International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:3, No:3, 2009

A Novel Cytokine Derived Fusion Tag for Over-Expression of Heterologous Proteins in *E. coli*

S. Banerjee, A. Apte Deshpande, N. Mandi and S. Padmanabhan

***Abstract*—**We report a novel fusion tag for expressing recombinant proteins in E. coli. The fusion tag is the C-terminus part of the human GMCSF gene comprising 45 amino acids, which aid in over expression of otherwise non expressible genes. Expression of hIFN α2b with this fusion tag also escapes the requirement of rare codons for expression. This is also a first report of a small fusion tag of human origin having affinity to heparin sepharose column facilitating the purification of fusion protein.

***Keywords*—**fusion tag, bacterial expression, rare codons, human GMCSF

I. Introduction

WITH the advent of recombinant DNA technology, production of heterologous proteins in *E. coli* has become the choice for expression systems to most of the researchers. It has certain advantages over other systems, viz., very well characterized, relatively simple genetics, high growth and production rate, low cost etc. However, the disadvantages are not less, like no post-translational modification and most importantly high level of expression causes the formation of insoluble protein aggregates (known as inclusion bodies) which are mostly inactive [1]. In order to get an active protein, optimization of the expression conditions or the refolding studies are required which could be time consuming and cost-intensive. On the other hand, many mammalian proteins cannot be expressed successfully in *E. coli*, which leaves researchers either to explore expression in a wide range of *E. coli* hosts or at different temperatures and/or fusion tags [2], [3]. One approach to deal with this difficulty has been to express proteins as N- or C- terminus fusions. Literature evidences show that formation of secondary structures in transcribed mRNA reduces expression of heterologous genes. These secondary structures interfere with the binding of ribosome with mRNA thereby prevent efficient translation initiation. These deleterious secondary structures occur due to short-range RNA-RNA interactions. Sequence determinants at both N- and C- termini of proteins can influence their stability towards protease degradation. In some cases, fusion tags might improve the yield of recombinant proteins by rendering them more resistant to intracellular proteases. There are several fusion tags available for the ease of expression and purification of recombinant proteins in bacterial system and the smallest fusion tag available is the his-tag (6-10 aa) [4]. There are potential problems of leakage of Ni^{2+} ions used during purification of His-tag proteins and also the matrix is expensive for large scale manufacturing purposes. The other tags available are thioredoxin (109 aa) [5], Glutathione S-transferase (236 aa) [6], maltose binding protein (363 aa) [7], NusA (435 aa) [8] etc. Most of these tags are affinity tags, large in size and of microbial origin. Mostly these tags facilitate purification of the fused protein and some of them (thioredoxin, NusA etc) are also reported to increase the solubility of the target proteins compared to unfused proteins when over expressed [1]. Here we report a fusion protein system having a very high potential to aid in expression of foreign genes. The tag protein is a 45 amino acid long peptide and is the C-terminus part of human Granulocyte Macrophage Colony Stimulating Factor (hGMCSF) gene product. hGMCSF is a glycoprotein growth factor that induces proliferation of hematopoetic proginator [9]. The processed hGMCSF polypeptide is 127 amino acid long and of molecular mass of 14.36 kDa. This is the first report of smallest fusion tag (GM) of human origin useful for expression of an otherwise non-expressible gene. GM-fusion protein can be purified by affinity chromatography using immobilized matrix and can escape rare codon requirement.

S. Banerjee (e mail: sampalibanerjee@lupinpharma.com), A. Apte Deshpande (e mail: anjalideshpande@lupinpharma.com), N. Mandi (e mail: naganathmandi@lupinpharma.com) and S. Padmanabhan (corresponding author phone: 91-20-66749489; fax: 91-20-66749560; e-mail: srirampadmanabhan@lupinpharma.com) are with Lupin Limited, Biotechnology R & D, 46A/47A, Nande Village, Mulshi Taluka, Pune-411042, India

II. MATERIALS AND METHODS

A. Construction of fusion tag vector (pCGM)

GM tag (the C-terminus domain of hGMCSF) was amplified from a full length human GM-CSF synthetic gene using gene specific primers (forward primer: 5' ccg ccg gaa ttc cat atg cac tac aag cag cac tgc cct cca 3'; reverse primer: 5' ccg ccg gaa ttc ttt atc atc atc atc gga tcc gac tgg ctc cca gca gtc 3'). PCR was performed in a total volume of 250µl containing 100 pg of synthetic gene (Gene bank accession no. BC108724), 3U of Taq DNA polymerase, 200 µM dNTPs (Bangalore Genei Pvt. Ltd. India) and 10 pmoles each of

primers (Sigma). Amplification was done in a two-step manner at 94 °C for 5 min followed by 5 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s; 25 cycles of 94 °C for 30 s, 62 °C for 30 s and 72 °C for 30 s and final primer extension at 72 °C for 5 min. The PCR product was cloned into pET21a vector (Novagen) and the constructed vector was designated as pCGM. The enterokinase (EK) cleavage site was introduced into the vector to obtain target protein with no extra amino acids at the N-terminus.

B. Cloning and protein expression

To construct the GM fusion proteins, human Granulocyte Colony Stimulating Factor (hGCSF), human interferon alpha 2b (hIFNα2b), human interleukine 2 (hIL-2) and human interleukine 11 (hIl-11) (Gene bank accession nos. DQ914891, AY255838, BC066256, NM_000641, respectively) were amplified using specific primers (Table1) and cloned into pCGM at BamHI/ EcoRI site (Bangalore Genei Pvt. Ltd., India).

Escherichia coli hosts {BL21 (DE3), procured from invitrogen, BL21 (DE3) codon plus from Stratagene, BL21 A1 from invitrogen, wherever applicable)}harboring different plasmids were grown and protein expression was induced by adding either 1mM IPTG (for BL21 (DE3), BL21 (DE3) codon plus) or 13mM L(+) Arabinose (for BL21A1). After 4h of induction at 37 °C, cells were harvested and soluble and insoluble fractions were separated. Protein expression and solubility profile were analysed on SDS-PAGE.

TABLE I PRIMER SEQUENCES FOR GENES UNDER STUDY

Gene	Primer sequence
hIFN α 2b	Forward: 5' ccg ccg gga tcc gat gat gat gat aaa atg tgt gac cta cca caa 3' Reverse: 5' ccg ccg gaa ttc aag ctt tta tca ttc ctt act tct taa act ttc 3'
hIL2	Forward: 5'ccg ccg gga tcc gat gat gat gat aaa cct act tca agt tct aca aag 3' Reverse: 5' ccg gaa ttc aag ctt tca agt cag tgt tga gat gat gct 3'
hIL11	Forward: 5'ccg ccg gga tcc gat gat gat gat aaa ggt cca cca cca cct gga cca c3' Reverse: 5'ccg gaa ttc aag ctt tca cag ccg agt ctt cag cag cag 3'
hGCSF	Forward: 5' ccg ccg gga tcc gat gat gat gat aaa acg cca tta ggc ccg gcc 3' Reverse: 5' ccg ccg gaa ttc aag ctt tta cgg ctg cgc taa atg acg 3'

C. Immunoblot

GM tagged fusion protein was analyzed by immunoblot using rabbit anti-hGMCSF and mouse anti-hGCSF antibodies. *E. coli* BL21 A1 cells harboring pET21a-GCSF and pCGM-GCSF were induced for expression with 13 mM L(+) Arabinose and cell lysates were run on SDS-PAGE followed by trasfer to nitrocellulose membrane. The blot was developed

by using goat anti-rabbit (for anti-hGMCSF) and rabbit anti-mouse (for anti-hGCSF) secondary antibodies conjugated with alkaline phosphatase. Specific bands were visualized using BCIP-NBT reagent (Bangalore Genei Pvt. Ltd.).

D. Affinity purification using heparin sepharose

GM-IL-11 inclusion body was resuspended in the denaturation buffer (50 mM CHES buffer, pH 9.0, containing 1M NDSB 256, 1 mM EDTA, 9 mM GSH-1mM GSSG, 240 mM NaCl and 1 mM KCl) and then refolded as per protocol of iFold protein refolding system kit 2 (Novagen, USA). The refolded protein was diluted with 50 mM sodium acetate buffer, pH 5.0 and loaded onto a Heparin sepharose column and the bound protein was eluted using salt gradient in 100 mM Tris, pH 8.8.

III. Results

A. Construction of pCGM vector and cloning genes of interest in pCGM

Several vector systems containing fusion tags have been described for expression, solubility and purification of recombinant proteins. The aim of this study was to construct an expression vector with a novel fusion tag. C-terminus part of human GMCSF was amplified by PCR and cloned into pET21a vector as NdeI/EcoRI fragment. The oligo nucleotides were designed to introduce EK site as well as two restriction sites, BamHI and EcoRI. This unique cloning strategy would generate protein of interest with authentic N-terminus after cleavage of the tag by EK. The constructed vector was designated as pCGM and sequence of the cGM tag is shown in fig. 1.

```
atgcactacaagcagcactgccctccaaccccggaaact
 M  H  Y  K  Q  H  C  P  P  T  P  E  T
tcctgtgcaacccagattatcacctttgaaagtttcaaa
 S  C  A  T  Q  I  I  T  F  E  S  F  K
gagaacctgaaggactttctgcttgtcatgccctttgac
 E  N  L  K  D  F  L  L  V  M  P  F  D
tgctgggagccagtcggatccgatgatgatgataaa
 C  W  E  P  V  G  S  D  D  D  D  K
```

Fig. 1 Nucleotide and corresponding amino acid sequence of GM tag The enterokinase cleavage site is shown in bold

Human GCSF, hIFNα2b, hIL-11 and hIL-2 were PCR amplified with respective primers (Table 1) and cloned into pET21a as NdeI/EcoRI fragments and into pCGM as BamHI/ EcoRI fragments. All the genes were also cloned into pET21a so as to express proteins without any tag for comparative studies.

B. Expression of the constructs

pCGM-hIFNα2b does not require rare codon supplementation

pET21a-hIFNα2b and pCGM-hIFNα2b were transformed into *E. coli* host BL21 (DE3) codon plus and expression was carried out in presence of 1mM IPTG for 4 hours at 37°C. After induction, soluble and insoluble fractions were separated and analysed on SDS PAGE. Fig. 2 shows IFN is completely insoluble when expressed alone (lane 2 and 6), in contrast, it is ~10% soluble when expressed as GM tag fusion (lane: 4). Yield of rhIFN production is ~ 20% higher (as estimated by densitometric scanning, data not shown) when expressed as GM-fusion compared to expression without fusion (lanes 6 and 8, respectively).

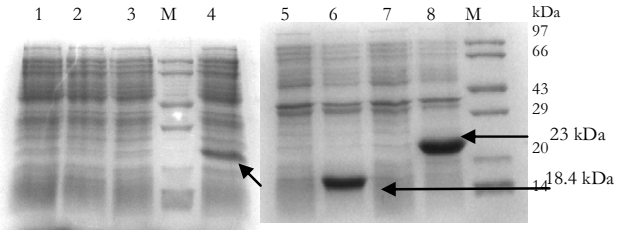

Fig. 2 Expression of hIFNα2b in BL21(DE3) codon plus cells. Lanes 1-4, soluble fraction, lanes 5-8, insoluble fraction. Lanes 1 and 5: pET21a; lanes 2 and 6: pET21a-hIFN; lanes 3 and 7: pCGM; lanes 4 and 8: pCGM-hIFN α 2b. M: medium molecular weight protein marker

Some proteins show expression in *E. coli* only after supply of rare codons in the host and hIFNα2b expression requires specific hosts that can supply rare codons mainly for AGG/AGA (rare arginine codons) [9], [10]. To check whether GM fusion could escape the codon requirement of IFN expression, both the constructs of IFNα2b (pET21a-hIFNα2b and pCGM-hIFNα2b) were transformed into BL21A1, BL21 (DE3) and BL21 (DE3) codon plus cells and expression was induced either with 1mM IPTG or 13mM arabinose. The first two cell lines do not provide tRNAs corresponding to rare codons. Whole cell lysates were analysed on SDS-PAGE and results are shown in fig. 3 and as seen in the figure, hIFNa2b does not express in any cell lines (lanes 1 and 2) except BL21 (DE3) codon plus cells (lane 3). On the other hand, when IFN expression was carried out as GM fusion tag, it was expressed in BL21A1 without supplementation of rare codons (lane 4).

C. GM-tag facilitates expression of difficult genes

Some of the genes especially those of human origin are very difficult to express in *E. coli* and one of the alternatives is to express them with either N- or C-terminus fusion tags. The expression of hIL-2 and hIL-11 was carried out in BL21A1 *E. coli* host from pET21a and pCGM vectors. Fig. 4 shows both hIL- 11 and hIL-2 did not express without any tag (lanes 1 and 4) but expressed in high amount as GM fusion (lanes 2 and 3).

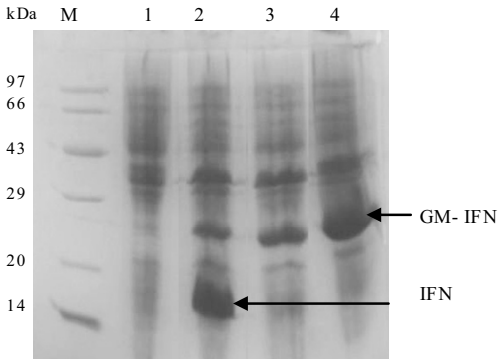

Fig. 3 Expression of hIFNα2b in different bacterial hosts. Lanes 1&2: pET21a-IFN clones and lanes 3&4: pCGM-IFN clones. Lanes 1 and 3: BL21 A1 derived IFN, lanes 2 and 4: BL21 (DE3) codon plus derived IFN. M: medium molecular weight protein marker. Lane 3 shows untagged IFN (~18 kDa) while lanes 3 and 4 indicate tagged IFN protein (~23 kDa).

The GM tag has also been found to have no effect on the activity of the fusion partner, as refolded hIL-11 was found to be biologically active when expressed as a GM fusion.

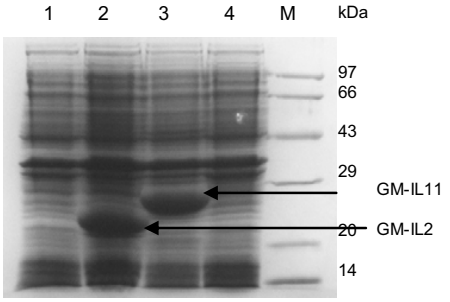

Fig. 4 Expression of hIL-2 and hIL-11 in BL21A1 cells. Lanes 1 and 2: hIL-2 expression without and with tag respectively and lanes 3 and 4: hIL-11 expression without and with tag respectively. M: medium molecular weight protein marker.

D. Immunoblot with anti-hGMCSF antibody

GM fusion proteins could be detected and quantified by immunoblot or ELISA with commercially available anti-hGMCSF antibody. Human GCSF was cloned in pCGM and pET21a vectors and expressed in BL21A1 E. coli host (fig. 5A). Immunoblot analyses of the same samples were carried out with both mouse anti-hGCSF and rabbit anti-hGMCSF antibodies. As shown in fig. 5B, GM-GCSF fusion protein is detected by GCSF (lane 1) as well as GMCSF antibodies (lane 3). As expected, untagged GCSF is detected only by GCSF antibody (lane2) and not by GMCSF antibody (lane 4).

E. Affinity purification using Heparin Sepharose 6 fast flow

The proteins eluted from the affinity column matrix was analysed on SDS-PAGE by silver staining and the protein was

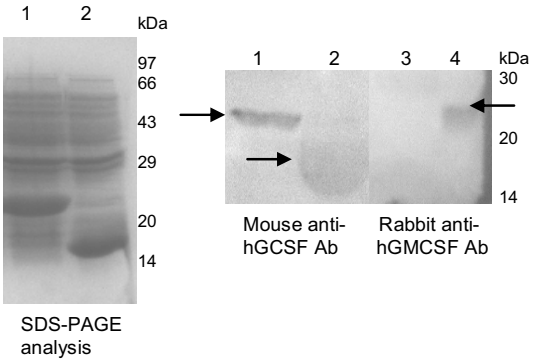

Fig. 5 SDS-PAGE and Immunoblot analyses of expressed GM-GCSF gene. (A) GCSF was expressed in BL21A1 with (lane 1) and without (lane 2) GM tag. (B) Immunoblot analyses of the same constructs analyzed with mouse anti-hGCSF and rabbit anti-hGMCSF antibodies. Lanes 1 and 4: GM-GCSF expressed in BL21A1 and lanes 2 and 3 show untagged GCSF expression. Arrows indicate respective expressed proteins visualized after NBT-BCIP treatment.

greater than 90% pure (fig. 6) and pure GM-hIL11 was found to be biologically active as tested by *in-vitro* activity assay.

IV. Discussion

Human granulocyte macrophage colony stimulating factor (hGM-CSF) is a glycoprotein growth factor responsible for proliferation of hematopoietic progenitor cells and functionally activates mature granulocytes and monocytes. Recently, it has been shown that hGMCSF has putative heparin binding site located in helix C (C-terminus helix) of hGMCSF [10]. Therefore, a fusion tag comprising the C-terminus part of hGMCSF could facilitate purification of

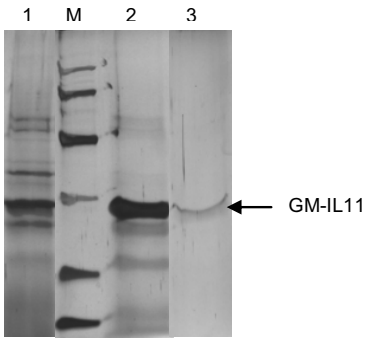

Fig. 6 Purification of GM-hIL11 using affinity chromatography. Lane1: Crude inclusion body preparation of GM-IL11, lane 2: Refolded protein, lane 3: Purified GM-IL11 (at 0.8M NaCl in 100 mMTris, pH 8.8). M: Protein molecular weight marker. The arrow indicates the purified protein of interest.

recombinant fusion protein through immobilized heparin affinity chromatography.

In the present study, we have constructed a prokaryotic inducible expression vector containing a novel fusion tag comprising C-terminus part of human GMCSF. The designated GM fusion tag is 45 amino-acid long peptide and has affinity towards heparin. This is the first report of a fusion tag of human origin useful for expression of an otherwise non-expressible gene (Fig. 4). When mature hIL-2 and hIL-11 genes were subjected to expression studies in *E. coli*, both the proteins did not express in any of the bacterial cell lines tested.

Lee et.al. reported that there is no correlation between codon bias and high level expression in case hIL-2 expression [11]. In our study also, we have found no improvement in the expression of hIL-2 and hIL-11 genes in BL21 (DE3) codon plus cells (data not shown). GM- fusion also has been found to increase protein yield of an expressing gene by ~ 20% (Fig. 2, as estimated by densitometry scanning) and in certain cases increases the solubility. It has been reported that fusion tags increase the expression of heterologous proteins either by reducing the probability of formation of secondary structures near the ribosome binding site, thereby helping in efficient translation initiation or by increasing mRNA or protein stability [12]. The GM tag of present study probably increases the expression of heterologous proteins by providing favourable context for efficient translation or by increasing mRNA or protein stability.

As the described fusion tag is small in size (45 amino acids), the molar ratio of the fusion tag versus fusion partner is always low which in turn results in higher yield of the proteins of interest. In addition to over expression, GM tag also escapes the requirement of rare codons supplement from host (Fig. 3). hIFNα2b gene cannot be expressed in regular *E. coli* host without supply of rare codons which are a prerequisite for interferon α2b expression in *E. coli* [13], [14]. Genes containing rare codons (with respect to *E. coli* codon preference) near the initiation codon, inhibit cell growth and protein synthesis due to ribosome stalling and premature release of specific peptidyl-tRNAs from the ribosome at the rare codons [13]. Also, presence of rare codon specific tRNAs in abundance is not always sufficient for expression of a gene with rare codons as RNA stability and secondary structure have immense effect on the expression of these genes [15].

Though GM-hIFNα2b expresses in ordinary *E. coli* host, at this point, it is not clear how this tag bypasses rare codon requirement. Most probably it alters the RNA secondary structure and hence increases RNA stability. As the fusion tag contains an enterokinase cleavage site, this construct allows the fusion partner to be obtained with no extra amino acid at N-terminus after enterokinase cleavage.

Interestingly, our preliminary studies also indicate that the GM fusion does not alter the biological activity of the fusion protein partner since refolded and purified GM-IL11 was biologically active using TF1 cells as described by Hateno et al. [16]. Moreover, we also found that this fusion protein was

equally active when compared to the commercially available rhIL-11 (data not shown).

V. CONCLUSION

In summary, we report a novel and a small fusion tag of human origin that aids in expression of "difficult to express" genes, escapes the use of special rare codons supplying *E. coli* hosts and has a high and selective affinity towards heparin that would ease the purification of recombinant fusion partner. In addition, all types of GM-fusion proteins could be easily detected by immuno-blot and quantified using commercially available anti-hGMCSF antibody. Efforts are also on to develop strategies to modify the GM-tag to achieve higher solubility of fusion partners. It is also important to mention that CGM does not have GMCSF activity.

ACKNOWLEDGMENT

The authors wish to thank Praveen Muneshwar for the refolding studies of GM-IL-11 and Dr. Veena P. Raiker for the bioactivity studies of GM-IL11 protein. Thanks are also due to Prakash Kotwal and Dr. Shardul Salunkhe for some protein expression studies.

REFERENCES

[1] R G Harrison, "Expression of soluble heterologous proteins via fusion with NusA protein", Innovations, vol 11, pp 4-7, 2000.
[2] D M Valeria, S Gunter, B Stephanie, D M Ario, "The solubility and stability of recombinant proteins are increased by their fusion to NusA.", Biochem. Biophys. Res. Commun., vol 322, pp 766-771, 2004.
[3] L D Cabrita, W Dai, S P Bottomley, "A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production", BMC biotechnology, vol 6, pp 12-19, 2006.
[4] MC Smith, TC Furman, TD Ingolia, and C Pidgeon , "Chelatingpeptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins", J. Biol. Chem., vol 263, pp 7211-7215, 1988.
[5] E R LaVallie, E A DiBlasio, S Kovacic, K L Grant, P F Schendel and J M McCoy, "A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the *E. coli* Cytoplasm", Biotechnol., vol 11, pp 187-193, 1993.
[6] D B Smith, K S Johnson, "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase", Gene, vol 67, pp 31-40, 1988.
[7] D C Guan, P Li, P D Riggs, H Inouye, "Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein", Gene, vol 67, pp 21-30, 1988.
[8] G D Davis, C Elisee, D M Newham and R G Harrison, "New fusion protein systems designed to give soluble expression in *Escherichia coli*", Biotechnol. Bioeng. Vol 65, pp 382-388, 1999.
[9] Y Kanakura, S A Cannistra, C B Brown, M Nakamura, G F Seelig, W W Prosise et. al., "Identification of functionally distinct domains of human granulocyte-macrophage colony-stimulating factor using monoclonal antibodies", Blood, vol 77, pp 1033-1043, 1991.
[10] A Sebollela, T C Cagliaris, G S C S Limaverde, A chapeaurouge, M H F Sorgine, T C Sampaio et. al., " Heparin binding sites in granulocyte macrophage colony stimulating factor", J. Biol. Chem., vol 280, pp 31049-31956, 2005.
[11] H W Lee, J H Joo, S Kang, I S Song, J B Kwon, M H Han and D S Na, "Expression of human interleukin-2 from native and synthetic genes in *E. coli*: No correlation between major codon bias and high level expression", Biotechnol. Lett., vol 14, pp 653-658, 1992.
[12] D S Waugh, "Making the most of affinity tags", Trends in biotech, vol 23, pp 316-320, 2005.
[13] J J Olivares-Trejo, J G Bueno-Martínez, G Guarneros,and J Hernández-Sánchez, "The pair of arginine codons AGA AGG close to the initiation codon of the lambda int gene inhibits cell growth and protein synthesis by accumulating peptidyl-tRNAArg4", Molecular Micribiology, vol 49, pp 1043-1049, 2003.
[14] O L Garcia, B Gonzalez, A Menendez, A E Sosa, J R Fernandez, H Santana and N Meneses, "The argU gene product enhances expression of the recombinant a2- Interferon in *Escherichia coli*", Annals New York Academy of Sciences, vol 782, pp 79-86, 1996.
[15] X Wu, H Jornvall, K D Berndt and U Oppermann, "Codon optimization reveals critical factors for high level expression of two rare codon genes in Escherichia coli: RNA stability and secondary structure", Biochem. Biophys. Res. Commun., vol 313, pp 89-96, 2004.
[16] S Hatano, H Asano, H Nagai, T Murate, N Mori, K Kawashima et. Al., " In vitro effect of recombinant human IL-11 in human malignant lymphoma cells", J. Clin. Exp. Hematopathol., vol 42, pp 55-60, 20.