Chemical synthesis and improved expression of recombinant human granulocyte colony-stimulating factor cDNA

S. Alrokayan[1,2]

[1]King Abdullah Institute for Nanotechnology, King Saud University, Riyadh, Saudi Arabia
[2]Center of Excellence in Biotechnology Research, College of Sciences, King Saud University, Riyadh, Saudi Arabia

Corresponding author: S. Alrokayan
E-mail: dr.salman@alrokayan.com

ABSTRACT. Recently, granulocyte colony-stimulating factor (G-CSF) has been recognized as a useful molecule for the treatment of a wide range of complex ailments, such as cancer, AIDS, H1N1 influenza, cardiac and neurological diseases. The vast therapeutic potential of G-CSF has induced scientists to develop biotechnological approaches for the synthesis of this pharmacologically active agent. We used a synthetic G-CSF cDNA molecule to produce the target protein by a simple cloning protocol. We constructed the synthetic cDNA using a DNA synthesizer with the aim to increase its expression level by specific sequence modifications at the 5' end of the G-CSF-coding region, decreasing the GC content without altering the predicted amino acid sequences. The identity of the resulting protein was confirmed by a highly specific enzyme-linked immunosorbent assay. In conclusion, a synthetic G-CSF cDNA in combina-

tion with the recombinant DNA protocol offers a rapid and reliable strategy for synthesizing the target protein. However, commercial utilization of this methodology will require rigorous validation and quality control.

Key words: Granulocyte colony-stimulating factor; Synthetic cDNA; Recombinant G-CSF; Cloning; Gene expression

INTRODUCTION

Granulocyte colony-stimulating factor (G-CSF) is a hematopoietic growth factor that stimulates differentiation and propagation of neutrophil precursor cells (Metcalf, 1985; Cohen et al., 1987; Liongue et al., 2009). The G-CSF gene has a length of 2.5 kb, with five exons, and it maps to chromosome 17. Two different human G-CSF mRNAs are generated by differential splicing of mRNA, resulting in the synthesis of two different polypeptides, composed of either 204 or 207 amino acids (Nagata et al., 1986). However, the mature protein is 174 or 177 residues long due to the cleavage of a 30-amino acid secretion signal sequence from the N-terminal end. Expression studies have shown that both these isoforms have authentic G-CSF activity, but the shorter form (174 residues) is at least 20-fold more potent in colony-stimulating activity than its longer counterpart (Nagata et al., 1986; Souza et al., 1986).

Recent studies have shown the promise of G-CSF for stem cell transplantations (Kim et al., 2009) and for treatment of a wide range of complex disorders, including AIDS (Heit et al., 2006), H1N1 influenza (Huang H et al., 2010), pulmonary alveolar proteinosis (Tazawa et al., 2010), epilepsy (Zhang et al., 2010), leukemia (Beekman and Touw, 2010), cancer (Gascon et al., 2010), myocardial infarction (Louzada et al., 2010), and cardiomyopathy (Macambira et al., 2009). The therapeutic potential of G-CSF against numerous life-threatening diseases has intensified the interest in developing a recombinant G-CSF (rG-CSF) molecule to cope with its future demand. Recombinant G-CSF has been shown to exhibit the biological properties of the native molecule (Souza et al., 1986). Several reports have reported the use of various cell lines and inducers for synthesis of G-CSF (Rotondaro et al., 1997; Jeong and Lee, 2001; Lasnik et al., 2001; Vanz et al., 2008; Jin et al., 2010; Babaeipour et al., 2010).

However, there is a growing need to develop biotechnological approaches to address the complexities of conventional cloning due to sequence bias of genes between prokaryotes and eukaryotes and hence to ensure efficient expression of human cDNA in an *Escherichia coli* system. Earlier, it has been shown that the codon-anticodon interaction seems to be so sticky that it interferes with the translation of hG-CSF in *E. coli*, due to the abundance of GC rich codons at the 5' end of hG-CSF cDNA (Delvin et al., 1988). Other relevant technological advancements include the use of PCR to produce the G-CSF gene and reverse transcription PCR to generate the cDNA to develop a prokaryote production system, avoiding the limitations of conventional cell-based methods (Kang et al., 1995).

We developed a different approach for constructing the synthetic cDNA, using a DNA synthesizer instead of RT-PCR, with the aim to increase its expression level by specific sequence modifications at the 5' end of the G-CSF-coding region and decreasing the GC content

without altering the predicted amino acid sequence.

MATERIAL AND METHODS

pETBlue-1 was purchased from Novagen; *E. coli* C41(DE3) and C43(DE3) cells were obtained from OverExpress. Restriction endonucleases and T4 DNA ligase were purchased from Roche. The human G-CSF ELISA kit was purchased from RayBiotech. All other reagents were of analytical grade.

General protocols

Methods for plasmid purification, subcloning, and bacterial transformation were followed as described earlier (Maniatis et al., 1982). Recombinant DNA methods, including transformations, agarose and polyacrylamide gel electrophoresis, small- and large-scale preparation of plasmids, restriction fragment and single-stranded DNA preparations, isolation of fragments from agarose gels, phosphorylation, dephosphorylation, ligation, and enzyme digestions were carried out using standard protocols (Maniatis et al., 1982).

Design and synthesis of hrG-CSF cDNA

A cDNA was designed with the help of the Vector NT software to code for hrG-CSF and to maximize translational initiation and protein elongation in *E. coli*. To facilitate the initiation of translation, the nucleotide sequence of the cDNA was adjusted to expose the start codon and Shine-Dalgarno sequence of its transcript in vector pETBlue-1. To facilitate protein elongation, the codons that were used were those found in highly expressed genes of *E. coli*. The nucleotide sequence of the synthetic gene is given in Figure 1 and the restriction map is shown in Figure 2. The custom synthesized gene was obtained from GeneArt (Germany).

Cloning of G-CSF cDNA

The G-CSF cDNA was incorporated into the pETBlue-1 vector using *Eco*RV and *Eco*RI restriction sites (Figure 3). The C43 (DE3) competent cells were used to transform into pETBlue1•G-CSF expressing cells. The cells took in the plasmid vector with high transformation efficiency. Single colonies formed on the LB (Luria Bertani) agar plates after incubation at 37°C overnight. Forty colonies were randomly selected from the plates while the screening was performed on the first 10 colonies to determine whether the cells had taken the plasmid. Testing was done by plasmid mini-preps, followed by restriction digestion. The QIAprep Spin Mini-Prep Kit (QIAGEN) protocol was used for the fast plasmid preparation on a small scale; plasmid DNA purified with this method is immediately ready for use. From the plasmid mini-preps, 3 µg of the DNA was digested with *Eco*RI and *Eco*RV in a total reaction mixture of 20 µL. The reaction mixtures were incubated at 37°C in a water bath for 1 h. Samples were then loaded on a 0.7% agarose mini-gel and electrophoresed for 1 h at 100 V. The electrophoresis showed the separation of the plasmid DNA into two bands; the upper fragment corresponded to the size of the pETBlue1 vector and the lower band corresponded to the G-CSF gene. The sizes of the two DNA fragments were checked by parallel running of a known marker.

```
  1  GCAAGGCGAT TAAGTTGGGT AACGCCAGGA TTCTCCCAGT CACGACGTTG
     CGTTCCGCTA ATTCAACCCA TTGCGGTCCT AAGAGGGTCA GTGCTGCAAC

 51  TAAAACGACG GCCAGCGAGA GATCTTGATT GGCTAGCAGA ATAATTTTGT
     ATTTTGCTGC CGGTCGCTCT CTAGAACTAA CCGATCGTCT TATTAAAACA

                                          M   T   P   L   G   P   A
101  TTAACTTTAA GAAGGAGATA TAGATATCAT GACCCCGCTG GGTCCGGCGA
     AATTGAAATT CTTCCTCTAT ATCTATAGTA CTGGGGCGAC CCAGGCCGCT

      S   S   L   P   Q   S   F   L   L   K   C   L   E   Q   V   R   K
151  GCAGCCTGCC GCAGAGCTTT CTGCTGAAAT GCCTGGAACA GGTGCGCAAA
     CGTCGGACGG CGTCTCGAAA GACGACTTTA CGGACCTTGT CCACGCGTTT

        I   Q   G   D   G   A   A   L   Q   E   K   L   C   A   T   Y   K
201  ATTCAGGGCG ATGGTGCGGC GCTGCAGGAA AAACTGTGCG CGACCTATAA
     TAAGTCCCGC TACCACGCCG CGACGTCCTT TTTGACACGC GCTGGATATT

          L   C   H   P   E   E   L   V   L   L   G   H   S   L   G   I
251  ACTGTGCCAT CCGGAAGAAC TGGTGCTGCT GGGCCATAGC CTGGGCATTC
     TGACACGGTA GGCCTTCTTG ACCACGACGA CCCGGTATCG GACCCGTAAG

          P   W   A   P   L   S   S   C   P   S   Q   A   L   Q   L   A   G
301  CATGGGCACC GCTGAGCAGC TGCCCGAGCC AGGCGCTGCA GCTGGCGGGC
     GTACCCGTGG CGACTCGTCG ACGGGCTCGG TCCGCGACGT CGACCGCCCG

          C   L   S   Q   L   H   S   G   L   F   L   Y   Q   G   L   L   Q
351  TGCCTGAGCC AGCTGCATAG CGGCCTGTTT CTGTATCAGG GCCTGCTGCA
     ACGGACTCGG TCGACGTATC GCCGGACAAA GACATAGTCC CGGACGACGT
          A   L   E   G   I   S   P   E   L   G   P   T   L   D   T   L
401  GGCGCTGGAA GGCATTAGCC CGGAACTGGG CCCGACCCTG GATACCCTGC
     CCGCGACCTT CCGTAATCGG GCCTTGACCC GGGCTGGGAC CTATGGGACG

          Q   L   D   V   A   D   F   A   T   T   I   W   Q   Q   M   E   E
451  AGCTGGATGT GGCGGATTTT GCGACCACCA TTTGGCAGCA GATGGAAGAA
     TCGACCTACA CCGCCTAAAA CGCTGGTGGT AAACCGTCGT CTACCTTCTT

          L   G   M   A   P   A   L   Q   P   T   Q   G   A   M   P   A   F
501  CTGGGCATGG CACCGGCGCT GCAGCCGACC CAGGGCGCGA TGCCGGCGTT
     GACCCGTACC GTGGCCGCGA CGTCGGCTGG GTCCCGCGCT ACGGCCGCAA

          A   S   A   F   Q   R   R   A   G   G   V   L   V   A   S   H
551  TGCGAGCGCG TTTCAGCGTC GTGCGGGCGG TGTGCTGGTT GCGAGCCATC
     ACGCTCGCGC AAAGTCGCAG CACGCCCGCC ACACGACCAA CGCTCGGTAG

          L   Q   S   F   L   E   V   S   Y   R   V   L   R   H   L   A   Q
601  TGCAGAGCTT CCTGGAAGTG AGCTATCGTG TGCTGCGTCA TCTGGCGCAG
     ACGTCTCGAA GGACCTTCAC TCGATAGCAC ACGACGCAGT AGACCGCGTC

651  CCGTAATGAA TTCCTGCCCG GGCGTTGTAA TCATAGTCAT AATCAATACT
     GGCATTACTT AAGGACGGGC CCGCAACATT AGTATCAGTA TTAGTTATGA

701  CCTGACTGCG TTAGCAATTT AACTGTGATA AACTACCGCA TTAAAGCTAT
     GGACTGACGC AATCGTTAAA TTGACACTAT TTGATGGCGT AATTTCGATA
```

Figure 1. Nucleotide sequence of the synthetic G-CSF cDNA. The corresponding protein sequence is shown above the gene sequence.

PstI (524)
NaeI (545)
PstI (605)
EcoRI (659)
NciI (669)
PstI (341)
PstI (401)
NciI (670)
NcoI (301)
NciI (421)
SmaI (670)
HinfI (30)   EcoRV (126)   PstI (227)   StyI (301)   PstI (452)

G-CSF gene EcoRV EcoRI
750 bp
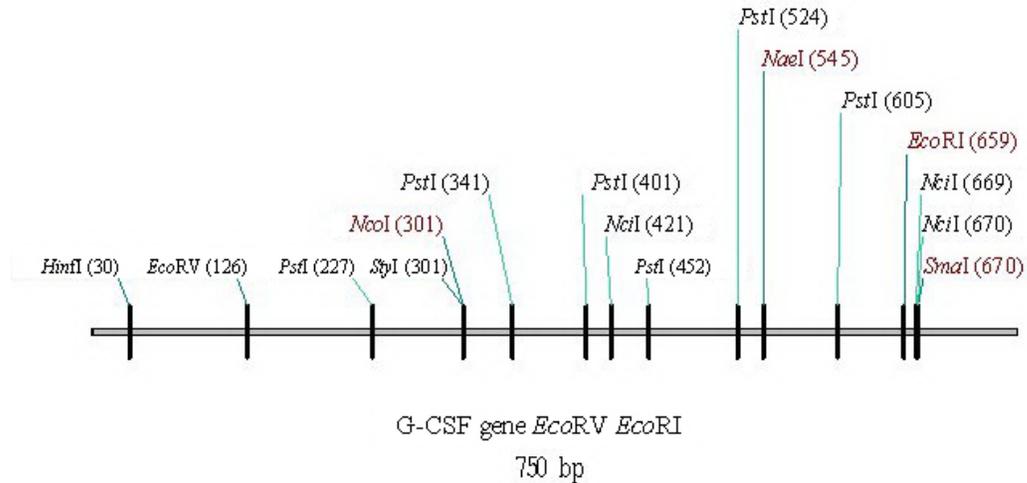
Figure 2. Restriction map of the synthetic G-CSF gene showing 35 unique restriction sites. Numbers in parentheses refer to the nucleotide position at the cutting site.

T7 transcription start
lac operator
lac operator
T7 promoter
RBS
EcoRV (279)
G-CSF cDNA
EcoRI (812)
pETBlue-1-G-CSF
4004 bp
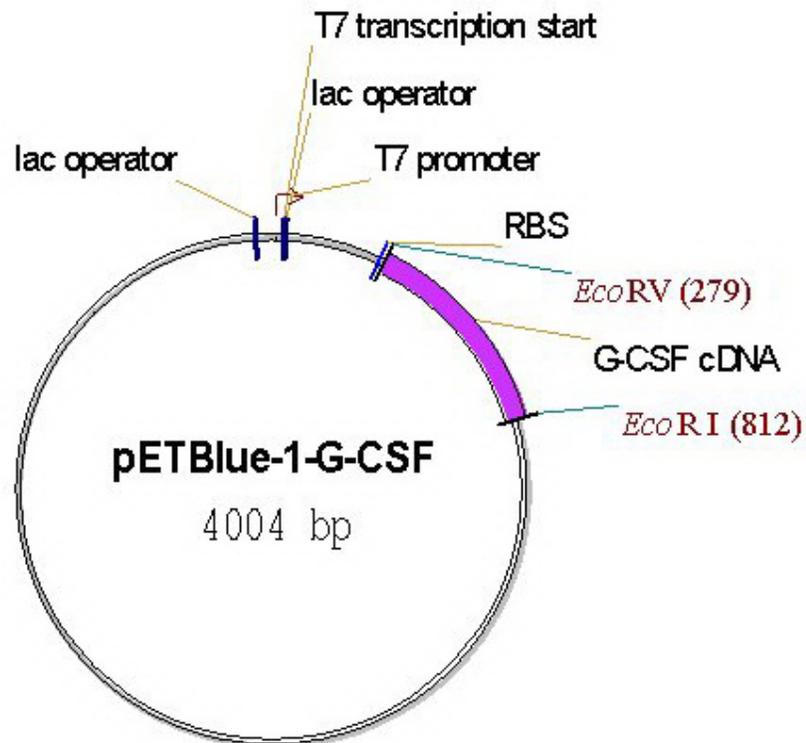
Figure 3. Structure of the expression vector (pET) [pETG-CSF construct]. The G-CSF insert is shown between the *Eco*RV and *Eco*RI sites.

Gene expression

All the 10 colonies with confirmed transformation for the expression vector were grown and cultivated on 50 mL LB media containing ampicillin. For the initial culture, 100 µL of the cells (which were stored in an aliquot with 80% glycerol) was added to 3 mL LB media and incubated overnight with shaking at 37°C. The following morning, 2.5 mL from the initial culture was added to 50 mL LB media and the cells were grown until an absorbance of 0.5-0.7 at 600 nm was reached. Ten microliters IPTG (100 mM) was then added to induce the production of hG-CSF. After 3 h, the cells were harvested by centrifugation at 4000 g for 20 min at 4°C.

Protein extraction and purification

Cell pellets were lysed using a lysis solution (50 mM sodium phosphate, 300 mM sodium chloride and 10 mM imidazole) and were solubilized in 2% digitonin and 20 mM Tris. Cellular debris was removed by centrifugation and the crude extract was loaded onto a hydroxylapatite column (2.5 x 8 cm) (Bio-Gel HTP, Bio-Rad) that was equilibrated with buffer A (25 mM KH_2PO_4, pH 6.9, 0.5 mM EDTA). Protein was eluted at 1 mL/min with 150 mL buffer A, followed by a linear gradient (200 mL) to 350 mM KH_2PO_4, pH 7.0, 0.5 mM EDTA. Column fractions of 5 mL were collected and those containing the target protein were pooled, dialyzed against 10 mM phosphate, pH 7.0, 0.5 mM EDTA, and concentrated with Aquacide II (Calbiochem).

Protein identification using ELISA

The authenticity of the protein was confirmed by a highly specific human G-CSF ELISA kit, using 50-fold diluted samples. This assay kit is highly sensitive, with a detection limit of as low as 20 pg G-CSF protein/mL. This kit employs an antibody specific for human G-CSF coated on a 96-well plate. All the reagents and samples were brought to room temperature before starting the analysis. One hundred microliters of standards and samples were pipetted into the wells and the plate was incubated for 2.5 h at room temperature, to allow the G-CSF present in the aliquot to bind to the wells onto the immobilized antibody. The solution was discarded and the wells were washed four times using 200 µL wash solution. Biotinylated antihuman G-CSF antibody (100 µL) was added to each well and the plate was incubated for 1 h at room temperature. After washing (four times with 200 µL wash solution), unbound biotinylated antibody, 100 µL HRP-conjugated streptavidin was added to each well and the plate was incubated for 45 min at room temperature. After discarding the solution, the wells were again washed five times using 200 µL wash solution, followed by the addition of 100 µL tetramethylbenzidine substrate solution and further incubation for 30 min at room temperature in the dark. Finally, 50 µL stop solution (changes the color from blue to yellow) was added to each well and the intensity of the color was measured at 450 nm.

RESULTS AND DISCUSSION

We were able to clearly demonstrate the suitability and validity of chemically synthe-

sized G-CSF cDNA (Figure 1) for large-scale cloning production purposes. Recently, it has been observed that rhG-CSF has better serum and whole blood stability and higher bioavailability than wild-type G-CSF (Liu and Jiang, 2010). Pavisic et al. (2010) documented a comprehensive protein stabilizing screening study, including conformational analysis of formulated rhG-CSF protein, to obtain information on its secondary structure conformational stability under the influence of stabilizing carbohydrates and polyols. Huang YS et al. (2010) designed an artificial gelatin-like protein and fused this hydrophilic GLK polymer to G-CSF to generate a chimeric fusion protein that displayed a slower plasma clearance rate and stimulated greater and longer lasting increases in circulating white blood cells than native G-CSF.

We chose the pETBlue-1 vector because it has high yield protein expression obtained with pET vectors. Expression of the target gene is driven by a T7lac promoter in the opposite orientation. pETBlue-1 facilitates the expression of unfused, native proteins from inserts that start with an ATG initiation codon at the 5' end and encode an open reading frame. The *Eco*RV (GATATC) blunt cloning site in this vector is optimally positioned relative to the strong T7 gene 10 ribosome binding site. Because the ATG start codon must be positioned 5-11 bp downstream of the ribosome binding site in order to obtain efficient translation initiation, the *Eco*RV cloning site must be used for T7 promoter mediated expression in pETBlue-1. There are two alternative designs of insert 5' ends that allow optimal expression. The first option is for the insert to begin with an ATG (Met) start codon, which creates 8 bp spacing between the RBS and ATG. The other option is for the insert to begin with a G-nucleotide that completes the ATG triplet, which creates 6 bp spacing between the ribosome binding site and ATG. Both configurations appear to produce similar expression levels.

We used the C41(DE3) and C43(DE3) strains of the *E. coli* for the cloning of pETBlue1•G-CSF expression vector; the cell showed comparable competencies. The bacterial strains, C41(DE3) and C43(DE3), were derived from the original strain to overcome many of its defects and no longer requires pLysS or pLysE. The new strains successfully expressed 10 globular and seven membrane proteins, which the original strain could not. Since then, many more proteins have been found that can be successfully over-expressed only in the improved strains C41(DE3) and C43(DE3). Even proteins that are expressed in BL21(DE3) are regularly better expressed (4- to 6-fold, on average) and more often soluble in the derived strains, C41(DE3) and C43(DE3).

In conclusion, we were able to demonstrate efficient use of synthetic G-CSF cDNA in combination with recombinant DNA protocols for rapid and reliable synthesis of the target genes and thus the commercial level end-product, the specific protein. Further studies are warranted to verify the functional activity of biotechnologically produced G-CSF, particularly as it affects cell proliferation.

ACKNOWLEDGMENTS

REFERENCES

Babaeipour V, Abbas MP, Sahebnazar Z and Alizadeh R (2010). Enhancement of human granulocyte-colony stimulating factor production in recombinant *E. coli* using batch cultivation. *Bioprocess Biosyst. Eng.* 33: 591-598.

Beekman R and Touw IP (2010). G-CSF and its receptor in myeloid malignancy. *Blood* 115: 5131-5136.

Cohen AM, Zsebo KM, Inoue H, Hines D, et al. (1987). *In vivo* stimulation of granulopoiesis by recombinant human granulocyte colony-stimulating factor. *Proc. Natl. Acad. Sci. U. S. A.* 84: 2484-2488.

Devlin PE, Drummond RJ, Toy P, Mark DF, et al. (1988). Alteration of amino-terminal codons of human granulocyte-colony-stimulating factor increases expression levels and allows efficient processing by methionine aminopeptidase in *Escherichia coli*. *Gene* 65: 13-22.

Gascon P, Fuhr U, Sorgel F, Kinzig-Schippers M, et al. (2010). Development of a new G-CSF product based on biosimilarity assessment. *Ann. Oncol.* 21: 1419-1429.

Heit B, Jones G, Knight D, Antony JM, et al. (2006). HIV and other lentiviral infections cause defects in neutrophil chemotaxis, recruitment, and cell structure: immunorestorative effects of granulocyte-macrophage colony-stimulating factor. *J. Immunol.* 177: 6405-6414.

Huang H, Li H, Zhou P and Ju D (2010). Protective effects of recombinant human granulocyte macrophage colony stimulating factor on H1N1 influenza virus-induced pneumonia in mice. *Cytokine* 51: 151-157.

Huang YS, Wen XF, Wu YL, Wang YF, et al. (2010). Engineering a pharmacologically superior form of granulocyte-colony-stimulating factor by fusion with gelatin-like-protein polymer. *Eur. J. Pharm. Biopharm.* 74: 435-441.

Jeong KJ and Lee SY (2001). Secretory production of human granulocyte colony-stimulating factor in *Escherichia coli*. *Protein Expr. Purif.* 23: 311-318.

Jin CY, Han SY, Kwon K, Yun EY, et al. (2010). Production of recombinant human granulocyte macrophage colony-stimulating factor from silkworm *Bombyx mori* Bm5 cells. *Z. Naturforsch. C.* 65: 153-156.

Kang SH, Na KH, Park JH, Park CL, et al. (1995). High level expression and simple purification of recombinant human granulocyte colony stimulating factor in *E. coli*. *Biotechnol. Lett.* 17: 687-692.

Kim SN, Moon JH, Kim JG, Chae YS, et al. (2009). Mobilization effects of G-CSF, GM-CSF, and darbepoetin-alpha for allogeneic peripheral blood stem cell transplantation. *J. Clin. Apher.* 24: 173-179.

Lasnik MA, Porekar VG and Stalc A (2001). Human granulocyte colony stimulating factor (hG-CSF) expressed by methylotrophic yeast pichia pastoris. *Pflugers Arch.* 442: R184-R186.

Liongue C, Wright C, Russell AP and Ward AC (2009). Granulocyte colony-stimulating factor receptor: stimulating granulopoiesis and much more. *Int. J. Biochem. Cell Biol.* 41: 2372-2375.

Liu XX and Jiang YP (2010). Pharmacokinetic study of a novel recombinant human granulocyte colony-stimulating factor in rats. *Chin. Med. Sci. J* 25: 13-19.

Louzada RA, Oliveira PF, Cavalcanti-de-Albuquerque JP, Cunha-Carvalho L, et al. (2010). Granulocyte-colony stimulating factor treatment of chronic myocardial infarction. *Cardiovasc. Drugs Ther.* 24: 121-130.

Macambira SG, Vasconcelos JF, Costa CR, Klein W, et al. (2009). Granulocyte colony-stimulating factor treatment in chronic Chagas disease: preservation and improvement of cardiac structure and function. *FASEB J.* 23: 3843-3850.

Maniatis T, Fritsch EF and Sambrook J (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Lab., Cold Spring Harbor, Nova York.

Metcalf D (1985). The granulocyte-macrophage colony stimulating factors. *Cell* 43: 5-6.

Nagata S, Tsuchiya M, Asano S, Yamamoto O, et al. (1986). The chromosomal gene structure and two mRNAs for human granulocyte colony-stimulating factor. *EMBO J.* 5: 575-581.

Pavisic R, Hock K, Mijic I, Horvatic A, et al. (2010). Recombinant human granulocyte colony stimulating factor pre-screening and screening of stabilizing carbohydrates and polyols. *Int. J. Pharm.* 387: 110-119.

Rotondaro L, Mazzanti L, Mele A and Rovera G (1997). High-level expression of a cDNA for human granulocyte colony-stimulating factor in Chinese hamster ovary cells. Effect of 3'-noncoding sequences. *Mol. Biotechnol.* 7: 231-240.

Souza LM, Boone TC, Gabrilove J, Lai PH, et al. (1986). Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells. *Science* 232: 61-65.

Tazawa R, Trapnell BC, Inoue Y, Arai T, et al. (2010). Inhaled granulocyte/macrophage-colony stimulating factor as therapy for pulmonary alveolar proteinosis. *Am. J. Respir. Crit. Care Med.* 181: 1345-1354.

Vanz AL, Renard G, Palma MS, Chies JM, et al. (2008). Human granulocyte colony stimulating factor (hG-CSF): cloning, overexpression, purification and characterization. *Microb. Cell Fact.* 7: 13.

Zhang Y, Wang W, Sun Z, Feng D, et al. (2010). Granulocyte colony-stimulating factor treatment prevents cognitive impairment following status epilepticus in rats. *Biol. Pharm. Bull.* 33: 572-579.