African Journal of Biotechnology Vol. 11(50), pp. 11104-11109, 21 June, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.133
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Purification of human recombinant granulocyte colony stimulating factor from *Escherichia coli*

M. Ramya[1] and E. Selvarajan[2]*

[1]School of Chemical and Biotechnology, SASTRA University, Thanjavur - 613402, Tamilnadu, India.
[2]School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore 632 014, Tamil-Nadu, India.

Accepted 21 May, 2012

In *Escherichia coli*, recombinant proteins were produced either as three dimensionally folded forms or as unfolded forms, inclusion body (IB). The formation of IB was a frequent consequence of high-level protein production and inadequacy of folding agents namely chaperones in the cytoplasm. The structure of the protein in inclusion bodies was poorly understood but it has been hypothesized that the protein may form misfolded β sheet aggregates. In order to procure the active protein, IBs must then be solubilised, refolded and purified. A study was done to determine the purity of recombinant protein which was produced as IBs in *E. coli*. The IBs were recovered by centrifugation and washed with salts and detergents to remove the host cell proteins and then solubilised with mild buffer at alkaline pH. The solubilised IBs were refolded under redox conditions, purified using cation exchange resins. It was experimentally verified that the final recovery of protein was 30% with 99% purity.

Key words: Granulocyte colony-stimulating factor (GCSF), inclusion body (IB), refolding, solubilisation, *Escherichia coli*.

INTRODUCTION

Recombinant technology has made possible the production of many therapeutically vital protein products and large number of other commercially significant proteins. Recombinant technology begins with the isolation of a gene of interest. The gene of interest [foreign deoxyribonucleic acid (DNA)] is integrated into the plasmid or phage, and this is referred to as recombinant DNA. When a recombinant deoxyribonucleic acid (rDNA) is inserted to an organism, it will produce proteins based on this recombinant DNA. This protein is called recombinant proteins and it code is carried by rDNA. Plasmids that have been constructed to obtain overproduction of individual target gene products in strain BL21 (DE3) (by addition of isopropyl-beta-D-thiogalactopyranoside as inducer) can directly be transformed into GJ1158. The NaCl induction regimen is also shown to be associated with a decreased propensity for sequestration of over expressed target protein within insoluble inclusion bodies (Bhandari and Gowrishankar, 1997).

Human granulocyte colony stimulating factor (HG-CSF or HGCSF), a single chain polypeptide containing 174 amino acid residues (MW = 18,800, pI = 6.1), is one of the hemopoietic growth factors which plays an important role in stimulating proliferation, differentiation, and functional activation of blood cells. It contains a free cysteine at position 17 and two intramolecular disulfide bonds (Wang et al., 2005; Abolghasemi et al., 2010). Some of the factors that have been associated with inclusion body formation in *Escherichia coli* are: high local concentration of the over expressed protein, reducing environment in the cytoplasm due to high levels of glutathione, preventing disulphide bond formation, lack of posttranslational modifications such as glycosylation, improper interactions with chaperones and other proteins participating in folding *in vivo*, intermolecular cross-linking via disulphides; however, proteins without cysteine residues also form aggregates and kinetics of protein translation in the context of rare codons (Susan and Petety, 2007).

*Corresponding author. E-mail: selrajan@gmail.com.

Proteins expressed as inclusion bodies are currently solubilized by the use of high concentration of chaotropic agents such as urea, guanidine hydrochloride, and thiocyanate salts, detergents such as sodium dodecyl sulfate (SDS), N-cetyl trimethyl ammonium chloride, and sarkosyl (sodium N-lauroyl sarcosine). The soluble denatured proteins are then refolded to their native state by dialyzing with buffer containing reducing and oxidizing agents. Renaturation of recombinant proteins from inclusion bodies into bioactive form is cumbersome, results in low recovery of the final product and also accounts for the major cost in overall production of recombinant proteins. The solubilization of inclusion body (IB) proteins by chaotropic agents in the presence of reducing agents such as dithiothreitol or β-mercaptoethanol allows the disruption of non-native disulphide bonds (Ahmed et al., 1975). Following solubilization, naturally disulphide bonded proteins have to be refolded under conditions, which permit the formation of their native disulphide bonds. Free cysteine residues can be oxidised by molecular oxygen, a redox reaction catalyzed by Cu^{2+} ions. 2-D chemical environment controls the refolding of the denatured green fluorescent proteins tethered to solid surfaces (Bryan et al., 2007). Refolding occurs readily on the repulsive poly (ethylene) glycol (PEG) functionalized surface but is inhibited on the attractive-NH_2 functionalized surface.

Each protein has a distinct and characteristic solubility in a defined environment and any changes to those conditions (buffer or solvent type, pH, ionic strength, temperature, etc.) can cause proteins to lose the property of solubility and precipitate out of solution. The environment can be manipulated to bring about a separation of proteins for example, the ionic strength of the solution can be increased or decreased, which will change the solubility of some proteins. A histogram of the solubilities, based on data from 3,173 translated proteins, revealed a clear bimodal distribution, indicating that the aggregation propensities are not evenly distributed across a continuum (Bei-Wen et al., 2006).

The purification method for growth factor was developed through intestinal epithelial cell (IEC) (Chrunyk et al., 1993). Protein refolding is a key step for the production of recombinant proteins, especially at large scales, and usually their yields are very low. In this work, recombinant human granulocyte colony-stimulating factor expressed in *E. coli* was renatured with simultaneous purification by IEC with a Q Sepharose fast flow (FF) column. Several chromatographic parameters affecting the refolding yield of the denatured/reduced r-growth factor, was investigated in detail and indicated that the urea concentration and the pH value were of great importance. At the optimal conditions, the renatured and purified r-growth factor was found to have a specific bioactivity of 3.0×10^8 IU/mg, a purity of 96%, and a mass recovery of 49%. Compared with the usual dilution method, the IEC method developed here is more effective for purification (Chaozhan et al., 2007).

MATERIALS AND METHODS

Expression system and batch cultivation

A pET vector was a bacterial plasmid designed to enable the quick production of a large quantity of any desired protein when activated. The *E. coli* strain BL21 (DE3) was transformed into pET23 vector. The gene was inserted into the target site. The transformants obtained by calcium chloride transformation was selected. Luria-broth medium was supplemented with 0.2% glucose, 50 mg/ml ampicillin and 34 mg/ml chloramphenicol. Batch cultivation was started with overnight seed culture grown in 30°C into the bioreactor with isopropyl-. -D-thiogalactopyranoside (IPTG) induction.

Isolation of inclusion bodies

The fermented broth was centrifuged. 35 g of *E. coli* pellet was washed with lysis buffer and sonication done in the time interval of 12 min 85 amplitude for 0.5 cycles. The samples were diluted to 50 to 100 folds and the lysis was analyzed by taking OD at 600 nm.

Recovery of inclusion bodies

The obtained IBs were washed with ionic and non- ionic detergents and sonication was done for two cycles with time interval of 10 min followed by centrifugation at 9000 rpm for 15 min. The first wash with 1% TritonX-100, 50 mM Tris-Hcl, 5 mM ethylenediaminetetraacetic acid (EDTA) at pH 8. The second wash with 50 mM Tris-HCl, 1 mM EDTA and pH 8.0 and third wash with 1% deoxycholate (DOC). The fourth wash with 0.5% of Triton X100, 0.5% of DOC, 50 mM Tris-HCl, 1 mM EDTA, and pH 8.0. The final wash was done with 2% TritonX-100 and 50 mM Tris-HCl. The actual IB weight was 5.1 g. The samples collected were stored at -20°C. Then the samples were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis by loading the samples in 15% gel and the run was done at 20 mA for 2 h. After the completion of run, the gel was stained in coomassie brilliant blue (CBB) staining for 1 h and kept for destaining.

Solubilisation and refolding

To that pellet solubilised buffer of 8.4 M urea was added and pH was adjusted to 12. The sample was kept for 45 min incubation in magnetic stirrer for complete solubilisation and then added to refolding buffer of 25 mM sodium acetate, 0.3% Tween 20, 5 µM copper sulphate. The sample was kept for 20 h incubation. It was stirred for 4 h and the pH was adjusted to 5.5. Then it was filtered with 3 µm filter and kept for incubation at 4°C. After 20 h incubation the sample was filtered with 0.45 µm filter paper.

Purification of r-protein by SP sepharose column

The refolded sample obtained in the previous step was loaded into the SP sepharose column. The flow rate was set as 10 ml/min and the pressure was 0.5 MPa, 500 ml of 25 mM sodium acetate, pH 5.5 was passed and the volume of wash column collected, then eluted with 25 mM sodium acetate, pH unadjusted and washed with 1 M NaCl.

Determination of purity by high performance liquid chromatography (HPLC)

The purity of protein was determined by high-performance liquid

Table 1. Ratio of pellet to wash buffers.

Sample	Pellet weight (g)	Volume of buffer (ml)
Triton wash	16	400
Tris wash	14	275
DOC wash	9	200
Triton+ DOC wash	8.3	200
Inclusion bodies	8.0	300
Triton wash	5.5	250
Tris wash	5.1	250

The table represented the volume of wash buffer added to the pellet. In each wash, there was a loss in the pellet and reduction from 16 to 5.1 g.

chromatography (HPLC) by using C18 column and equilibrated with buffers. The flow rate was adjusted for 1 ml/min and the run time was set for 45 min. After the sample was kept, the injection volume was adjusted to 40 µl and the run time was adjusted to 45 min and the column temperature was set as 40°C and the different wavelength was detected in UV.

Identification of protein by isoelectric focusing

The monomer solution (24.25% (w/v) acryl amide, 0.75% (w/v) bis (N,N′-methylene bisacrylamide) and catalyst solutions (0.1% riboflavin -5′-phosphate (FMN), 10% (w/v) ammonium persulfate (APS) and 25% glycerol (w/v) were prepared and the solutions was pipetted into the tray where the gel has been kept. 2 µl of sample was applied over the template and allowed to diffuse into the gel for 5 min. The gel with adsorbed samples placed on the graphite electrodes and focussed. The gel was immersed in the staining solution and destained.

RESULTS AND DISCUSSION

Cell lysis and IB isolation

Initially the pellet was mixed with the lysis buffer in the ratio of 1:20 (g/ml). As the target protein was intracellular, it had to be released from the cells by sonication. Lysis was done in four cycles and the lysis efficiency was observed by taking OD at 600 nm. The OD values decreased gradually in each cycles and at 4^{th} cycle there was over 95% lysis confirmed by microscope. Once the lysis optimization was done, recovery of IBs from cell homogenates were accomplished by centrifugation. The exact conditions for centrifugation should be evaluated for each protein since the size of the inclusion bodies can vary depending on the protein and expression level and their size will affect their rate of sedimentation.

Recovery of IBs

Washing of IBs was to improve the ease with which the desired protein can be subsequently refolded and purified. Here, various combinations of detergents were used for IBs washing. Initially four washes were done and analysed in SDS PAGE which still showed impurities. Some of the proteins (impurities) that were present along with inclusion bodies include outer membrane, RNA polymerase, ribosomal subunits and also heat shock proteins. These contaminants may be because of co-sedimentation or due to weak adsorption to the surface of IBs. Membrane fragments present in cytoplasmic inclusion bodies and other ionic impurities were removed by ionic and non-ionic detergents. Proteolysis was a problem until Triton X 100 washing step and the addition of protease inhibitor was included (Table 1). After four washes Tris and Triton washes were included and the resulting washed samples were loaded in 15% SDS-PAGE. From the gel it was evident that the impurities were eluted and minimized. The washing efficiency was almost 80% (Figure 1).

Solubilisation and refolding

The IB protein was solubilised using alkaline buffer and refolded in redox conditions. In general, proteins expressed as IBs in *E. coli* were solubilized by the use of chaotropic agents like urea and guanidine-HCl and then refolded to their native state by dialyzing in buffer containing reductants and oxidants. Solubilization of IB protein using mild conditions without disturbing the native structure was a key to high throughput recovery of r-proteins. 40 ml of 8 M urea was used at pH 12 for the complete solubilization for 45 min. Then overnight refolding was carried out in the ratio of 1:20 dilution with 800 ml of refolding buffer. Refolding of the protein was done from the pre-existing structure rather than from the unfolded molecules. Pulsatile refolding was carried out to reduce the protein aggregation. This combination of non-denaturing solubilization and pulsatile refolding resulted in improved recovery of r-protein.

Purification of r-HGCSF by SP sepharose column

SP sepharose fast flow was a strong cationic exchanger with excellent flow properties and high capacity for proteins of all pI values. The refolded samples went through three runs. In run 1 and 2 the sample pH was adjusted to 5.5 and the samples were collected (flow-through (FT) and wash). Elution was done at higher pH (7.5) with the same equilibration buffer. The flow rate (10 ml/min) was maintained throughout the run to get better resolution. In run 1 the FT showed higher UV and less protein recovery was observed in Lowry's. In run 2 it was observed that most of the proteins came in elution than in FT. To get better results, 3rd run was carried out. As the pI value of target protein was 5.9, 3rd run was done at pH

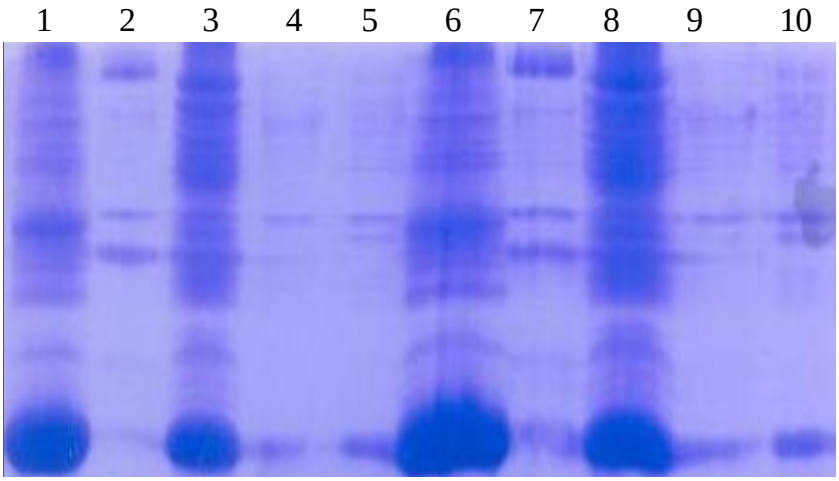

Figure 1. Ionic and non-ionic washes. The recovery of IBs after each washes were analysed by SDS PAGE with gel conditions used 15% reduced. Load volume was 20 µl. Lane 1, Tris IB; lane 2, Tris wash; lane 3, Triton IB; lane 4, triton wash; lane 5, IB solubilised; lane 6, Tris IB; lane 7, Tris wash; lane 8, Triton wash; lane 9, Triton IB; lane 10, IB solubilised. The stain used was commassie blue staining.

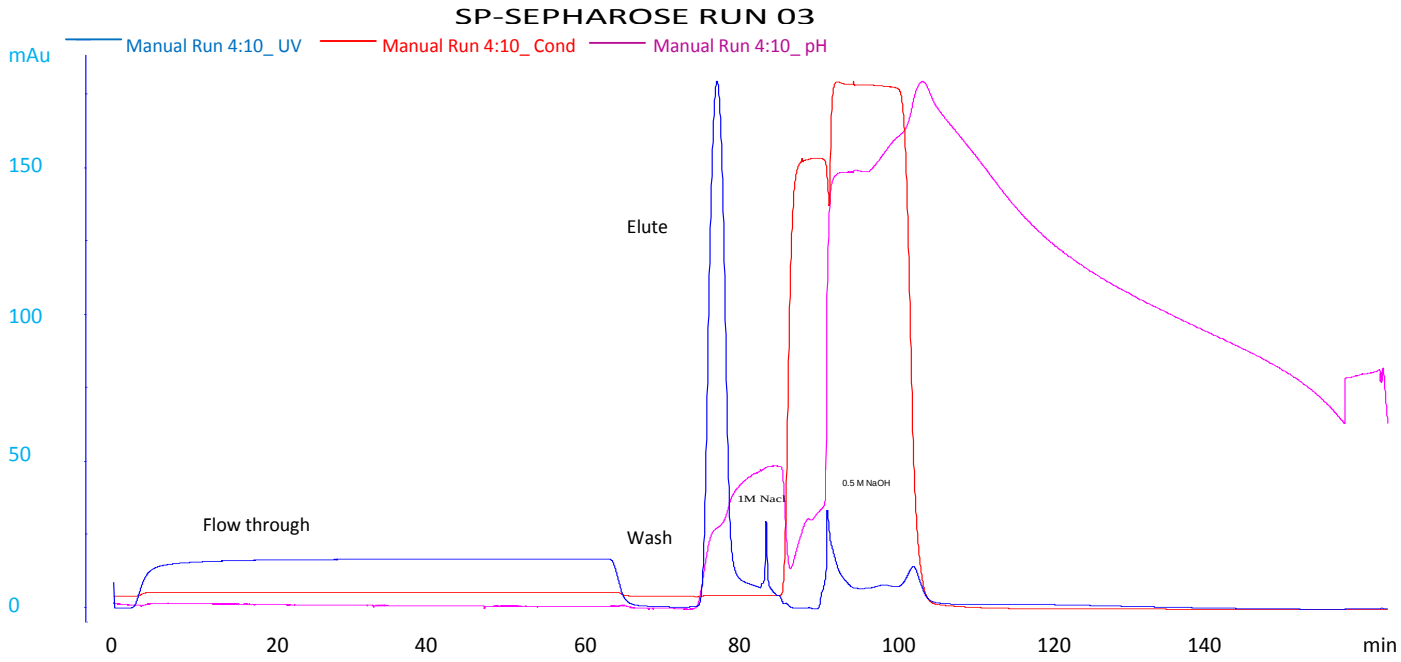

Figure 2. Chromatogram of SP-sepharose. The peak elute showed that all the refolded proteins had eluted and no protein was eluted in flow-through. NaOH wash was included to elute the protein which binds strongly to the column load volume: 20 µl.

5. Almost all the proteins came in elution and no peak was observed in NaOH wash in the chromatogram (Figure 2). All the samples were analysed in 15% SDS-PAGE. Silver staining was done to detect the nano level of proteins present in the samples (Figure 3). The highest recovery was observed in 3rd run which exhibits 98% of protein got refolded.

Determination of purity by HPLC

This method used a gradient high-performance reversed-phase liquid chromatography (HPLC) separation procedure coupled with UV absorbance detection for determining the purity of r-protein. Purity determination of r-protein was based on a peak area versus total area

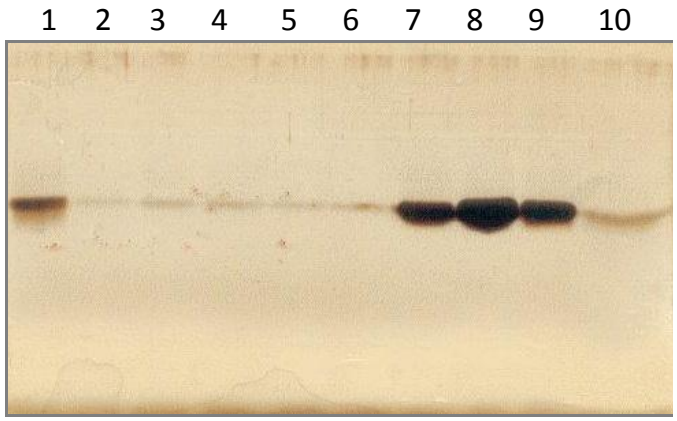

Figure 3. Recovery of refolded protein in run. The samples collected from all the three runs of Sp Sepharose were loaded in 15% reduced SDS PAGE gel and stained with silver staining. Load volume was 20 µl. Lane 1, Load; lane 2, flow-through 1; lane 3, flow-through 2; lane 4, flow-through 3; lane 5, flow-through 5; lane 6, wash; lane 7, elute fraction 1; lane 8, elute fraction 2; lane 9, 1M NaCl; lane 10, 0.5 NaOH.

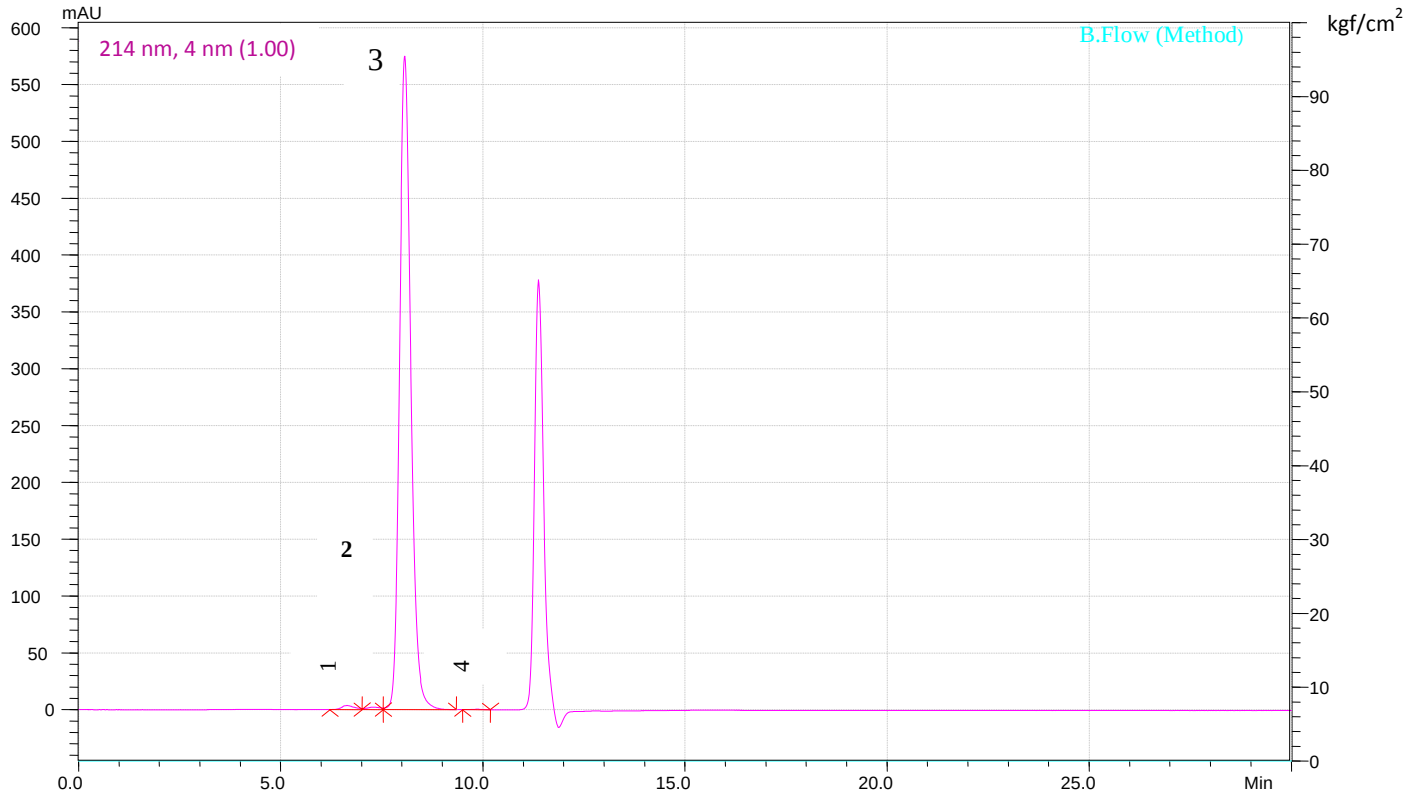

Figure 4. Chromatogram of HPLC. The 3rd peak showed higher area with 98% purity at 8th min.

calculation. The size exclusion column was used to determine whether any aggregates and degradants were present. 3rd peak showed higher area with purity of 98%. The peak was observed at the 8th min (Figure 4).

Identification of protein by isoelectric focusing (IEF)

An IEF technique separates the molecules according to their electric charge differences. The sample was loaded

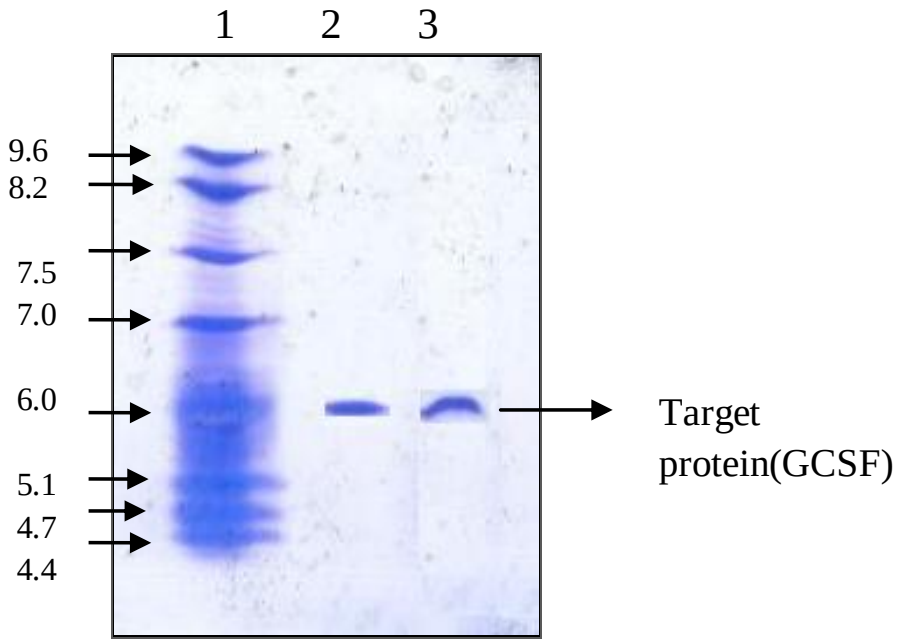

Figure 5. IEF gel. Lane 1, Marker; lane 2, standard GCSF; lane 3, purified GCSF. When compared with standard the target protein found to be closer to 6.

with 0.2% ampholytes which had a particular pK value. A pH gradient was established before loading the sample. The protein formed a band where it had no net charge. The protein Pi value was found to be 5.9 (Figure 5).

Conclusion

Large scale production of soluble proteins in *E. coli* has been hindered due to the formation of IBs. Many r-proteins form insoluble aggregates when synthesised at high levels in *E. coli*. However, studying the effects of the sequence and environment in all the possible permutations and combinations which exist in nature, on protein aggregation is a rather daunting task. Lysis of IBs was optimized at 4th cycle during sonication. Recovery of IBs by detergent washes were done and solubilised under alkaline condition made the protein more soluble without denaturing its native structure. Refolding under redox condition followed by purification using cation exchanger optimized to achieve the high throughput recovery of desired protein with purity more than 95%.

ACKNOWLEDGEMENT

The authors acknowledge Inbiopro Solutions Pvt, Bangalore, India for providing the laboratory facilities for carrying out the research work.

REFERENCES

Bhandari P, Gowrishankar J (1997). An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. J. Bacteriol., 13: 4403-4406

Wang CZ, Liu JF, Geng XD (2005). Refolding with Simultaneous Purification of Recombinant Human (GCSF) from *E. coli* Using Strong Anion Exchange Chromatography, Chinese Chem. Lett., 16(3):389-392.

Abolghasemi DS, Babaeipour V, Mofid MR, Divsalar A, Faraji F (2010). An efficient purification method for high recovery of Recombinant Human Granulocyte Colony Stimulating Factor from recombinant *E. coli*. International J. Environ. Sci. Dev., p. 1 & 2.

Susan IT, Petety VB (2007). Protein aggregation: A perspective from amyloid and inclusion-body formation. Curr. Sci., 92: 6.

Ahmed AK, Schaffer SW, Wetlaufer DB (1975). Non enzymatic reactivation of reduced buffers. J. Biochem., 250: 8477-8482.

Bryan H, Yini W, Xiao-Yang Z, Athens G (2007). Denaturing and refolding of protein molecules on surfaces. Proteomics, 17: 1771-1774.

Bei-Wen Y, Tatsuya N, Katsuyo S, Wenzhen J, Shoji T, Takuya U, Hideki T (2006). Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. Proc. Natl. Acad. Sci. U.S.A., 106: 4201-4206**.**

Chrunyk BA, Evans J, Lillquist J, Young P, Wetzel R (1993). Inclusion body formation and protein stability in sequence variants of interleukin-1 beta. J. Biol. Chem., 268(24): 18053-18061.

Chaozhan W, Lili W, Xindu G (2007). Renaturation with simultaneous purification of rGH-CSF from *Escherichia coli* by ion exchange chromatography". Biomed. Chromatogr., 21(12): 1291-1296.