumetric activities of ~ 140 mg/litre. The fact that fermentation times are much longer in this case compared to *E. coli*, has to be counterpoised with the benefits of getting the protein out into the culture supernatant. We were thus able to employ a simple 2-step purification strategy to get pure protein. This stabilising effects of glycosylation is an added advantage. Moreover it should be remembered that the GMSCF expressed in *Pichia* is not a fusion protein as is the case with *E. coli* and the removal of the fusion partner (either the 6-His tag or MBP) is not a problem. The fact that we could run the fermentor at high cell densities and still get values of specific activities close to that obtained in shake flasks, points to the success of the bioprocess strategy in maintaining the cells in an environment conducive to best levels of expression and also to the comparatively high levels of stability of GM-SCF in the supernatant. If either criteria is not met specific activities decline with increase in biomass and the expected increase in volumetric activities is not obtained. Thus from the shake flask to the fermenter the culture OD$_{600}$ increased more than 40 fold while the specific activity fell only by ~ 20 % leading to more than a 30 fold increase in volumetric activity.