Chapter 4.

CLONING AND EXPRESSION OF hGM-CSF IN *E. COLI*
The cloning of the hGM-CSF gene was done by amplifying it from a human source rather than obtaining it from other research groups. Given its commercial importance it is anyway difficult to obtain this gene. Moreover, if at a later stage it was decided to transfer this technology to industry, having an indigenously cloned gene, has a distinct advantage. It was also important to find out whether hGM-CSF cloned from an indigenous source would show any difference (both at the level of the gene and the protein) to standard hGM-CSF, given the heterogeneity exhibited by many cytokines among different populations.

4.1. Stimulation of PBMCs to produce GM-CSF:

Human peripheral blood mononuclear cells (PBMCs) are known to produce a host of cytokines including GM-CSF under the effect of molecular stimulators and antigens (Bagnara et al., 1993). While theoretically an RT-PCR reaction can be used to amplify the low basal level of GM-CSF mRNA present in PBMCs it is always preferable to work with larger quantities of the template mRNA. Thus, in order to ensure the success of RT-PCR strategy, it was important to induce the cells and determine from this, the time course of induction of GM-CSF. To find out the best time point (which gives maximum GMCSF expression), venous blood was drawn, collected on a heparinized tube (diluted with an equal volume of incomplete RPMI-1640 medium and layered over histopaque). This was then centrifuged for 25-30 minutes at room temperature (RT) at 250 X

![Fig 4.1. Bar diagram showing GM-CSF production by PBMCs on stimulation by PHA.](image-url)
g. The PBMC layer on the interface between histopaque and plasma was harvested, washed with incomplete RPMI, cell number counted and adjusted to $10^6$ /ml and cultured in 10% FCS-RPMI. Cultured cells were stimulated with 10 mg/ml Phytohemagglutinin (PHA). Cell culture supernates were collected every 6 hours for a period of 48 hours for assay.

These samples were used to determine the GM-CSF concentration in the supernatant by sandwich ELISA. A standard curve was plotted using GMCSF standards of known concentrations (in the range 15 – 500 picograms/ml). The color developed was correlated with GMCSF activity using this standard curve. The ELISA data revealed that the GM-CSF concentration increased post induction and reached a maximum of 125 picograms/ml after 48 hours of induction (Fig. 4.1).

4.2. Isolation of total RNA from stimulated PBMCs:

After determining the time point at which GM-CSF concentration reached a peak, the induction experiment was repeated keeping all the conditions exactly the same as before. These PBMCs were then harvested after 48 hours. Cell lysis was accomplished in a special lysis buffer, in which the RNases were deactivated. The nucleic acids were bound consequently to the surface of glass fibers in the presence of a chaotropic salt (guanidine hydrochloride). Contaminating DNA was removed by digesting the sample with DNasel and the bound total RNA was eluted by a simple wash. The amount of the total RNA thus isolated was determined (by measuring optical density at 260 nm) to be 10 μg which was more than sufficient for the RT-PCR reaction.

4.3. Design of primers for PCR:

Primers were designed to amplify the GM-CSF cDNA (from the total RNA isolated above) having Nde I and Bam HI restriction sites so that it could be cloned in a pRSET-A plasmid. This plasmid contains the T7 promoter and is
routinely used for getting high-level expression of recombinant proteins in *E. coli*. Human GM-CSF is synthesized as a 144 amino acid (a. a.) prepeptide which is processed during its release from the host cell. A 17 a.a. signal peptide is cleaved off producing a mature peptide of 127 a.a. We designed the primer so as to amplify the gene corresponding to this 127 a.a. peptide. An AUG codon is needed for translational initiation, which was automatically introduced through the Ndel site. Two primers with the following sequences were synthesized

\[ 5' - CCT\ GCA\ GCC\ ATA\ TGG\ CAC\ CCG\ CCC\ GCT\ CG - 3'\ (\text{primer\ number\ 1}) \]

\[ \text{Nde I} \]

and

\[ 5' - GGG\ GAT\ CCT\ CAC\ TCC\ TGG\ ACT\ GG - 3'\ (\text{primer\ number\ 2}) \]

\[ \text{Bam H I} \]

4.4. cDNA synthesis:

1 µg of total RNA and 200ng each of the two primers were used to set up an RT-PCR under the following conditions which were optimized after a few trials.

Step 1. Reverse Transcription - The sample mix was placed in a thermocycler equilibrated at 60°C and incubated for 30 minutes. This step allowed the cDNA formation from the mRNA by the C. therm polymerase.

Step 2. PCR reaction – Amplification of the cDNA using a combination of C. therm. Polymerase and Taq DNA polymerase.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C</td>
<td>2 minutes</td>
</tr>
<tr>
<td>35</td>
<td>95°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td></td>
<td>40°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>1</td>
<td>72°C</td>
<td>7 minutes</td>
</tr>
</tbody>
</table>
The 406 base pair amplified DNA fragment was visualized on a 2% agarose gel by Ethidium Bromide staining alongside DNA molecular weight marker.

4.5. Construction of the plasmid pRSET-GMCSF:

1 μg of the plasmid pRSET-A (2.86 kb) was double digested with restriction enzymes Ndel and BamHI. The amplified cDNA from the RT-PCR reaction was also doubly digested with the same enzymes under the same conditions. The digested samples were run on a 1.5% agarose gel and the digested bands excised from the gel. The gel fragments were cleaned and DNA eluted by phenol extraction and ethanol precipitation. The eluted fragments were quantified by visual estimation on an ethidium bromide stained 1.5% agarose gel. 5 ng of pRSET-A and 2.5 ng of GM-CSF cDNA (in a vector: insert molar ratio of 1:3.5) were taken in a reaction mix having a total volume of 10 μl and ligated at 22°C for 16 hrs (Fig. 4.2). 5 μl of this ligated sample was then used to transform 200 μl of freshly prepared competent E. coli DH5-α cells and the entire transformation mix was spread on LB agar containing 1X ampicillin (100 μg/ml) and incubated overnight. A number of colonies which appeared on the plate were inoculated on the next day in 5 ml LB ampicillin medium in separate flasks and grown for 18 hrs. Plasmids were isolated from these cultures and resolved on 1.2% agarose gel. All plasmids exhibiting lesser mobility than the control pRSET-A plasmid were selected for further screening. These were then digested by Ndel and BamHI. The digested samples were electrophoresed on a 2% gel and the positive clones showed the presence of a 390 bp insert corresponding to the size of the GM-CSF gene (Fig. 4.3).
Fig. 4.2. Schematic diagram showing the construction of the plasmid pRSET-GMCSF
4.6. Sequencing of the hGM-CSF cDNA:

In order to confirm the cloning of the hGM-CSF gene, DNA sequencing was done. The plasmid pRSET-gmcsf has a T7 promoter, which drives the transcription of the rhGM-CSF gene. Consequently the universal T7 primer was used to sequence the cloned cDNA. The standard Sanger's di-deoxy nucleotide method was used. Radioactive P$^{32}$-γATP was used to label the primer rather than silver staining the sequencing gel. The sequence obtained was submitted to GenBank where it is available with the accession number ‘AF510855’. It has a difference of two nucleotides from the first published cDNA sequence for human GM-CSF (accession number NM_000758, Wong et al., 1985). At positions 239
and 315 there is an A and a C respectively in our sequence compared to Wong et al, where there are Ts on both positions. This implies that at the protein level at position number 80 there is a Lysine and at position number 106 there is a Leucine in our sequence compared to the standard sequence, where there is a Methionine and a Phenylalanine at the two respective positions. This difference raises an interesting question about the heterogeneity of GM-CSF in different human populations. This aspect was not fully explored given that our focus was on high level expression of the gene. However it should be noted that this GM-CSF was biologically active as demonstrated in chapter 9.

4.7. Expression of rhGM-CSF in E. coli BL21-DE3 strain

The plasmid pRSET-GMCSF was used to transform E. coli BL21-DE3 competent cells. A single transformed colony was inoculated in 5 ml LB containing 1X ampicillin and grown overnight at 37°C with continuous shaking at 200 rpm. After 16 hours, 1 ml of the grown culture was used to inoculate 100 ml LB ampicillin in a flask which was incubated under the same conditions. The optical density of the culture was monitored continuously. At an OD_{600} of 0.6, the culture was induced with 0.5 mM IPTG. 1 ml samples were drawn from the induced culture every hour, centrifuged at 10,000 rpm, supernatant discarded and cell pellets stored at -20°C. The frozen cell pellets were resuspended in phosphate buffered saline (PBS), 6X gel loading buffer was added to make the final concentration 1X and boiled for three minutes to lyse the cells. These samples consisting of total cellular protein were then loaded on a 15% denaturing polyacrylamide gel (SDS-PAGE) and analyzed by coomassie staining. No perceptible induction in the form of a protein band was observed at the 14.5 kDa range which is the expected molecular weight of the protein (Fig. 4.4).
Fig 4.4. SDS-PAGE showing total cell lysate of BL21 (DE3) cells carrying pRSET-GMCSF at different time points after induction. Lane M. protein molecular weight markers; Lanes 1 to 8. Uninduced, 1hr, 2hr, 3hr..upto 7 hrs of induction

4.8. Quantifying rhGM-CSF expression by ELISA:

Since the level of expression was too low to be visualized on SDS-PAGE it was decided to do an ELISA since it was a more sensitive assay. An E. coli BL21-DE3 culture carrying the plasmid pRSET-GMCSF was therefore induced as mentioned earlier with 0.5

Fig 4.5. rhGM-CSF production in BL21(DE3) cells after induction with IPTG.
mM IPTG and hourly samples collected. The pellets were resuspended in PBS and incubated at 37°C for 15 minutes in the presence of 200 μg/ml lysozyme. The samples were then lysed by sonication. This was followed by centrifugation at 14000 rpm for 20 minutes, which pelleted the cell debris. The supernatant consisting of the soluble cytoplasmic fraction was collected and analyzed through sandwich ELISA. The rhGM-CSF concentration reached a maximum of 500 ng/ml one hour after induction and progressively decreased with time to reach zero after about 5 hours (Fig. 4.5).

4.9. Study on possible inclusion body formation:

Since this level of rhGM-CSF in the soluble cytoplasmic fraction was very low, a possibility existed that the recombinant protein was aggregating in the cells as insoluble inclusion bodies and getting pelleted along with the cell debris. To find out whether this was happening with rhGM-CSF, cells from 50 ml of culture was

Fig 4.6. SDS-PAGE showing inclusion body formation of rhGM-CSF. Lane M: protein molecular weight marker; Lane 1: Cell debris suspended in 8M urea; Lane 2: Sonicated cell lysate supernatant.
harvested after 3 hours of induction with 0.5 mM IPTG in triplicate. The pelleted cells were suspended in PBS, incubated with 200μg/ml lysozyme and then sonicated. The cell lysates were then centrifuged at 14,000 g for 30 minutes to pellet the cell debris and inclusion bodies if any. The supernatants were decanted and the cell debris were resuspended separately in 1% SDS, 8M Urea and 6M Guanidine Hydrochloride respectively and boiled. These were analysed on SDS PAGE by coomassie staining. We observed a band in the correct position (corresponding to the GM-CSF protein) which is possibly because of inclusion body formation though the amount of protein did not seem to be significantly high (Fig. 4.6).

4.10. Effect of IPTG concentration on production of rhGM-CSF

To find out the effect of varying concentrations of IPTG on the level of product formation, an overnight culture carrying pRSET-GMCSF was used to inoculate three separate 100 ml LB ampicillin in 3 separate shake flasks. Each of them was induced at an OD_{600} of 0.6 but with different IPTG concentrations of 0.1 mM, 0.5 mM and 1 mM, chosen to represent a broad range. Hourly samples were kept at duplicates. One was lysed by boiling in SDS buffer and analyzed on SDS-PAGE by coomassie staining and by Western Blot. The other was used to quantify the production level of production by ELISA. The Coomassie stained gel did not show any induced protein band as before. Western Blot however revealed that rhGM-CSF was indeed being
induced, the maximum being at an IPTG concentration of 1 mM (Fig. 4.8). The cell pellets were also suspended in the same volume of PBS and after sonication the supernatant was serially diluted and quantitated by ELISA. The product accumulation pattern as measured by ELISA was identical in all the three cases, increasing till 1 hour after induction and then gradually tapering off. The maximum rhGM-CSF concentration obtained 1 hour after induction with 1 mM IPTG was 800 ng/ml of culture volume (Fig. 4.7). This represents a fairly low value of 0.5% of total cellular protein.

![Western Blot showing the effect of IPTG on rhGM-CSF production BL21(DE3) cells. Lane M: protein molecular weight marker; Lanes 1,2,3,4: uninduced, 1hr, 2hr, 3hr after induction with 1mM IPTG; Lanes 5,6,7,8: same as in 1,2,3,4, with 0.5mM IPTG; Lanes 9,10,11,12: same, with 0.1mM IPTG](image)

One of the main advantages of *E. coli* as an expression system has been the high level of specific activity obtained with recombinant proteins. Expression levels routinely reach 10-20% of TCP and many research groups have obtained values as high as 40% of TCP for some recombinant proteins. *E. coli* can also avoid the apparent toxicity of the product by packaging it off as inclusion bodies. As a result the cell growth pattern undergoes a minor shift but a high amount of the product is ensured. In the case of rhGM-CSF however we see that the product formation is extremely poor compared to a lot of other proteins. It is almost 100 fold lower than recombinant protein yields obtained in our lab for eg.
Asparaginase (source, *E. coli*), Streptokinase (source, *Bacillus*), Interferon Gamma (source, human) etc. GM-CSF thus seems to be one of those "difficult to express" proteins. As pointed out earlier, other groups working on GM-CSF have also reported fairly low expression levels. It was therefore decided to investigate the causes behind this low expression.

4.11. Growth rate studies:

One of the most important measures of the state of cellular health is the specific growth rate of the organism. Recombinant proteins vary in their levels of cytotoxicity and this is reflected quantitatively by a change in the doubling time of the cell. Wild type *E. coli* has a doubling time of 20 minutes in complex media. However during recombinant protein production we usually work with highly constrained systems with larger doubling times.

A suitable negative control would therefore be a BL21(DE3) strain carrying only the pRSET plasmid (without the GM-CSF gene). We however observed that the specific growth rate of the rhGMCSF recombinant was lower than the one without rhGM-CSF even before the cultures were induced. We concluded that a leaky background expression of rhGM-CSF (which is normal with BL21-DE3 strain) retards the growth of the cells. Therefore a non-induced culture of the same recombinant strain was used as a negative control. The test cultures were all induced at an OD ranging between 0.65 to 0.7. The non-induced culture exhibited a steady rise in cell density and reached an OD of 2.76 in 4 hours whereas all the induced cultures expressing rhGM-CSF barely reached an OD of 1 (Fig 4.9.A, B). It is evident from the growth curve that rhGM-CSF expression severely restricts bacterial growth. This could be one of the major reasons for the low level of specific activity. From the graph, the specific growth rate ($\mu$) was calculated for each curve. The uninduced culture had a $\mu$ of 0.47 h$^{-1}$ corresponding to a doubling time of 88 minutes, while the induced cultures all had a $\mu$ of 0.09 h$^{-1}$ corresponding to a doubling time of 462 minutes.
Fig 4.9.A. Plot of $O_{D_{600}}$ vs Time showing the effect of various IPTG concentrations on the growth curve of BL21(DE3) cells expressing rhGM-CSF.

B. Plot of $\ln O_{D_{600}}$ vs Time showing specific growth rates ($\mu$) at different points of time with different IPTG concentration.
4.12. Effect of reduced growth temperature on induction:

In an earlier observation it was seen that the after induction, the rhGM-CSF producing cells doubled once, stopped growing and eventually lysed (Berges et al., 1996). One way to prevent this from happening is to attenuate the

![Graph A](image)

![Graph B](image)

Fig 4.10.A. OD\textsubscript{600} vs Time plot showing the effect of reduced growth temperature (30°C) on growth curve.

B. lnOD\textsubscript{600} vs time showing comparative μ values at different points.
growth/division rate so as to give more time for product accumulation. A simple way to do that is to grow the cells at a reduced growth temperature. We chose an optimal temperature of 30°C which would lower the specific growth rate without seriously hampering cellular activities. The culture was induced at an OD of 0.65 and the growth curve plotted in comparison to that of an uninduced culture, also grown at 30°C. The growth curve looked exactly like the one measured at 37°C with only a marginally lower specific growth rate. The uninduced culture grew to reach an OD of 2.25 after four hours measured from the time of induction, while the induced culture could only reach an OD of 0.87 in the same time. The product concentration measured by ELISA was 350 ng/ml, lower than that of the culture grown at 37°C (Fig. 10.A, B). From the above observations we can conclude the high metabolic burden imposed on the host cell by GMCSF cannot be ameliorated by reducing the temperature of cultivation.

4.13 Using a heat-shock based induction strategy:

One possible bottleneck for GM-CSF expression could be at the transcriptional level. In that case overexpression of the T7 RNA polymerase would increase transcription rates and thus lead to more product formation. This was indeed observed with increasing IPTG concentrations. Another way to get high levels of T7 RNA polymerase is to have it under a multi copy plasmid and induce it by the stronger λP_L promoter. To this end, a two-plasmid system was used for expression. The plasmid pGP1-2 carries a T7 RNA polymerase gene under a heat-shock λP_L promoter, which is silent below 30°C but starts producing T7 RNA polymerase above that temperature. It also carries a kanamycin resistance gene. Consequently, BL21(DE3) was co-transformed with two plasmids, pRSET-GMCSF and pGP1-2. Cultures were grown in LB medium with both ampicillin and kanamycin at 30°C and the cells were briefly given a heat shock by shifting the culture to 42°C for 2 minutes and then transferring it to a shaker working at 37°C for continuous synthesis of T7 RNA polymerase. The results were analyzed on SDS-PAGE and ELISA. Surprisingly no induction could be observed even
after repeated attempts. This demonstrates that transcriptional controls are not an important factor controlling hGM-CSF expression. Rather the increased metabolic burden associated with producing large amount of T7 RNA polymerase along with GM-CSF adversely effects cellular health and lowers productivity.

4.14. The morphology of rhGM-CSF expressing colonies:

An important indicator of cellular health is the appearance of the colonies expressing the protein. RhGM-CSF expressing bacterial colonies exhibit a marked difference in their morphology compared not only to the parent strain but also to other recombinant strains (Fig.4.11.).

![Photograph showing the morphology of E. coli colonies expressing rhGM-CSF compared to healthy colonies.](image)

A: Plate with healthy recombinant colonies expressing rhIFN-γ
B: Plate with rhGM-CSF expressing colonies.
As shown in the picture the GM-CSF recombinants had a smaller colony size (inspite of a longer incubation time), were extremely thin layered and they appeared almost transparent and irregular in shape compared to the round, firm, white colonies typical of *E. coli*. In addition they appeared to be granular in texture and this granularity increased with increasing time of incubation. Compared to them, parent *E. coli* strain colonies were smooth and maintained their smoothness even as the colony size increased. Another interesting feature was the loss of viability of the rhGM-CSF expressing colonies with increasing shelf time at 4° C. This loss occurred at a much faster rate than with colonies of other recombinant strains. All these observations point out that rhGM-CSF has a highly toxic effect on the host cells.

4.15. RhGM-CSF expression in other *E. coli* strains:

To determine whether the low expression was a problem specific to the BL21 (DE3) strain we transformed two other *E. coli* strains viz JM109 (DE3) (PROMEGA) and ER2566 (NEB) both inducible by IPTG. The same morphological features were observed in both the transformed strains as seen earlier in BL21(DE3) transformed with pRSET-GMCSF. Both were inoculated separately on LB medium with ampicillin and induced with IPTG as before and the results analyzed on SDS-PAGE as well as ELISA. The level of rhGM-CSF production was almost identical to that of BL21 (DE3) cells.

The lacZ promoter makes a small amount of T7 RNA polymerase without induction. That is why all these strains exhibit a leaky low-level accumulation of the recombinant protein. For a toxic protein, this disturbs the cellular health, ultimately affecting the productivity after induction. *E. coli* posseses a sharp stress response system (Goff et al., 1985) which brings about global changes in its protein expression profile after the expression of abnormal or toxic proteins. This response might prevent the high level accumulation of the target protein. To reduce basal (background) expression of a cloned gene, we make use of the fact
that T7 lysozyme is an inhibitor of T7 RNA polymerase. By expressing this gene at low levels, the pre-induction level of T7 RNA polymerase can be further reduced. Post induction the cells express enough T7 RNA polymerase so that the small amount of degradation by T7 lysozyme is not significant, thus allowing the target gene to be expressed. This system permits toxic proteins to be expressed without adversely affecting cellular health by having a much tighter control of expression.

An E. coli BL21 (DE3) pLysE strain was consequently transformed with pRSET-GMCSF and grown on LB media containing 1X ampicillin and chloramphenicol and induced with IPTG as before. The rhGM-CSF concentration, estimated by ELISA, was 500 ng/ml which was identical to that obtained with BL21 (DE3) strain. The pattern of product formation was also same as before, increasing for the first hour and then gradually declining with time.

4.16. Studies on plasmid stability:

Problems of plasmid instability can be exacerbated when a gene whose product is toxic to the host cell is cloned in the plasmid. In such a situation, cells that lack the plasmid can rapidly overgrow the culture whenever the selection pressure is removed. Incidentally β-lactamase is made in substantial amounts and is secreted into the medium by the plasmid containing cells, where it can degrade the ampicillin in the medium. In addition, ampicillin is susceptible to hydrolysis under the acidic media conditions created by bacterial metabolism. This means that a culture in which the cells exhibit a high degree of instability, ampicillin selection will work only until enough β-lactamase has been secreted to destroy the ampicillin in the medium. From that point on, cells that lack plasmid will not be killed and will begin to overgrow the culture.

The relative growth rate advantage of the plasmid free cells is very high when the recombinants are expressing toxic proteins, hence they tend to quickly overtake
the recombinant cell population thereby leading to poor recombinant protein yields.

Since the specific growth rate of the induced cells falls drastically (due to the metabolic burden associated with GMCSF production) it is possible that plasmid free cells may emerge in the population given the tremendous growth rate advantage they have over the GMCSF producing cells. To study this possibility two identical cultures of BL21 (DE3) carrying pRSET-GMCSF were grown and one of then was induced with 1 mM IPTG while the other was allowed to grow without induction. Just before induction a sample was drawn and spread on an LB plate without any antibiotic after appropriate dilution. Likewise after induction samples were drawn from both the cultures, diluted and spread on LB plates after every hour. These plates were incubated overnight at 37°C. On the following day colonies that appeared on the LB plates were patched on LB ampicillin plates to measure the fraction of plasmid containing cells.

The data obtained showed that there was about 5% instability in both the cultures at the time of induction. The non-induced culture did not demonstrate any appreciable degree of plasmid instability and after four hours the instability increased to 8%. The stability pattern for the plasmid carrying was radically different. The first hour sample showed 30% instability which went upto 90% in just 4 hours (Fig 4.12). This high level of instability is difficult to explain unless one postulates an active mechanism of shedding the plasmid by the host cells. We have earlier observed a similar pattern of instability with streptokinase producing recombinants (Yazdani et al., 2002). The most probable explanation is that with the build up of the toxic protein the cells loose their colony forming ability and hence get underrepresented on the plate. Thus the colonies growing on the plate are primarily those which have lost the gene for the toxic protein, even if they actually form a small fraction of the total population in the culture
Fig. 4.12. Graph showing comparison of % plasmid stability between induced and non-induced cells carrying the plasmid pRSET-GMCSF.

4.17. Increasing plasmid stability:

In a strategy similar to that used for streptokinase to increase stability and expression (Yazdani et al., 2002) it was decided to increase the selective pressure. As discussed earlier, β-lactamase is secreted out into the media and ampicillin is degraded by it. One way to prevent all the ampicillin from getting degraded is to add it in excess. Excess ampicillin in the media will not allow any plasmid free cells to grow preventing it from overtaking the recombinant cell population. The fraction of plasmid containing cells will be more, though the growth rate will be affected due to the high antibiotic concentration. Two cultures of BL21 (DE3) with pRSET-GMCSF were grown in LB medium, one containing 1X ampicillin as control and the other containing 5X ampicillin. They were both induced at an OD₆₀₀ of 0.6 with 1 mM IPTG. While the first one was induced directly, the second one was centrifuged, pelleted down, suspended in fresh LB with 5X ampicillin and then induced. Samples were collected after every hour and
analyzed on Western Blot (Fig. 4.13) and ELISA. After 5 hours the cells of the second culture showed visible lysis.

In the control culture the results were the same as observed before, the protein production increasing in the first hour and subsequently declining. The second one exhibited a steady accumulation of rhGM-CSF which increased every hour and after the fourth hour reached 4 μg/ml (ELISA), a value 5 times greater than the maximum reached with 1X ampicillin.

4.18. The stability of the rhGM-CSF mRNA by quantitative RT-PCR:

There are many ways in which the mRNA of the recombinant protein can affect product yield. One of course can be a low transcriptional efficiency. However, the T7 promoter has been observed to be a very strong promoter, producing copious amounts of mRNA upon induction. A second case might be of low stability of the mRNA. The half-life of the mRNA could be extremely short thereby limiting the recombinant protein production. One might expect the instability of target mRNA
to limit expression in cases where recombinant production is low, although in each case that has been examined, substantial amounts of target mRNA seem to accumulate. Sometimes secondary structure in the mRNA transcript can interfere with the AUG translation initiation codon and/or the ribosome-binding site (Tessier et al., 1984; Looman et al., 1986; Lee et al., 1987). The set is never exhaustive, and there could be numerous other reasons besides these for low recombinant protein production. One simple way to find out the stability of the mRNA would be to quantitate the mRNA at different time points and look for a possible reduction of the level. There is one problem with this approach. It has been seen that because of plasmid instability, the total number of copies of plasmid present in a cell population reduces over time. This would also effect the mRNA level. So the reduction in the mRNA level might not really be indicative of the instability of the mRNA. It has also been seen that the rhGM-CSF reaches a maximum one hour after induction. A simple approach would be to make a comparative quantification between the rhGM-CSF mRNA and a second recombinant mRNA where expression levels are very high (that of recombinant human Interferon-gamma gene) one hour after induction. That would indicate whether mRNA level has any role to play in the low production of rhGM-CSF.

Two cultures of BL21 (DE3), one carrying pRSET-GMCSF and the other carrying pRSET-IFN-γ, were grown on LB ampicillin media at 37°C and induced with 1 mM IPTG at an OD_{600} of 0.6. The cells were harvested after 1 hour and a 10 ml pellet (~ 10^{10} cells) was used to prepare total RNA. The RNA was quantitated by UV spectrophotometry by observing the OD at 280 nM. Equal amounts of total RNA (100 ng) was used to set up a reverse transcription reaction. Equal volumes of the RT mixture were used to set up PCR reactions for 5, 10, 15 and 20 cycles with both samples. The results were analyzed on a 2% agarose gel (Fig 4.14).

In all cases it was found that the PCR product for rIFN-γ (432 bp) was marginally more than that of rhGM-CSF (406 bp). This marginal difference is not commensurate with the huge difference in the protein level as observed on SDS-
PAGE. Thus mRNA stability may be affecting rhGM-CSF production marginally but cannot possibly be the major reason for its low expression.

4.19. Studies on possible codon bias:

Excessive rare codon usage in the target gene has also been implicated as a cause for low-level expression (Zhang et al., 1991; Sorensen et al., 1989). The effect seems to be most severe when multiple rare codons occur near the amino terminus (Chen and Inouye, 1990). A number of studies have indicated that high usage of the Arg codons AGA and AGG can have severe effects on protein yield. The impact appears to be highest when these codons are present near the N-
terminus and when they appear consecutively (Brinkmann et al., 1989; Hua et al., 1994; Schenk et al., 1995; Zahn, 1996, Calderone et al., 1996). In the rhGM-CSF gene, since there is only one AGA codon, this does not seem to be the cause of such a drastic reduction in the recombinant protein level. On the other hand there are 6 rare CCC codons (Proline) of which the first 4 are on the N-terminal. This could be a likely reason for low expression. Recombinant proteins from human sources often carry rare codons (with respect to E. coli), which affect the translational efficiency thereby reducing production.

Two strains E. coli BL21-CodonPlus(DE3)-RIL and E. coli BL21-CodonPlus(DE3)-RP (STRATAGENE) have been devised to express genes that are codon compromised. The former carries extra copies of argU, ileY and leuW tRNA genes on a 3.5 kb plasmid with a chloramphenicol marker. These genes encode tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA and the leucine codon CUA, respectively. The latter contain extra copies of of the argU and proL genes on the same plasmid backbone. These genes encode tRNAs that recognize the arginine codons as above and the proline codon CCC respectively. These codons have the rarest usage in E. coli. Incidentally the rhGM-CSF gene has one AGA, one CUA and 6 CCC codons.

Competent cells of both the strains were prepared and transformed with pRSET-GMCSF. Single colonies of transformed cells were inoculated in LB media with 1X ampicillin and 1X chloramphenicol. Both cultures were induced as before and 1 ml aliquots harvested every hour. The results were analysed on SDS-PAGE alongside induced samples from transformed BL21(DE3) cells.

Contrary to expectation, BL21(DE3) cells showed the maximum induction of rhGM-CSF. In the CodonPlus strains too, the pattern of induction was the same as observed in BL21(DE3) cells, increasing for the first hour and then gradually decreasing (Fig. 4.15). This observation is in contrast with Hua et al., (1994) who reported that when supplemented with a minor tRNA(AGA/AGG)Arg encoded by
Fig. 4.15. Western Blot showing comparison of rhGM-CSF production between BL21(DE3), CodonPlus(RIL) and CodonPlus(RP) strains. Lane M: protein molecular weight marker; Lanes 1,2,3: BL21 cells, 1hr, 2hr, 3hr after induction; Lanes 4,5: CodonPlus-(RIL) cells, 1hr and 2hr after induction; Lanes 6,7P: CodonPlus-(RP) cells, 1hr,2hr after induction.

the E. coli argU gene, the expression level of hGM-CSF was raised about 3-4-fold. The above experiment demonstrates that codon bias was not a factor affecting the expression of rhGM-CSF.

4.20. Multiple induction using 5X antibiotic:

One consistent feature of rhGM-CSF production has been the profile of product formation after induction. Instead of accumulation, there seems to be a reduction of the recombinant protein over time. The possibility of the presence of a degradative mechanism which reduces the inducer concentration could be responsible for the observed phenomenon. To study this possibility it was decided to maintain high level of induction by giving multiple IPTG doses.
Three growing cultures of BL21 (DE3) carrying pRSET-gmcsf with 5X ampicillin were given pulse inductions of 1 mM IPTG after every 30, 45 and 60 minutes respectively. The results were analyzed by Western Blot (Fig. 4.16.) as well as by ELISA. The level of rhGM-CSF went up to 5 μg/ml in all the cases and remained consistent at that level throughout the period of induction. Therefore multiple inductions increased the level of rhGM-CSF production by 25%.

4.21. Optimization of growth media:

Minimal media helps in preventing byproduct accumulation and is therefore often used to get high cell density culture, while complex media (the richer the media the better the specific growth rate) often counters the toxic effect of recombinant proteins.

To study the effect of media on rhGM-CSF production, BL21 (DE3) cells with pRSET-GMCSF were grow in 4 different media (all containing ampicillin), viz. M9 (minimal medium), M9CA (M9 media with casamino acids, semi defined), LB
Fig. 4.17. Western Blot showing the effect of media composition of the expression of rhGM-CSF. Lane M: protein molecular weight maker; Lanes 1,2,3,4: uninduced, 1hr, 2hr, 3hr after induction in M9 medium; Lanes 5,6,7,8: same, in M9CA medium; Lanes 9,10,11,12: same, in LB medium; lanes 13,14,15,16: same in TB medium.

(complex medium) and TB (Terrific Broth, highly enriched). All cultures were induced at an OD$_{600}$ of 0.6. The samples were analyzed as before on SDS-PAGE by Western Blot (Fig.4.17.) as well as by coomassie staining (Fig.4.18.).

The cell densities obtained with M9 and M9CA was comparatively low. OD$_{600}$ was about 0.75 after 3 hours. In the case of both LB and TB the OD$_{600}$ reached

Fig. 4.18. SDS-PAGE showing comparative expression of rhGM-CSF in M9 and TB media. Lanes 1,2,3,4: Uninduced, 1hr, 2hr, 3hr after induction in M9 medium; Lane M: protein molecular weight marker; Lanes 5,6,7,8,9: Uninduced, 1hr, 2hr, 3hr, 4hr after induction in TB medium.
1 in the same time. The expression results for LB was as observed earlier and the M9CA medium also gave similar results. Remarkably, in minimal M9 medium the rhGM-CSF production went up to 4% of the total protein and became clearly visible on SDS-PAGE after coomassie staining. The highest levels of expression were observed in TB. There the rhGM-CSF at its maximum was estimated to be around 6-7% of the total protein. This would imply a cytoplasmic production of about 8-10 mg/litre production at the level of shake flask at OD600 of 1. This demonstrates the ability of TB to counter the 'toxic' affects of rhGM-CSF.

4.22. Comparison between Dextrose and Glycerol as carbon source:

The increased expression observed in TB could be because of various reasons. One of the reasons could be the presence of a slowly utilizable carbon source in the form of glycerol. To compare the effectiveness of the two carbon sources in terms of getting better expression, two cultures were grown in TB supplemented with 0.2% Dextrose and 0.4% Glycerol respectively. After induction the samples were analyzed on SDS-PAGE by coomassie staining (Fig.4.19.). Both carbon source exhibited similar rate of rhGM-CSF production.

![Fig. 4.19. Comparison of rhGM-CSF production between Dextrose and Glycerol as carbon sources. Lane M: protein molecular weight marker: Lanes 1,2,3,4: Uninduced, 1hr,2hr,3hr after induction with Dextrose; lanes 5,6,7,8: same (as lanes 1,2,3,4) with Glycerol.](image-url)
4.23. Discussions:

All the observations described so far demonstrate that the specific yield of rhGM-CSF is quite low. There can be numerous reasons for this. The fact that the protein is highly toxic to the host cell is obvious from the growth curve data. However, through a range of observations we were able to rule out transcriptional and translational bottlenecks as reasons for the low expression of rhGM-CSF.

Thus higher concentrations of IPTG increased expression marginally. Multiple inductions similarly sustained the production over time, which tended to decrease otherwise, but even here the increase was not very significant. On the other hand heat induction strategy using a two plasmid system did not work which indicates that T7 RNA polymerase level was not important in this context.

Translation was not a bottleneck as demonstrated by similar levels of specific activity obtained with CodonPlus cells. On similar lines, but contrary to expectations, the studies on BL21(De3)pLys cells did not give better yields than BL21(DE3) cells. This demonstrated that the assumption that an uninduced background expression of a toxic protein like GM-CSF could contribute towards low expression post induction was not correct. The studies on plasmid stability indicated that plasmid stability was a factor, clearly pointing to the toxicity of the protein being responsible for low expression and consequently the expression was enhanced when a higher concentration of antibiotic was used. This clearly demonstrated that the metabolic burden on the rhGM-CSF expressing cells was high, enabling plasmid free cells to emerge and grow at a much faster rate than their plasmid carrying counterparts.

The most successful study so far was the expression of rhGM-CSF using different media. Both minimal media and enriched media gave higher yields than the standard LB with enriched medium showing higher expression than minimal
medium. Between glucose and glycerol as carbon sources, there was no perceptible difference showing carbon source was not a critical factor.

Overall, the toxicity of hGM-CSF seemed to be the major factor behind poor expression. Therefore avenues had to be found to export rhGM-CSF out of the cytoplasm which could help in reducing the toxic effect of GM-CSF towards the host cell. These strategies are discussed in the next chapter.