

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

04 July 2019 (04.07.2019)



(10) International Publication Number

WO 2019/130344 A1

(51) International Patent Classification:

C07K 14/535 (2006.01) A61K 38/00 (2006.01)

(21) International Application Number:

PCT/IN2018/050873

(22) International Filing Date:

24 December 2018 (24.12.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

201711022403 27 December 2017 (27.12.2017) IN

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(71) Applicant: COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH [IN/IN]; Anusandhan Bhawan, 2 Rafi Marg, New Delhi 110 001 (IN).

(72) Inventors: DATTA, Sonal; Institute of Microbial Technology, Post Box No. 1304, Sector 39A, Chandigarh 160036 (IN). SAHNI, Girish; Institute of Microbial Technology, Post Box No. 1304, Sector 39A, Chandigarh 160036 (IN).

(74) Agent: SINGH, Manisha; LexOrbis, 709-710 Tolstoy House, 15-17 Tolstoy Marg, New Delhi 110001 (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: A POLYPEPTIDE EXHIBITING GRANULOCYTE-COLONY STIMULATING FACTOR ACTIVITY

(57) Abstract: The present disclosure relates to a polypeptide exhibiting granulocyte-colony stimulating factor activity. The polypeptide comprises at least one non-native cysteine residue at a site selected from the group consisting of T₁CP₂ (SEQ ID NO: 25), P₂CL₃ (SEQ ID NO: 26), L₃CG₄ (SEQ ID NO: 27), G₄CP₅ (SEQ ID NO: 28), P₅CA₆ (SEQ ID NO: 29), A₆CS₇ (SEQ ID NO: 30), S₉₆CP₉₇ (SEQ ID NO: 31), P₉₇CE₉₈ (SEQ ID NO: 32), L₉₉CG₁₀₀ (SEQ ID NO: 33), P₁₀₁CT₁₀₂ (SEQ ID NO: 34), E₁₂₂CE₁₂₃ (SEQ ID NO: 35), L₁₂₄CG₁₂₅ (SEQ ID NO: 36), M₁₂₆CA₁₂₇ (SEQ ID NO: 37), P₁₃₈CA₁₃₉ (SEQ ID NO: 39), A₁₄₃CF₁₄₄ (SEQ ID NO: 40), R₁₄₆CR₁₄₇ (SEQ ID NO: 41), R₁₆₉CH₁₇₀ (SEQ ID NO: 42), H₁₇₀CL₁₇₁ (SEQ ID NO: 43), L₁₇₁CA₁₇₂ (SEQ ID NO: 44), A₁₇₂CQ₁₇₃ (SEQ ID NO: 45), and Q₁₇₃CP₁₇₄ (SEQ ID NO: 46) in an amino acid sequence having at least 90% sequence identity to sequence set forth in SEQ ID NO: 2.



WO 2019/130344 A1

A POLYPEPTIDE EXHIBITING GRANULOCYTE-COLONY STIMULATING FACTOR ACTIVITY

FIELD OF THE INVENTION

5

The present invention generally relates to variants of Granulocyte-Colony Stimulating Factor ('G-CSF'), conjugates of the variants of G-CSF, methods for preparing such variants and conjugates and use of such variants or conjugates in therapy particularly neutropenia and leucopenia.

10 BACKGROUND OF THE INVENTION

Chemotherapy constitutes an indispensable component of the treatment of various forms of lymphomas and metastatic cancers. Chemotherapy is believed to suppress the hematopoietic system and thus weaken the host immune system. Neutropenia is the most critical
15 haematological toxicity associated with the chemotherapy [Crawford, J. et al., Cancer, 2004. 100(2): p. 228-37]. Chemotherapy induced neutropenia is characterized by reduction in the number of neutrophils, which leads to predisposition of cancer patients to fatal infections and sepsis [Lyman, G.H., Clin Cornerstone, 2006. 8 Suppl 5: p. S12-8] Clinical management of neutropenia includes a lengthy hospital stay, along with administration of high-end
20 antibiotics, which leads to tremendous escalation in cost of cancer treatment [Kuderer, N.M., et al., Cancer, 2006. 106(10): p. 2258-66]. The prophylactic and therapeutic administration of Colony-Stimulating Factors (CSFs) has proven to be extremely effective at significantly reducing the risk of neutropenia in patients receiving dose-intensive chemotherapy [Lyman, G.H. et al., Am J Med, 2002. 112(5): p. 406-11].
25 CSFs are cytokines that guide the hematopoietic system to generate specific types of white blood cells [Metcalf, D., Cancer, 1990. 65(10): p. 2185-95]. These factors were discovered in effort to grow the bone marrow cells *in vitro* [Bradley, T.R. et al. Nature, 1967. 213(5079): p. 926-7; Pluznik, D.H. et al. Experimental cell research, 1966. 43(3): p. 553-63 and Bradley, T.R. et al. The Australian journal of experimental biology and medical science,
30 1966. 44(3): p. 287-99]. Two types of CSFs are employed for prophylaxis of neutropenia namely - Granulocyte Macrophage-CSF (GM-CSF) and G-CSF. The most commonly used CSF is recombinant human G-CSF (rHuG-CSF). The administration of rHuG-CSF stimulates the production of mature functional neutrophils and thus reduces the risk of neutropenia [Welte, K., et al., Blood, 1996. 88(6): p. 1907-29]. Despite, the discovery of the

CSFs in 1960s, the human G-CSF was first purified and characterized by Moore and colleagues in 1985 from bladder carcinoma cell line 5637 [Welte, K., et al., Blood, 1996. 88(6): p. 1907-29]. This naturally produced G-CSF from the bladder carcinoma cell line 5637 is O-glycosylated and has molecular weight of 19.6 kDa that yield a product of 18.8 kDa upon treatment with O-glycanase [Souza, L.M., et al., Science, 1986. 232(4746): p. 61-5]. Nagata and co-workers utilized N-terminal amino acid sequencing to identify the cDNA sequence of G-CSF [Nagata, S., et al., Nature, 1986. 319(6052): p. 415-8]. Souza et al., further cloned, sequenced and expressed the identified sequence. The recombinant protein resulting from the expression of G-CSF cDNA in *Escherichia coli* was found to be capable of inducing proliferation and differentiation of human bone marrow cells [Souza, L.M., et al., Science, 1986. 232(4746): p. 61-5.]]. The administration of rHuG-CSF was approved by the United States Food and Drug Administration (the 'FDA') in 1991 and since then has been actively used for prevention of neutropenia, not only for patients receiving chemotherapy but also for patients suffering from leucopenia, AIDS, sepsis, and patients undergoing bone marrow transplantation [Keating, G.M., Drugs. 71(6): p. 679-707]. G-CSF binds to the extracellular immunoglobulin like domain and the cytokine receptor homologue domain on the Granulocyte-Colony Stimulating Factor Receptor (G-CSFR) This binding leads to homodimerization of G-CSFR and initiates the activation of JAK-STAT and mitogen-activated protein kinase pathways to execute the effects of G-CSF [Van de Geijn, G.J., et al., Rev Physiol Biochem Pharmacol, 2003. 149: p. 53-71]. The recombinant G-CSF is often produced in *E. coli* in the non-glycosylated form (e.g. Filgrastim) although a glycosylated protein (e.g. Lenograstim) produced in Chinese hamster ovary cells (CHO) is also in use [Fernandez-Varon, E. et al. Vet J, 2007. 174(1): p. 33-41]. Both non-glycosylated form and glycosylated forms have been reported to have equivalent activities and similar pharmacokinetics in cynomolgus monkeys[Tanaka, H., et al., Cytokine, 1997. 9(5): p. 360-9]. Filgrastim and the glycosylated Lenograstim, which is CHO derived G-CSF enhances the proliferation and differentiation of neutrophil precursors, migration of neutrophils in blood and tissues and increases the activity of mature neutrophils to prevent neutropenia. However, filgrastim has a half-life of 3–4 hr and needs to be administered daily [Kuwabara, T., et al. Drug metabolism reviews, 1996. 28(4): p. 625-58]

G-CSF is thought to be cleared using several different mechanisms, including receptor-mediated endocytosis [Kuwabara, T., et al., The American journal of physiology, 1995. 269(1 Pt 1): p. E1-9.], renal clearance [Kuwabara, T., et al., Pharmaceutical research, 1995. 12(10): p. 1466-9] and enzymatic degradation mechanism [El Ouriaghli, F., et al., Blood, 2003.

101(5): p. 1752-8.]. Several approaches have been employed to increase the serum half-life of rHuG-CSF. Various modifications such as conjugation with Polyethylene glycol ('PEG'), known as PEGylation [Molineux, G., Current pharmaceutical design, 2004. 10(11): p. 1235-44], conjugation with the sialic acid [Andersen, D.C. et al. Biotechnology and
5 bioengineering, 1995. 47(1): p. 96-105], attachments with the human serum albumin [Halpern, W., et al., Pharmaceutical research, 2002. 19(11): p. 1720-9] etc. have been utilized. However, PEGylation has emerged as the method of choice for this purpose. PEGylation increases serum half-life of therapeutic proteins, including masking the protein's surface to shield it from proteases, antibodies and antigen processing cells and increasing the molecular
10 size to reduce the renal ultrafiltration. Furthermore, PEG imparts favourable attributes on the polypeptides to improve the biological distribution and solubility [Harris, J.M. et al. Drug discovery, 2003. 2(3): p. 214-21]. In the year 2002, the FDA approved the use of PEGylated rHuG-CSF or Pegfilgrastim as a prophylactic drug in chemotherapy. PEGylated G-CSF has significantly improved serum half-life and thus is administered once per cycle of
15 chemotherapy [Crawford, J., Drugs, 2002. 62 Suppl 1: p. 89-98] compared to daily dose of filgrastim.

The currently used PEGylated G-CSF conjugated with 20 kDa PEG molecule at the first amino acid on the N-terminal of the human G-CSF protein using reductive alkylation method [US 5,824,784]. Although, this method results in efficient PEGylation of G-CSF, it is
20 associated with heterogeneous PEGylation at the amino group of lysine residues present in protein in addition to PEGylation at the N-terminal amino group. G-CSF variants having multiple PEGylation have also been created for administration on the same-day treatment of chemotherapy-induced neutropenia [US 20090203601]. However, this method and other methods using conjugation of PEG molecule at the amino group of the N-terminal leads to a
25 heterogeneous population of conjugated G-CSF due to variable levels of PEGylation. Such heterogeneous populations of drug molecules create difficulty in accurately predicting the biological activity. Recently, new methods of the site directed PEGylation had been described [Crawford, J., Drugs, 2002. 62 Suppl 1: p. 89-98, Goodson, R.J. and N.V. Katre, Bio/technology, 1990. 8(4): p. 343-6-30 and Xiong, C.Y., et al., Protein engineering, design
30 & selection : PEDS, 2006. 19(8): p. 359-67] generating more homogenously PEGylated proteins.] One of the common methods of PEGylation uses cysteine of the protein or a suitably introduced cysteine residue for PEGylation [US 5766897A].

G-CSF variants having multiple PEGylation have also been created for administration on the same-day treatment of chemotherapy-induced neutropenia [US 20090203601] . However,

this method leads to a heterogeneous population of conjugated G-CSF due to variable levels of PEGylation. Such heterogeneous populations of drug molecules create difficulty in accurately predicting the biological activity.

Using this approach, cysteine 17 of the G-CSF was exploited for generating PEGylated version of G-CSF [Ishikawa, M., et al., Cell structure and function, 1992. 17(1): p. 61-5]. However, since the cysteine 17 is partially buried in a hydrophobic pocket, this method requires denaturation followed by PEGylation and the renaturation of the protein.

There is a need for new and improved polypeptides that exhibit G-CSF activity and conjugates thereof that can be used in therapy, for example in the treatment of neutropenia and leucopenia and at the same time do not suffer from the disadvantages of currently available polypeptides and their conjugates.

SUMMARY OF THE INVENTION

The present disclosure relates to a polypeptide exhibiting G-CSF activity. The polypeptide comprises at least one non-native cysteine residue at a site selected from the group consisting of T₁CP₂ (SEQ ID NO: 25), P₂CL₃ (SEQ ID NO: 26), L₃CG₄ (SEQ ID NO: 27), G₄CP₅ (SEQ ID NO: 28), P₅CA₆ (SEQ ID NO: 29), A₆CS₇ (SEQ ID NO: 30), S₉₆CP₉₇ (SEQ ID NO: 31), P₉₇CE₉₈ (SEQ ID NO: 32), L₉₉CG₁₀₀ (SEQ ID NO: 33), P₁₀₁CT₁₀₂ (SEQ ID NO: 34), E₁₂₂CE₁₂₃ (SEQ ID NO: 35), L₁₂₄CG₁₂₅ (SEQ ID NO: 36), M₁₂₆CA₁₂₇ (SEQ ID NO: 37), P₁₃₈CA₁₃₉ (SEQ ID NO: 39), A₁₄₃CF₁₄₄ (SEQ ID NO: 40), R₁₄₆CR₁₄₇ (SEQ ID NO: 41), R₁₆₉CH₁₇₀ (SEQ ID NO: 42), H₁₇₀CL₁₇₁ (SEQ ID NO: 43), L₁₇₁CA₁₇₂ (SEQ ID NO: 44), A₁₇₂CQ₁₇₃ (SEQ ID NO: 45), AND Q₁₇₃CP₁₇₄ (SEQ ID NO: 46) in an amino acid sequence having at least 90% identity to sequence set forth in SEQ ID NO: 2.

In an embodiment of the present disclosure, the polypeptide further comprises short linkers sequence at N- and /or C-terminal. Examples, of the short linked sequences include but are not limited to GC, GGSC, GGGGSC and SGGSGGC at C-terminal. Wherein the linker sequence consist from the group consist of (G)_nC (SEQ ID NO: 47), (GGS)_nC (SEQ ID NO: 48), (GGGGS)_nC (SEQ ID NO: 49) and (SGGSGG)_nC (SEQ ID NO: 50) at N- and /or C-terminal in an amino acid sequence having at least 90% identity to sequence set forth in SEQ ID NO: 3.

The present disclosure also relates to a nucleic acid construct encoding a polypeptide exhibiting granulocyte-colony stimulating factor activity and having sequence set forth in SEQ ID NO. 1.

The present disclosure also relates to polypeptide conjugates comprises the said polypeptide covalently attached to at least one molecule of polyethylene glycol ('PEG').

The present disclosure further relates to an expression vector comprising the said nucleic acid construct.

The present disclosure also relates to a host cell comprising the said expression vector.

The present disclosure also relates to a method of treating a pateint suffering from neutropenia. The method comprises administering to the pateint a therapeutic amount of the said polypeptide or the said conjugate.

The present disclosure also relates to use of the said polypeptide or the said conjugate for the preparation of a pharmaceutical compostion for treating neutropenia.

The present disclosure also relates to a pharmaceutical composition comprising the said polypeptide or conjugate and at least one pharmaceutically acceptable carrier or excipient.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1. Depicts the purified rHuG-CSF and PEGylated rHuG-CSF. The engineered human G-CSF construct was overexpressed in *E. coli* cells. The purified rHuG-CSF protein was conjugated with the 20 kDa PEG-aldehyde derivative. The conjugated product was then purified using cation exchange chromatography followed by hydrophobic interaction chromatography (HIC) or Size-exclusion chromatography. The purified rHuG-CSF (indicated by thin arrow) and PEG-conjugated rHuG-CSF (indicated by thick arrow) were analysed on non-denaturing PAGE and stained with Coomassie brilliant blue. The purified G-CSF and N-terminal 20 kDa PEGylated rHuG-CSF are found to be more than 95% pure, when analysed by SDS-PAGE.

Fig. 2. Depicts the representative profile of PEGylation of G-CSF variants. The rHuG-CSF variant was overexpressed in *E. coli* cells and purified using methods described earlier. The purified protein was conjugated with the 20 kDa PEG using PEG-maleimide derivative. The conjugated product was then purified using cation exchange chromatography (a) followed by HIC or size-exclusion chromatography (b). The PEG-conjugated G-CSF was then analysed

using MALDI-TOF mass spectrometry (c). It was observed that the majority of G-CSF was conjugated to the 20 kDa PEG and had molecular weight of approximately 40 kDa. (d) The purified G-CSF PEGylated with 20 kDa was then analysed on non-denaturing PAGE and stained with barium iodide that specifically stains PEG (panel I). Native G-CSF without PEG conjugation was not detected while the PEG conjugated protein was clearly visible. The samples were also analysed with Coomassie brilliant blue staining (panel II). Native G-CSF was only detected in Coomassie staining. An increase in the molecular weight was clearly visible upon conjugation with 20 kDa PEG. It was observed that the purified G-CSF conjugated with the 20 kDa run at the expected size of 40 kDa.

Fig. 3. Depicts representative profile of conjugation of G-CSF variant with 20 kDa, 30 kDa and 40 kDa PEG. As described above, the novel G-CSF variant was conjugated with 20 kDa, 30 kDa and 40 kDa PEG using PEG-maleimide derivatives. The PEG conjugated variants were purified using the cation exchange chromatography followed by HIC or size-exclusion chromatography purification steps. The purified conjugated products along with unmodified control were then run on 12% non-reducing PAGE and stained with barium iodide staining for visualizing PEG (a) or with Coomassie brilliant blue staining for protein (b). Corresponding increase in the molecular weight was observed upon conjugation with 20 kDa, 30 kDa and 40 kDa PEG and non-PEGylated control was detected only in Coomassie stained gel.

Fig. 4. Depicts comparison between *in vivo* activity of standard and G-CSF variant conjugated with 20 kDa, 30 kDa and 40 kDa methoxy PEG maleimide.

DETAILED DESCRIPTION

The present disclosure relates to polypeptides exhibiting G-CSF activity, conjugates of the said polypeptides and nucleic acid sequences encoding the said polypeptides. The present disclosure further relates to pharmaceutical compositions comprising said polypeptides and conjugates, methods of treatment using the said polypeptides, conjugates and compositions and use of said polypeptides in manufacture of preparation of a pharmaceutical composition for treating neutropenia.

In the present disclosure, amino acid names are used as defined by the Protein Data Bank (PDB) (www.pdb.org), which is based on the IUPAC nomenclature (IUPAC Nomenclature

and Symbolism for Amino Acids and Peptides (residue names, atom names etc.), Eur. J. Biochem., 138, 9-37 (1984) together with their corrections in Eur. J. Biochem., 152, 1 (1985).

Thus, the following symbols have been used for the amino acids.

Amino Acid	Symbol
Alanine	Ala or A
Cysteine	Cys or C
Aspartic acid	Asp or D
Glutamic acid	Glu or E
Phenylalanine	Phe or F
Glycine	Gly or G
Histidine	His or H
Isoleucine	Ile or I
Lysine	Lys or K
Leucine	Leu or L
Methionine	Met or M
Asparagine	Asn or N
Proline	Pro or P
Glutamine	Gln or Q
Arginine	Arg or R
Serine	Ser or S
Threonine	Thr or T
Valine	Val or V
Tryptophan	Trp or W
Tyrosine	Tyr or Y

- 5 The terminology used for identifying amino acid positions/substitutions is illustrated as follows

P5 indicates that position number 5 on the amino acid sequence of the disclosed polypeptide is occupied by proline.

- 10 P₅CA₆ indicates that a non-native cysteine residue is inserted between proline at position number 5 and alanine at position number 6 of the amino acid sequence with SEQ IN NO: 3.

The term cysteine derivative, cysteine variant and or G-CSF variant is used for the polypeptide with cysteine 17 replaced with serine or alanine and consist of cysteine substitution and or addition at selected site/s.

5 The term “exhibiting G-CSF activity” is intended to indicate that the polypeptide or conjugate has one or more of the functions of native G-CSF, in particular rHuG-CSF with the amino acid sequence shown in SEQ ID NO: 2 including the capability to bind to a G-CSF receptor (Fukunaga et al., J. Bio. Chem, 265:14008, 1990).

10 The present disclosure relates to a polypeptide exhibiting G-CSF activity. The polypeptide comprises at least one non-native cysteine residue at a site selected from the group consisting of T₁CP₂ (SEQ ID NO: 25), P₂CL₃ (SEQ ID NO: 26), L₃CG₄ (SEQ ID NO: 27), G₄CP₅ (SEQ ID NO: 28), P₅CA₆ (SEQ ID NO: 29), A₆CS₇ (SEQ ID NO: 30), S₉₆CP₉₇ (SEQ ID NO: 31), P₉₇CE₉₈ (SEQ ID NO: 32), L₉₉CG₁₀₀ (SEQ ID NO: 33), P₁₀₁CT₁₀₂ (SEQ ID NO: 34), E₁₂₂CE₁₂₃ (SEQ ID NO: 35), L₁₂₄CG₁₂₅ (SEQ ID NO: 36), M₁₂₆CA₁₂₇ (SEQ ID NO: 37), P₁₃₈CA₁₃₉ (SEQ ID NO: 39), A₁₄₃CF₁₄₄ (SEQ ID NO: 40), R₁₄₆CR₁₄₇ (SEQ ID NO: 41),
 15 R₁₆₉CH₁₇₀ (SEQ ID NO: 42), H₁₇₀CL₁₇₁ (SEQ ID NO: 43), L₁₇₁CA₁₇₂ (SEQ ID NO: 44), A₁₇₂CQ₁₇₃ (SEQ ID NO: 45), and Q₁₇₃CP₁₇₄ (SEQ ID NO: 46) in an amino acid sequence having at least 90% identity to sequence set forth in SEQ ID NO: 3.

In an embodiment of the present disclosure, more than one non-native cysteine residue is inserted at two or more aforementioned sites. The particular number of cysteine residues to
 20 be inserted depends upon the desired nature and degree of conjugation.

In an embodiment, the amino acid sequence comprises substitution of cysteine at C17 with serine residue.

25 In another embodiment, the polypeptide further comprises short linkers sequence at N- and /or C-terminal. Examples, of the short linker sequences include but are not limited to GC, GGSC, GGGGSC and SGGSGGC. The significance of these linker sequence is covered in subsequent example.

30 The present disclosure also relates to a polypeptide conjugate. The polypeptide conjugate comprises the said polypeptide covalently attached to at least one molecule of PEG.

The insertion of cysteine residue in the amino acid sequence makes the latter more susceptible to conjugation. Also, it allows optimization of the conjugation pattern. In an

embodiment the at least one molecule of PEG is a methoxy PEG maleimide derivative. The size of the PEG molecule is in the range of 5,000 to 40,000 daltons. In accordance with specific embodiments the size is selected from 20,000, 30,000 and 40,000 daltons.

The present disclosure also relates to a nucleic acid sequence encoding the disclosed peptide.

5 The nucleic acid construct comprises a sequence set forth in sequence with SEQ ID NO: 1

The present disclosure also relates to an expression vector comprising the nucleic acid construct. The expression vector comprises other elements necessary for the expression of the nucleic acid in a host cell for example having strong promoter such as T7 RNA polymerase.

10 The present disclosure also relates to a host cell expressing the disclosed polypeptide. The host cell is obtained by transforming a suitable cell with the disclosed expression vector. Any suitable method of transformation of host cell may be used. Examples, of suitable host cells include prokaryotic host cells such as *E. coli* but are not limited to it.

15 In other aspect, the present disclosure relates to a pharmaceutical composition comprising the polypeptide or the polypeptide conjugate and at least one pharmaceutically acceptable carrier or excipient. The pharmaceutical composition may be formulated in a variety of forms. Examples of such forms include a liquid or gel, or lyophilized, or any other suitable form. The polypeptide or the polypeptide conjugate can be formulated into pharmaceutical compositions in a manner known per se in the art to result in a polypeptide pharmaceutical that is sufficiently storage-stable and is suitable for administration to humans or animals.

20 In yet another aspect, the present disclosure relates to method for treating various forms of leucopenia or neutropenia using the disclosed polypeptide, conjugate or composition. In particular, the disclosed polypeptide, conjugate or composition may be used to prevent infection in cancer patients undergoing certain types of radiation therapy, chemotherapy bone marrow transplantations and in liver regeneration.

25 In another aspect, the present disclosure relates to use of the disclosed polypeptide, conjugate or the polypeptide for the preparation of a pharmaceutical composition for treating various forms of neutropenia or leucopenia.

The invention is further described in the non-limiting examples below.

EXAMPLES

Example 1. The nucleic acid construct synthesis, cloning and expression

The cDNA sequence of human G-CSF was codon optimized for expression in *E. coli* (Genescript). This sequence was cloned into the BamHI and HindIII sites of the expression vector pET23a (Novagen) using forward primer IMT1: 5' GGATCCATGACGCCGCTGGGTCCG 3' with SEQ ID NO: 51 and reverse primer IMT2: 5' AAGCTTTTACGGCTGTGCCAGGTGAC 3' with SEQ ID NO: 52. This construct was used to transform BL21 (DE3) *E. coli* strain from Novagen. In order to increase the yield of the codon optimized rHuG-CSF sequence, *in silico* analysis of DNA and RNA sequence of codon-optimized construct was performed. This analysis suggested formation of hairpins and highly stable secondary structure at 5' prime end of the mRNA transcript, raising possibility that the mRNA transcripts might be hindering the translation. Translationally silent mutagenesis of the gene at 5' prime end was performed to disrupt and/or reduce the mRNA secondary structures by replacement of GC rich codons (that are more likely to promote secondary structure in the mRNA transcript) with AT rich codons at suitable positions. Through a series of translationally silent mutagenesis, several sequences were created and analysed for increase in the protein yield compared to that of the native sequence.

IMT3: 5' ATG ACT CCA CTG GGT CCG GCG 3'

IMT4: 5' ATG ACT CCA TTA GGT CCG GCG AGT 3'

IMT5: 5' ATG ACT CCA TTA GGT CCA GCG AGT 3'

IMT6: 5' ATG ACT CCA TTA GGT CCG GCA AGT AGC CTG 3'

IMT7: 5' ATG ACT CCA TTA GGT CCA GCA AGT AGC CTG 3'

IMT8: 5' ATG ACT CCA TTA GGT CCA GCA TCT AGC CTG CCG CAA 3'

IMT9: 5' ATG ACT CCG CTG GGT CCG GCA TCT AGC CTG CCG CAA 3'

IMT10: 5' ATG ACT CCG TTA GGT CCG GCA TCT AGC CTG CCG CAA 3'

IMT8 (with SEQ ID NO: 58) resulted in maximal destabilisation of the secondary structure in the mRNA and resulted in significantly higher protein expression. This sequence was

cloned into the NdeI and HindIII sites of the expression vector pET23a and pET9b (Novagen). This engineered construct has resulted in 2.5 fold increase in rHuG-CSF protein yield.

Human sequence 5' **ATG ACC CCC CTG GGC CCT GCC AGC TCC CTG**

5 *E. Coli* Codon optimized 5' **ATG ACG CCG CTG GGT CCG GCG AGT AGC CTG**
(GC Content-74%)

Silent mutations (IMT8) 5' **ATG ACT CCA TTA GGT CCA GCA TCT AGC CTG**
(GC Content-48%)

Amino acids M T P L G P A S S L

10 **Example 2. Over-expression and purification of G-CSF and its variants**

The codon optimized cDNA sequence of human G-CSF with incorporated IMT8 sequence (SEQ ID NO: 58) at the 5' prime end was cloned into the NdeI and HindIII sites of the expression vector pET23a & pET9b. This engineered construct (SEQ ID NO: 1) was then used to transform BL21 (DE3) strain of *E. coli*.

15 This construct was utilized in subsequent G-CSF variants engineering. G-CSF variants were created by using standard protocols of site directed mutagenesis or by PCR using primers with desired changes for introduction or substitution of cysteine at a particular position in the coding sequence. The purified rHuG-CSF expressed in *E. coli* was analysed using the reducing and non-reducing PAGE. This analysis revealed that the purified G-CSF exhibits a
20 single protein band at the right size, and is comparable to the commercial product (Figure 1). Broadly, the rHuG-CSF protein isolation and purification involved following steps.

- i. Overnight cell culture was used to inoculate 1 litre media and once the OD₆₀₀ reached ~0.4- 0.6, the cell culture was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG).
- 25 ii. After 4 hrs, the cells were harvested by centrifugation. The cells were lysed using sonication and the inclusion bodies were isolated.
- iii. The inclusion bodies from step ii were washed thrice using following buffer components (50 mM Tris-Cl, pH 8.0, 1 mM EDTA buffer with 2% Triton X-100, 1% Na-deoxycholate or 1 M NaCl).

- iv. Washed inclusion bodies from step iii were solubilised using urea, wherein the preferred concentration is in the range of about 2 M to about 4 M urea.
- v. The isolated G-CSF was subjected to two-step refolding process.
- vi. The refolded G-CSF was purified using cation exchange chromatography.
- 5 vii. The cation exchange purified sample was subjected to hydrophobic interaction chromatography (HIC) and/or size exclusion chromatography (SEC). The resins and columns were purchased from GE Healthcare Life Sciences.
- viii. In case of cysteine variants of G-CSF, the purified G-CSF variant from step vi was used to conjugate PEG of varying size.
- 10 ix. The 5 fold molar excess of thiol specific PEG was used to conjugate rHuG-CSF variant in sodium phosphate or Tris buffer, preferred pH range of the buffer was about 6.5 to about 7.5.
- x. The N-terminal amino acid specific PEG was conjugated to rHuG-CSF in sodium phosphate, sodium acetate or sodium citrate buffer wherein the preferred buffer concentration was in the range of 25 mM to about 100 mM. The reaction is performed in the presence of reducing agent, sodium cyanoborohydrate (wherein the preferred concentration range is about 15 mM to about 25 mM). The preferred pH for the reaction was in the range of 4.0-5.5. The PEG's for both the chemistry were from JenKem Technology, USA.
- 15 xi. The PEGylated species were purified by cation exchange chromatography using SP Sepharose or more preferably MacroCap SP resins (from GE Healthcare Life Sciences), followed by HIC and/or SEC.
- xii. The samples were run on SDS PAGE and stained with barium iodide for visualizing the PEGylated G-CSF (Figure 2).

25 **Example 2. Selection of G-CSF protein regions for generating cysteine variants**

G-CSF is the primary growth factor involved in the proliferation, maturation, and differentiation of the neutrophilic-precursor cells to effector neutrophils. Extensive structural and functional studies over the years have gathered vast information about the regions of G-CSF, that play a critical role in its binding with the G-CSF receptor (G-CSFR) to initiate the signal transduction cascade that play an important role in the neutrophil proliferation. The structure of G-CSF complexed with the ligand-binding region of the G-CSF receptor in a 2:2 conformation has been solved [Aritomi, M., et al., Nature, 1999.

401(6754): p. 713-7 and Tamada, T., et al., Proceedings of the National Academy of Sciences of the United States of America, 2006. 103(9): p. 3135-40]

The solution structure of the G-CSF has also been solved using NMR spectroscopy [Zink, T., et al., Biochemistry, 1994. 33(28): p. 8453-63]. The G-CSF possesses four alpha-helical bundle structure, and these helices are labelled as A, B, C & D starting from N-terminal. There are three primary sites on the G-CSF that interact with G-CSFR protein. Another important feature of the G-CSF is the presence of five cysteine residues; four of those are involved in disulphide bonds. G-CSF has one free cysteine at position 17 and has intramolecular disulphide bonds at position 36-42 and 64-74. These disulphide bonds are necessary for biological activity of G-CSF. Whereas, the substitution of cysteine at position 17 with serine yield a mutant G-CSF protein that is fully functional [U.S. 4810643]. In the current invention, all the cysteine substitution variants have been derived from the G-CSF variant in which cysteine 17 has been changed to serine or alanine. The recombinant human G-CSF protein sequence has been assigned SEQ ID NO: 2. The cysteine 17 replaced to serine 17 variant protein sequence is assigned SEQ ID NO: 3. For all subsequent cysteine variant generation, G-CSF template with SEQ ID NO: 3 was used.

Cysteine mutations are utilized for PEGylation to increase the *in vivo* half-life of the therapeutic proteins [28]. Currently used PEGylated-G-CSF is conjugated to a 20 kDa PEG molecule at the N-terminal using reductive alkylation. However, covalent PEG modification can also be performed at the rationally selected residues of the G-CSF to further improve the half-life of G-CSF. To select the specific residues in G-CSF, or the functionally irrelevant regions of G-CSF for cysteine substitution, computational biology approach was utilized for detecting the surface accessible amino acids. The existing structural information from G-CSF structural studies was also employed to locate the regions suitable for cysteine substitution. The preferred sites for PEGylation in region proximal to Helix A are- T1, P2, L3, G4, P5, A6, S7 and S8; in Helix A are R22, E33, K34; in AB loop K40, L61; in Helix B Q90; in BC loop P97, E98, L99; in Helix C P101, Q119, E122 and E123; CD loop, P128 and P138; in Helix D R146, R147, R169, H170, L171 and A172; and in region distal to Helix D Q173 and P174. These preferred sites for cysteine substitution of native amino acid are given in Table 1.

This example provide most preferred sites for cysteine substitution- T1C (SEQ ID NO: 4), L3C (SEQ ID NO: 5), G4C (SEQ ID NO: 6), P5C (SEQ ID NO: 7), Q90C (SEQ ID NO: 8), P97C (SEQ ID NO: 9), E98C (SEQ ID NO: 10), P101C (SEQ ID NO: 11), Q119C (SEQ ID

NO: 12), E122C (SEQ ID NO: 13), E123C (SEQ ID NO: 14), P128C (SEQ ID NO: 15), P138C (SEQ ID NO: 16), R146C (SEQ ID NO: 17), R147C (SEQ ID NO: 18), R169C (SEQ ID NO: 19), H170C (SEQ ID NO: 20), L171C (SEQ ID NO: 21), A172C (SEQ ID NO: 22) Q173C (SEQ ID NO: 23), P141C (SEQ ID NO: 24) .

- 5 In one aspect, provided is a method to further confirm the solvent accessibility of the G-CSF in solution, wherein, protease degradation mapping was performed. Herein, G-CSF was subjected to protease digestion with several proteases such as trypsin, chymotrypsin and elastase using both *in silico* and *in vitro* analyses. The N- or C- terminal of the digested fragments were sequenced to identify the most prominent site/s of protease digestion. The
- 10 protein degradation revealed the surface exposed regions which could be more accessible for PEGylation. Several residues in these region were utilized in this example, wherein cysteine substitution and addition was selected for generation of G-CSF variants for improved PEGylation efficiency. The structural integrity of G-CSF variants where analysed by computational biology approach. The secondary structure of the G-CSF variants was
- 15 analysed using Circular Dichroism (CD) spectroscopy. The variants having similar structures to the wild type G-CSF protein and wherein their structural integrity is maintained could be used for PEG conjugation. The cysteine substitution and insertion variants in close proximity or at the potential protease sites used for site specific PEGylation could impart protease resistance and prolong the *in vivo* circulation half-life.
- 20 In another aspect of this invention, provided is a method wherein instead of cysteine substitution, cysteine addition was preferred. Most preferably in unstructured loop regions which is not involved in G-CSF receptor binding and thus will not impede the biological activity. The predictions made using computational biology was combined with the structural data of G-CSF and residues for cysteine insertion mutagenesis were selected. Values of
- 25 absolute surface accessibility were considered for selecting the specific residues for cysteine substitution and as well as addition. Table 2, provides the details of the most preferable positions for cysteine additions to create G-CSF variant are listed. This was ensured that the structural integrity of G-CSF is not compromised due to substitution or addition of cysteine. The solvent accessibility would increase the efficiency of the PEG conjugation of G-CSF,
- 30 besides shielding the protein from the protease to increase the *in vivo* half-life. The most preferable sites for cysteine addition are- T₁CP₂ (SEQ ID NO: 25), P₂CL₃ (SEQ ID NO: 26), L₃CG₄ (SEQ ID NO: 27), G₄CP₅ (SEQ ID NO: 28), P₅CA₆ (SEQ ID NO: 29), A₆CS₇ (SEQ ID

NO: 30), S₉₆CP₉₇ (SEQ ID NO: 31), P₉₇CE₉₈ (SEQ ID NO: 32), L₉₉CG₁₀₀ (SEQ ID NO: 33), P₁₀₁CT₁₀₂ (SEQ ID NO: 34), E₁₂₂CE₁₂₃ (SEQ ID NO: 35), L₁₂₄CG₁₂₅ (SEQ ID NO: 36), M₁₂₆CA₁₂₇ (SEQ ID NO: 37), P₁₃₈CA₁₃₉ (SEQ ID NO: 39), A₁₄₃CF₁₄₄ (SEQ ID NO: 40), R₁₄₆CR₁₄₇ (SEQ ID NO: 41), R₁₆₉CH₁₇₀ (SEQ ID NO: 42), H₁₇₀CL₁₇₁ (SEQ ID NO: 43),
 5 L₁₇₁CA₁₇₂ (SEQ ID NO: 44), A₁₇₂CQ₁₇₃ (SEQ ID NO: 45), and Q₁₇₃CP₁₇₄ (SEQ ID NO: 46).

Example 3. Modification of N- and /or C-terminal of G-CSF to create variants for efficient PEGylation

G-CSF has four helix connected with loops and also N- and C-terminal regions have unstructured regions. Importantly, computational analysis has suggested that these N- and C-terminal of G-CSF are solvent accessible. Addition of flexible amino acid linker sequence at
 10 the C terminal of the protein could also increase the flexibility and solubility of the region. The flexible linkers are generally rich in small or polar amino acids such as glycine and serine but can also consists of amino acids such as threonine, alanine, lysine and glutamic acid.

Using this information, in this example, short flexible linker sequences containing cysteine was added at N- and/or C- terminal of the G-CSF to further increase the flexibility and solvent accessibility of cysteine added for PEG conjugation. These cysteine variants were further modified by conjugating cysteine reactive methoxy PEG maleimide. The addition of linker sequences containing cysteine would not alter the overall conformation of molecule
 20 and thus would not reduce the activity of the therapeutic protein. However, due to better solvent accessibility, these variants could possess higher PEG conjugation efficiency. Furthermore, in silico analysis also indicated that the cysteine with smaller amino acids such as glycine or a serine linker could enhance the solvent accessibility. The most preferred cysteine containing linker sequences are- (G)_nC (SEQ ID NO: 47), (GGS)_nC (SEQ ID NO:
 25 48), (GGGGS)_nC (SEQ ID NO: 49) and (SGGSGG)_nC (SEQ ID NO: 50) (as given in Table 3). In these sequences linker length is n = 1 to 4. These variants will impart improved solvent accessibility of cysteine for PEGylation. The cysteine variants generated have been assigned unique SEQ IDs. Table 3, provides the details of the most preferable positions for N and C terminal cysteine additions to create G-CSF variant.

Moreover, the solved structure of the G-CSF suggests that the N- and C-terminal unstructured loops are in close proximity. Two cysteine residues at both the two terminals could result in formation of disulphide bond under appropriate conditions. Such a disulphide bond formation

will result in a circularized variants of the G-CSF. Earlier published literature suggests that circularization of protein enhances the thermal stability, protease resistance and in vivo half-life.

Example 4. Conjugation of PEG to rHuG-CSF to increase half-life

5 PEGylation is one of the important methods used to create modified variants of the therapeutic proteins for improving their overall half-life in vivo. Currently used PEGylated G-CSF is conjugated to a 20 kDa PEG molecule at the N-terminal using reductive alkylation. Cysteine mutations have been demonstrated to facilitate PEGylation of the therapeutic proteins. In this invention several cysteine variants of G-CSF were provided to facilitate site
10 specific PEGylation (Table 4). These G-CSF variants are expected to have comparable biological activity to the commercially available G-CSF since the core of the protein involved in interaction with the G-CSF receptor has not been changed. However, due to their design and PEGylation, they would possess the advantage of improved bioavailability due to slow degradation. Such PEGylated variants of G-CSF are expected to possess longer serum half-
15 life.

Furthermore, different sizes of PEG ranging from 5000 daltons-40,000 daltons could be conjugated to these variants to increase their half-life and bioavailability. To confirm efficient PEGylation of newly designed variants MALDI-TOF analysis could be performed. Figures 2 and 3 show the purification strategy evolved for PEGylated G-CSF variant, as is
20 seen by the resolution of purified PEGylated product in SDS-PAGE stained with Coomassie brilliant blue for visualization of proteins and barium iodide for visualization of the PEGylated variants. The higher molecular weight of the PEGylated variant was also confirmed with Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) Mass Spectrometry. To access the secondary structure of the variants, CD spectroscopy was
25 performed.

Example 5. Analysis of bioactivity of G-CSF and PEGylated variants by cell-based assay

To analyse the biological activity of rHuG-CSF and its variants, cell proliferation assays were performed. The biological activity of rHuG-CSF was determined by its ability to proliferate
30 murine myeloblastic NFS-60 cells. In these assays the metabolic activity of proliferating cells is measured through reduction of tetrazolium reagent such as XTT. The cells were treated with various concentrations of standard, rHuG-CSF and its PEGylated variants for 48 hr and

their metabolic activity was then assessed using the XTT reagent. The biological activity of commercially available filgrastim and the lab produced rHuG-CSF and its variant/s was found to be comparable. The result from this could demonstrate that the mutations in the G-CSF has not resulted in the compromise of the biological activity.

5 **Example 6. *In vivo* biological activity and half-life of the novel G-CSF variants**

Male BALB/c mice, 12-14 weeks old, were used in the current study for analysing the biological activity and *in vivo* half-life of the novel G-CSF variants engineered in this study. Towards this, mice were acclimatized for a week, and neutropenia was induced in mice using intra-peritoneal injection of cyclophosphamide (200 mg/kg) as per standard procedures. To
10 confirm induction of neutropenia, blood was withdrawn and total leucocytes counts (TLC) were measured. One-day, post induction of neutropenia, “Sham” or Mock (having buffer only), therapeutically active G-CSF (commercially available), and the engineered variants were independently administered as the single subcutaneous dose (up to 1 mg/Kg) or multiple dosages of 125 µg/Kg for the span of 4-7 days. After G-CSF treatment, the blood samples
15 were withdrawn, and TLC counts are determined for following 5-10 days. The lab-made native-like, and commercial filgrastim exhibited comparable specific activity. Interestingly, we observed that the treatment of neutropenia with the PEGylated G-CSF variants (i.e. newly constructed PEGylated cysteine variants) resulted in accelerated recovery from neutropenia. Similarly, for estimating the *in vivo* half-life, independent groups of mice were injected with
20 either the Sham sample, commercially available G-CSF and novel G-CSF variants. Blood samples were withdrawn, and the presence of G-CSF in serum was estimated using the commercially available G-CSF Elisa kit. These data clearly showed that the variant possess higher biological activity and longer *in vivo* half-lives compared to original unmodified G-CSF (Figure 4).

25 **ADVANTAGES OF THE INVENTION**

- The disclosed polypeptides are more susceptible to conjugation. Also, the disclosed polypeptides allow optimization of the conjugation pattern. The disclosed conjugates exhibit significantly longer serum half-life, a therapeutically advantageous property.
30 The disclosed conjugates, therefore, can be used on the same day of administration of chemotherapy. Site specific PEGylation overcomes the problem of product heterogeneity and loss of biological activity that arises due to the conventionally used amine-PEGylation.

- Site specific PEGylation allows the protein to be selectively conjugated with PEG at a unique single, double or more, predetermined, site(s). Such sites make the characterization of the polypeptide more robust and the biological activity can be relatively accurately predicted and made commensurate with different regimens of treatments. Further, the site of PEGylation is at residues that are solvent accessible making PEGylation highly efficient and the product more homogenous.
- The present invention provides cysteine variants of G-CSF and their application to generate PEG conjugated variants, with potential for significantly longer serum half-life, a therapeutically advantageous property, and thus could be used on the same day of administration of chemotherapy.

REFERENCES:

1. Crawford, J., D.C. Dale, and G.H. Lyman, *Chemotherapy-induced neutropenia: risks, consequences, and new directions for its management*. Cancer, 2004. **100**(2): p. 228-37.
- 5 2. Lyman, G.H., *Risks and consequences of chemotherapy-induced neutropenia*. Clin Cornerstone, 2006. **8 Suppl 5**: p. S12-8.
3. Kuderer, N.M., et al., *Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients*. Cancer, 2006. **106**(10): p. 2258-66.
4. Lyman, G.H., N.M. Kuderer, and B. Djulbegovic, *Prophylactic granulocyte colony-stimulating factor in patients receiving dose-intensive cancer chemotherapy: a meta-analysis*. Am J Med, 2002. **112**(5): p. 406-11.
- 10 5. Metcalf, D., *The colony stimulating factors. Discovery, development, and clinical applications*. Cancer, 1990. **65**(10): p. 2185-95.
6. Bradley, T.R., D. Metcalf, and W. Robinson, *Stimulation by leukaemic sera of colony formation in solid agar cultures by proliferation of mouse bone marrow cells*. Nature, 1967. **213**(5079): p. 926-7.
- 15 7. Pluznik, D.H. and L. Sachs, *The induction of clones of normal mast cells by a substance from conditioned medium*. Experimental cell research, 1966. **43**(3): p. 553-63.
8. Bradley, T.R. and D. Metcalf, *The growth of mouse bone marrow cells in vitro*. The Australian journal of experimental biology and medical science, 1966. **44**(3): p. 287-99.
- 20 9. Welte, K., et al., *Filgrastim (r-metHuG-CSF): the first 10 years*. Blood, 1996. **88**(6): p. 1907-29.
10. Welte, K., et al., *Purification and biochemical characterization of human pluripotent hematopoietic colony-stimulating factor*. Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(5): p. 1526-30.
- 25 11. Souza, L.M., et al., *Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells*. Science, 1986. **232**(4746): p. 61-5.
12. Nagata, S., et al., *Molecular cloning and expression of cDNA for human granulocyte colony-stimulating factor*. Nature, 1986. **319**(6052): p. 415-8.
- 30 13. Keating, G.M., *Lenograstim: a review of its use in chemotherapy-induced neutropenia, for acceleration of neutrophil recovery following haematopoietic stem cell transplantation and in peripheral blood stem cell mobilization*. Drugs. **71**(6): p. 679-707.
14. van de Geijn, G.J., et al., *Granulocyte colony-stimulating factor and its receptor in normal hematopoietic cell development and myeloid disease*. Rev Physiol Biochem Pharmacol, 2003. **149**: p. 53-71.
- 35 15. Fernandez-Varon, E. and L. Villamayor, *Granulocyte and granulocyte macrophage colony-stimulating factors as therapy in human and veterinary medicine*. Vet J, 2007. **174**(1): p. 33-41.
16. Tanaka, H., et al., *Three types of recombinant human granulocyte colony-stimulating factor have equivalent biological activities in monkeys*. Cytokine, 1997. **9**(5): p. 360-9.
- 40 17. Kuwabara, T., S. Kobayashi, and Y. Sugiyama, *Pharmacokinetics and pharmacodynamics of a recombinant human granulocyte colony-stimulating factor*. Drug metabolism reviews, 1996. **28**(4): p. 625-58.
18. Kuwabara, T., et al., *Receptor-mediated clearance of G-CSF derivative nartograstim in bone marrow of rats*. The American journal of physiology, 1995. **269**(1 Pt 1): p. E1-9.
- 45 19. Kuwabara, T., et al., *Renal clearance of a recombinant granulocyte colony-stimulating factor, nartograstim, in rats*. Pharmaceutical research, 1995. **12**(10): p. 1466-9.

20. El Ouriaghli, F., et al., *Neutrophil elastase enzymatically antagonizes the in vitro action of G-CSF: implications for the regulation of granulopoiesis*. Blood, 2003. **101**(5): p. 1752-8.
21. Molineux, G., *The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta)*. Current pharmaceutical design, 2004. **10**(11): p. 1235-44.
- 5 22. Andersen, D.C. and C.F. Goochee, *The effect of ammonia on the O-linked glycosylation of granulocyte colony-stimulating factor produced by chinese hamster ovary cells*. Biotechnology and bioengineering, 1995. **47**(1): p. 96-105.
23. Halpern, W., et al., *Albugranin, a recombinant human granulocyte colony stimulating factor (G-CSF) genetically fused to recombinant human albumin induces prolonged myelopoietic effects in mice and monkeys*. Pharmaceutical research, 2002. **19**(11): p. 1720-9.
- 10 24. Harris, J.M. and R.B. Chess, *Effect of pegylation on pharmaceuticals*. Nature reviews. Drug discovery, 2003. **2**(3): p. 214-21.
25. Crawford, J., *Pegfilgrastim administered once per cycle reduces incidence of chemotherapy-induced neutropenia*. Drugs, 2002. **62 Suppl 1**: p. 89-98.
- 15 26. Kinstler, O.B., et al., *N-terminally chemically modified protein compositions and methods*, 1998, Google Patents.
27. Soni, B., et al., *Method for the treatment of neutropenia by administration of a multi-pegylated granulocyte colony stimulating factor (G-CSF) variant*, 2008, Google Patents.
- 20 28. Braxton, S.M., *Cysteine-pegylated proteins*, 1998, Google Patents.
29. Goodson, R.J. and N.V. Katre, *Site-directed pegylation of recombinant interleukin-2 at its glycosylation site*. Bio/technology, 1990. **8**(4): p. 343-6.
30. Xiong, C.Y., et al., *Development of tumor targeting anti-MUC-1 multimer: effects of di-scFv unpaired cysteine location on PEGylation and tumor binding*. Protein engineering, design & selection : PEDS, 2006. **19**(8): p. 359-67.
- 25 31. Ishikawa, M., et al., *The substitution of cysteine 17 of recombinant human G-CSF with alanine greatly enhanced its stability*. Cell structure and function, 1992. **17**(1): p. 61-5.
32. Kozlowski, A. and J.M. Harris, *Improvements in protein PEGylation: pegylated interferons for treatment of hepatitis C*. Journal of controlled release : official journal of the
- 30 Controlled Release Society, 2001. **72**(1-3): p. 217-24.
33. Zhai, Y., et al., *Enhanced circulation half-life of site-specific PEGylated rhG-CSF: optimization of PEG molecular weight*. Journal of biotechnology, 2009. **142**(3-4): p. 259-66.
34. Aritomi, M., et al., *Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme*. Nature, 1999. **401**(6754): p. 713-7.
- 35 35. Tamada, T., et al., *Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(9): p. 3135-40.
36. Zink, T., et al., *Structure and dynamics of the human granulocyte colony-stimulating factor determined by NMR spectroscopy. Loop mobility in a four-helix-bundle protein*.
- 40 Biochemistry, 1994. **33**(28): p. 8453-63.
37. Berna, M. and F.M. Veronese, *G-CSF conjugates with peg*, 2005, Google Patents.

PATENT REFERENCES:

	US 6,608,183	8/2003	Cox, III; George N	530/399
	US 7,309,781	12/2007	Cox, III; George N	530/399
5	US 7,253,267	8/2007	Cox, III; George N	530/399
	US 7,214,779	5/2007	Cox, III; George N	530/399
	US 7,232,885	6/2007	Cox, III; George N	530/351
	US 7,345,149	3/2008	Cox, III; George N	530/351
	US 7,314,921	1/2008	Cox, III; George N	530/399
10	US 7,629,314	12/2008	Cox, III; George N	514/1.1
	US 7,795,396	9/2010	Cox, III; George N	530/351
	US 7,732,572	6/2010	Cox, III; George N	530/351
	US 7,964,184	6/2011	Cox, III; George N	424/85.5
	US 8,133,480	3/2012	Cox, III; George N	424/85.2
15	US 7,153,943	12/2012	Cox, III, et al.	530/399
	US 7,824,669	11/2010	Cox, III, et al.	424/85.1
	US 8,148,500	4/2012	Cox, III, et al.	530/399
	US 7,947,655	5/2011	Cox, III, et al.	541/21.2
	US 7,560,101	7/2009	Cox, III, et al.	424/85.1
20	US 8,455,434	6/2013	Cox; George N.	514/7.6
	US 8,841,426	9/2014	Kang, et al.	530/399
	US 5,824,784	10/1998	Kinstler, et al.	530/399
	US 8,093,032	1/2012	Kumar, et al.	435/216
	US 20090203601	8/2009	Soni, et al.	514/1.1
25	US 20130165637	6/2013	Xiaoqiang, et al.	530/362
	US6831158	12/2004	Nissen, et al.	530/397
	US7994124B2	8/2011	Cox, et al.	514/7.6
	EP0668354A1	4/1989	Shaw et al.	C12N 15/27

We claim:

1. A polypeptide exhibiting granulocyte-colony stimulating factor activity, the polypeptide comprising an amino acid sequence having at least 90% sequence identity to sequence set forth in SEQ ID NO: 2 and having at least one non-native cysteine residue at a site selected from the group consisting of T₁CP₂ (SEQ ID NO: 25), P₂CL₃ (SEQ ID NO: 26), L₃CG₄ (SEQ ID NO: 27), G₄CP₅ (SEQ ID NO: 28), P₅CA₆ (SEQ ID NO: 29), A₆CS₇ (SEQ ID NO: 30), S₉₆CP₉₇ (SEQ ID NO: 31), P₉₇CE₉₈ (SEQ ID NO: 32), L₉₉CG₁₀₀ (SEQ ID NO: 33), P₁₀₁CT₁₀₂ (SEQ ID NO: 34), E₁₂₂CE₁₂₃ (SEQ ID NO: 35), L₁₂₄CG₁₂₅ (SEQ ID NO: 36), M₁₂₆CA₁₂₇ (SEQ ID NO: 37), P₁₃₈CA₁₃₉ (SEQ ID NO: 39), A₁₄₃CF₁₄₄ (SEQ ID NO: 40), R₁₄₆CR₁₄₇ (SEQ ID NO: 41), R₁₆₉CH₁₇₀ (SEQ ID NO: 42), H₁₇₀CL₁₇₁ (SEQ ID NO: 43), L₁₇₁CA₁₇₂ (SEQ ID NO: 44), A₁₇₂CQ₁₇₃ (SEQ ID NO: 45), and Q₁₇₃CP₁₇₄ (SEQ ID NO: 46).
2. The polypeptide as claimed in claim 1, wherein the polypeptide comprises substitution of C17 with a serine/alanine residue set forth in SEQ ID NO: 3.
3. The polypeptide as claimed in any of claims 1 or 2, wherein the polypeptide further comprises short linkers sequence selected from the group consisting of (G)_nC (SEQ ID NO: 47), (GG)_nC (SEQ ID NO: 48), (GGGG)_nC (SEQ ID NO: 49) and (SGGSGG)_nC (SEQ ID NO: 50) at N- and/or C-terminal.
4. The polypeptide as claimed in any of claims 1 to 3, wherein the polypeptide is conjugated to at least one molecule of polyethylene glycol.
5. The polypeptide as claimed in claim 4, wherein the polyethylene glycol is a methoxy PEG maleimide derivative.
6. The polypeptide as claimed in claim 4 or 5, wherein size of the polyethylene molecule ranges from 5,000 to 40,000 daltons.
7. A nucleic acid construct encoding a polypeptide exhibiting granulocyte-colony stimulating factor activity, the construct comprising a sequence set forth in a sequence selected from the group consisting of SEQ ID NO: 1.

8. An expression vector comprising the nucleic acid construct as claimed in claim 7.

9. A host cell comprising the expression vector as claimed in claim 8.

5

10. A method of treating a pateint suffering from neutropenia, the method comprising administering to the pateint a therapeutic amount of the polypeptide as claimed in any of claims 1 to 6.

10 11. Use of the polypeptide as claimed in any of claims 1 to 6 for the preparation of a pharmaceutical compostion for treating neutropenia.

12. A pharmaceutical composition comprising the polypeptide as claimed in any of claims 1 to 6 and at least one pharmaceutically acceptable carrier or excipient.

15

20

25

30

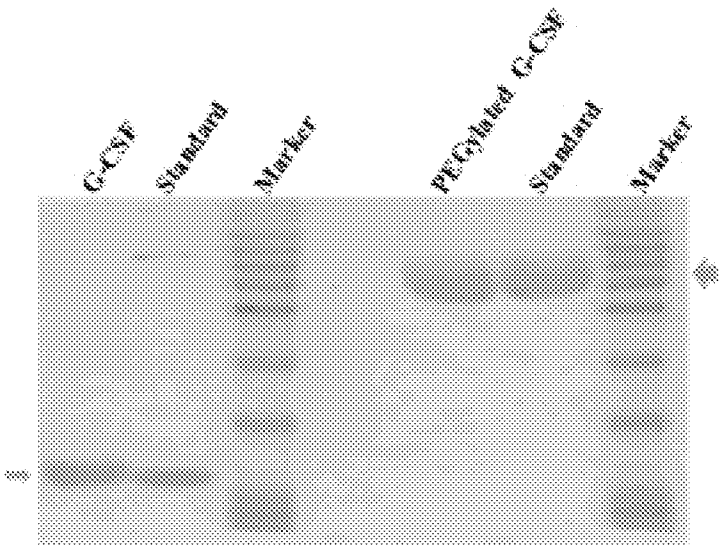


Figure 1

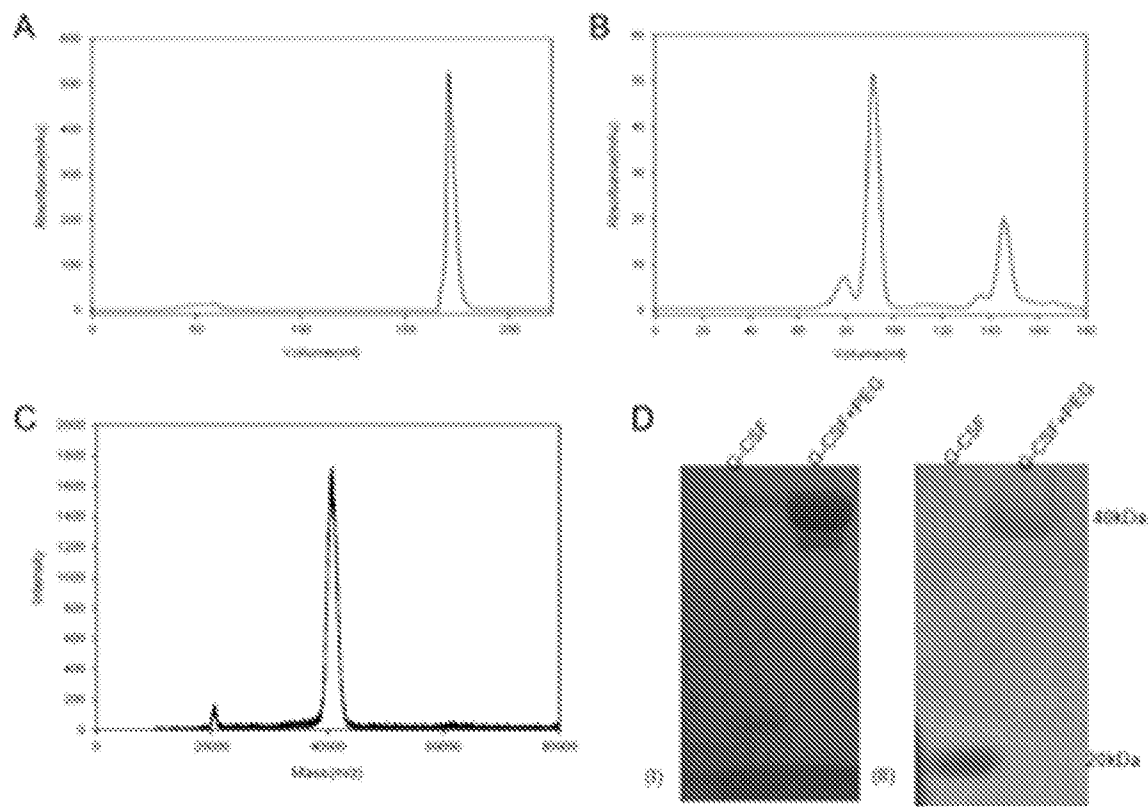


Figure 2

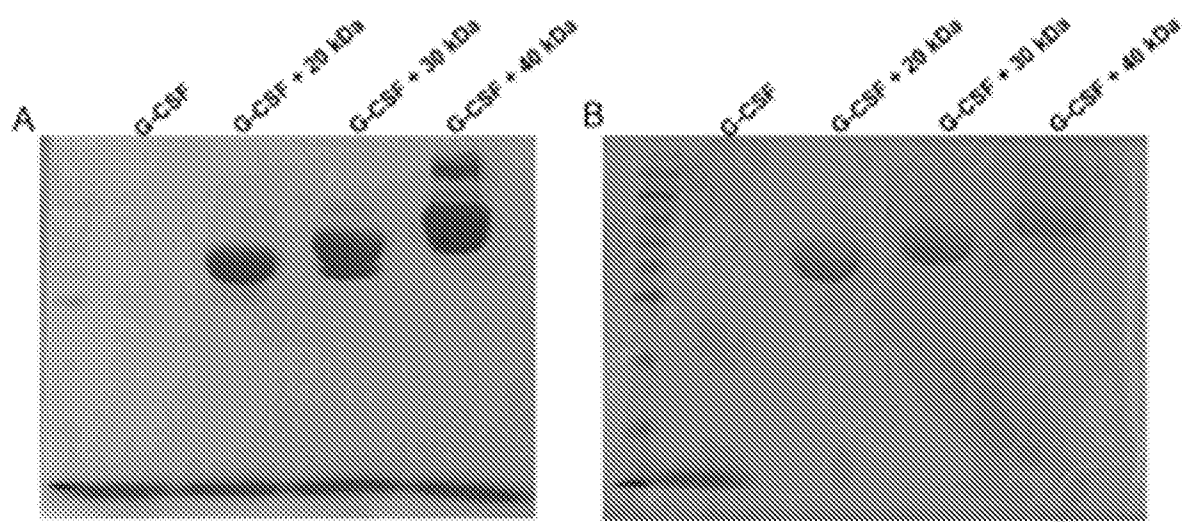


Figure 3

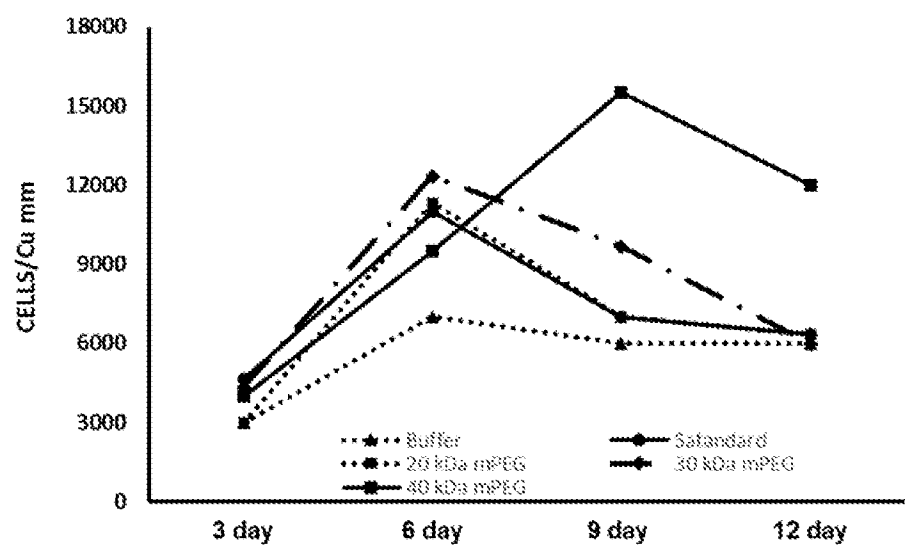


Figure 4

Table 1. Preferred position for cysteine substitution for site specific PEGylation of G-CSF

Proximal to helix A Amino acid 1-10	T1, P2 ,L3, G4, P5, A6, S7, S8
Helix A Amino acid 11-39	E33, K34
AB loop Amino acid 40-70	K40, L61
Helix B Amino acid 71-91	Q90
BC loop Amino acid 92-99	P97, E98, L99
Helix C Amino acid 100-123	P101, E122, E123
CD Loop Amino acid 124-142	P128, P138
Helix D Amino acid 143-172	R146, R147, R169, H170, L171, A172
Distal to helix D Amino-acid 173-174	Q173, P174

Table 2. Preferred positions for cysteine insertion for site specific PEGylation of G-CSF

Region	Insertion at site
Proximal to helix A Amino acid 1-10	MCT ₁ , T ₁ CP ₂ , P ₂ CL ₃ , L ₃ CG ₄ , G ₄ CP ₅ , P ₅ CA ₆ , A ₆ CS ₇
Helix A Amino acid 11-39	None
AB loop Amino acid 40-70	None
Helix B Amino acid 71-91	None
BC loop Amino acid 92-99	S ₉₆ CP ₉₇ , P ₉₇ CE ₉₈ , L ₉₉ CG ₁₀₀
Helix C Amino acid 100-123	P ₁₀₁ CT ₁₀₂ , M ₁₂₁ CE ₁₂₂ , E ₁₂₂ CE ₁₂₃
CD Loop Amino acid 124-142	L ₁₂₄ CG ₁₂₅ , M ₁₂₆ CA ₁₂₇ , Q ₁₃₄ CG ₁₃₅ , P ₁₃₈ CA ₁₃₉
Helix D Amino acid 143-172	A ₁₄₃ CF ₁₄₄ , R ₁₄₆ CR ₁₄₇ , R ₁₆₉ CH ₁₇₀ , H ₁₇₀ CL ₁₇₁ , L ₁₇₁ CA ₁₇₂ , A ₁₇₂ CQ ₁₇₃
Distal to helix D Amino-acid 173-174	Q ₁₇₃ P ₁₇₄ C, Q ₁₇₃ CP ₁₇₄

Table 3. N- & C- terminal cysteine insertion and substitution sites for PEGylation

Region	N- & C- terminal cysteine insertion and substitution sites for PEGylation
Proximal to helix A Amino acid 1-10	MCT ₁ T1, P2, L3, G4, P5, A6, S7, S8 N terminal linkers- (G)nC, (GGS)nC, (GGGGS)nC, (SGGSGG)nC
Distal to helix D 173-174	Q173C, P174C Q ₁₇₃ P ₁₇₄ C C terminal linkers- (G)nC, (GGS)nC, (GGGGS)nC, (SGGSGG)nC

Table 4. Sequence list of variants of G-CSF for cysteine substitution and insertion.

SEQUENCE ID	Modification
SEQ ID NO 1	DNA (G-CSF)
SEQ ID NO 2	Polypeptide sequence of Recombinant Human G-CSF
SEQ ID NO 3	C17S
SEQ ID NO 4	T1C
SEQ ID NO 5	L3C
SEQ ID NO 6	G4C
SEQ ID NO 7	P5C
SEQ ID NO 8	Q90C
SEQ ID NO 9	P97C
SEQ ID NO 10	E98C
SEQ ID NO 11	P101C
SEQ ID NO 12	Q119C
SEQ ID NO 13	E122C
SEQ ID NO 14	E123C
SEQ ID NO 15	P128C
SEQ ID NO 16	P138C
SEQ ID NO 17	R146C
SEQ ID NO 18	R147C
SEQ ID NO 19	R169C
SEQ ID NO 20	H170C
SEQ ID NO 21	L171C
SEQ ID NO 22	A172C
SEQ ID NO 23	Q173C
SEQ ID NO 24	P174C
SEQ ID NO 25	T1CP2
SEQ ID NO 26	P2CL3
SEQ ID NO 27	L3CG4
SEQ ID NO 28	G4CP5
SEQ ID NO 29	P5CA6
SEQ ID NO 30	A6CS7
SEQ ID NO 31	S96CP97
SEQ ID NO 32	P97CE98
SEQ ID NO 33	L99CG100
SEQ ID NO 34	P101CT102
SEQ ID NO 35	E122CE123
SEQ ID NO 36	L124CG125
SEQ ID NO 37	M126CA127
SEQ ID NO 38	C134CG135
SEQ ID NO 39	F138CA139
SEQ ID NO 40	A143CP144
SEQ ID NO 41	R145CR147

SEQ ID NO 42	R169CH170
SEQ ID NO 43	H170CL171
SEQ ID NO 44	L171CA172
SEQ ID NO 45	A172CQ173
SEQ ID NO 46	Q173CP174
SEQ ID NO 47	GC
SEQ ID NO 48	GGSC
SEQ ID NO 49	GGGGSC
SEQ ID NO 50	SGGSGGC
SEQ ID NO 51	IMT1
SEQ ID NO 52	IMT2
SEQ ID NO 53	IMT3
SEQ ID NO 54	IMT4
SEQ ID NO 55	IMT5
SEQ ID NO 56	IMT6
SEQ ID NO 57	IMT7
SEQ ID NO 58	IMT8
SEQ ID NO 59	IMT9
SEQ ID NO 60	IMT10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IN2018/050873

A. CLASSIFICATION OF SUBJECT MATTER
C07K14/535, A61K38/00 Version=2019.01

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K, A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

TotalPatent One, IPO Internal Database

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US20030153049A1, (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY), 14 AUG 2003 (14-08-2003), Whole document	7-9
X	WO2007011166A1, (MOGAM BIOTECHNOLOGY RESEARCH INSTITUTE), 25 JAN 2007 (25-01-2007), Whole document	7-9
Y	---do---	1-6, 10-12
X	WO1999003887A1, (BOLDER BIOTECHNOLOGY INC), 28 JAN 1999 (28-01-1999), Abstract, Example 5, claims 21-23	7-9
Y	---do---	1-6, 10-12
X	WO2006135176A1, (CJ CORPORATION), 21 DEC 2006 (21-12-2006), Paragraphs [021]-[024],[054]; fig.1	7-9
Y	---do---	1-6, 10-12

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

28-03-2019

Date of mailing of the international search report

28-03-2019

Name and mailing address of the ISA/

Indian Patent Office
Plot No.32, Sector 14, Dwarka, New Delhi-110075
Facsimile No.

Authorized officer

Ezhil Nilamani G

Telephone No. +91-1125300200

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IN2018/050873

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Continuation of Observations where unity of invention is lacking(Box III)

This International Searching Authority found multiple inventions in this international application, the reason and inventions are as follows

The subject matter of present claims relates to variants of Granulocyte-Colony Stimulating Factor ('G-CSF'), conjugates of the variants of G-CSF, nucleotide construct, methods for preparing such variants and conjugates and use of such variants or conjugates in therapy particularly neutropenia.

The common technical feature linking the claims is the polypeptides exhibiting GCSF activity having at least 90% identity to SEQ ID NO: 2 and at least one non native cysteine insertion. However these features are known from D2 which discloses the mutants of GCSF which has non native cysteine insertion and has at least 90% identity to SEQ ID NO: 2. Hence the technical feature linking the claims are known from the prior art document. Since no other feature could be identified which might be considered to represent a special technical feature in the sense of Rule 13.1 PCT to link the subject-matters of the different inventions, the present application does not meet the requirements of Rule 13 PCT with respect to unity of invention. Therefore claims 1 to 12 lacks unity of invention "a posteriori". Thus the GCSF variant polypeptide of SEQ ID NO: 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46 are not unified.

The following sets of separate invention have been identified.

Invention 1: Claims 1-12 (partially) directing to a polypeptide exhibiting granulocyte-colony stimulating factor activity, the polypeptide comprising an amino acid sequence having at least 90% sequence identity to sequence set forth in SEQ ID NO: 2 and having at least one non-native cysteine residue at a site selected from the group consisting of T1CP2 (SEQ ID NO: 25), nucleic acid construct, use of the polypeptide and composition comprising the polypeptide.

Invention 2: Claims 1-12 (partially) directing to SEQ ID NO: 26

Invention 3-21: Claims 1-12 (partially) directing to SEQ ID NO: 27-37 and seq id no: 39-46. Each sequence represents a separate invention.

As all the claims could be searched without effort justifying additional fees, the present search and written opinion is with respect to the claims 1-12.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IN2018/050873

Citation	Pub.Date	Family	Pub.Date
US 20030153049 A1	14-08-2003	WO 0173081 A1	04-10-2001
		EP 1185675 A1	13-03-2002
		CN 1366551 A	28-08-2002
		KR 1020010094652 A	01-11-2001
WO 2007011166 A1	25-01-2007	CN 102153643 A	17-08-2011
		EP 2174951 A2	14-04-2010
		JP 2009501789 A	22-01-2009
		KR 1020070010817 A	24-01-2007
		US 20080200657 A1	21-08-2008
WO 9903887 A1	28-01-1999	US 20040265269 A1	30-12-2004
		AU 2008202371 A1	19-06-2008
		NZ 513077 A	26-03-2004
		JP 2009096810 A	07-05-2009
		EP 1881005 B1	03-04-2013
WO 2006135176 A1	21-12-2006	KR 100694994 B1	07-03-2007
		US 20100143288 A1	10-06-2010
		CN 101198357 A	11-06-2008
		JP 2008543304 A	04-12-2008
		IN 4991KOLNP2007 A	02-01-2009
		EP 1896081 A1	12-03-2008