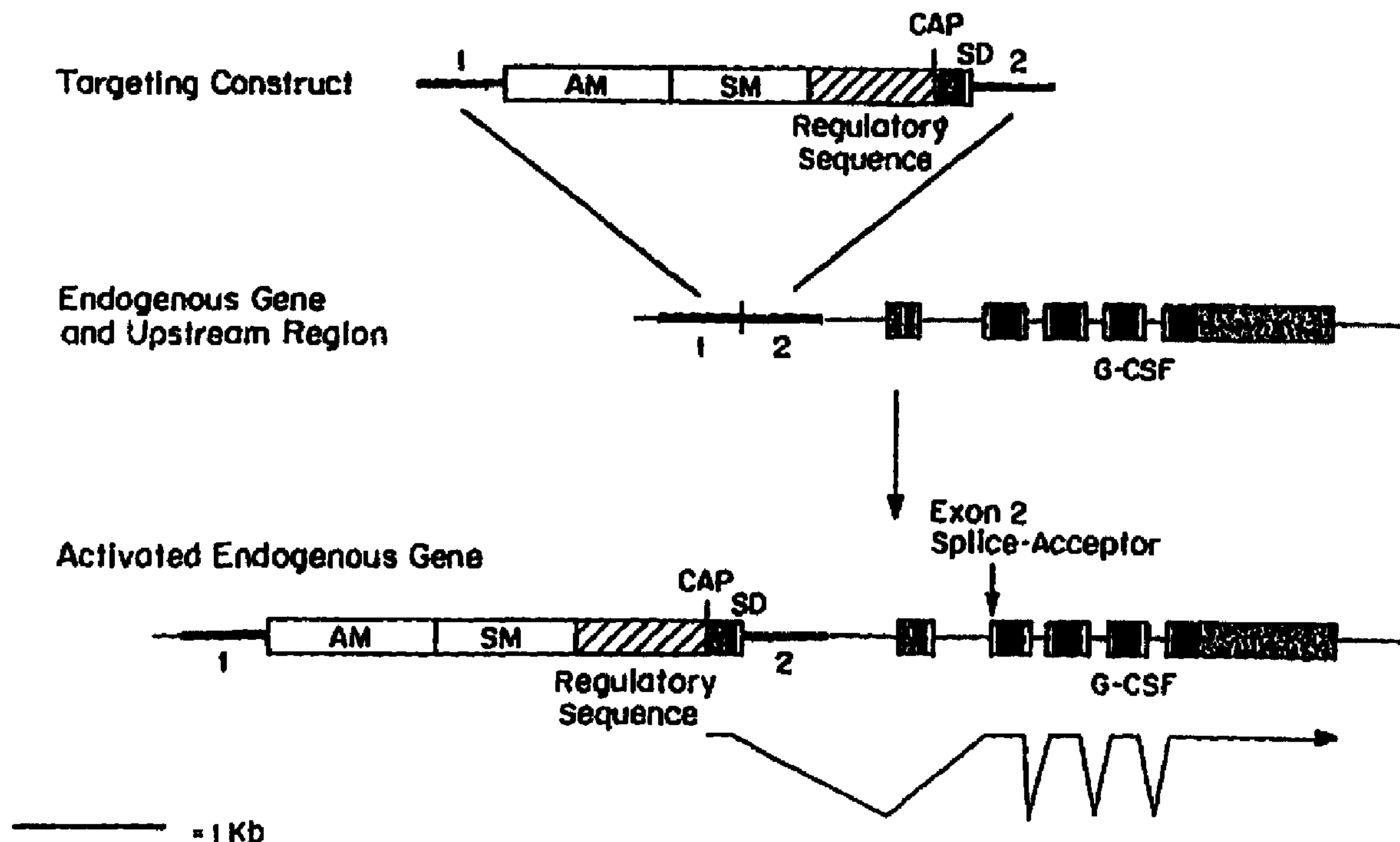




- (72) TRECO, DOUGLAS A., US
(72) HEARTLEIN, MICHAEL W., US
(72) SELDEN, RICHARD F., US
(71) TRANSKARYOTIC THERAPIES, INC., US
(51) Int.Cl.⁶ C12N 15/67, C12N 15/90, C12N 15/85, C07K 14/535,
C12N 15/27, C12N 5/10, A61K 48/00
(30) 1998/05/07 (60/084,649) US
(54) **SEQUENCES GENOMIQUES EN AMONT DE LA REGION
CODANTE DU GENE G-CSF DESTINEES A LA PRODUCTION
ET A L'ADMINISTRATION DE PROTEINES**
(54) **GENOMIC SEQUENCES UPSTREAM OF THE CODING
REGION OF THE G-CSF GENE FOR PROTEIN PRODUCTION
AND DELIVERY**



(57) L'invention concerne une molécule isolée d'acide nucléique capable de s'hybrider, dans des conditions précises, ou de présenter une identité de séquence d'au moins 80 % avec une région génomique définie en amont de la région codant du gène G-CSF. L'invention concerne également une structure d'ADN contenant cette molécule d'ADN comme séquence cible pour la recombinaison homologue.

(57) An isolated nucleic acid molecule that hybridizes under stringent conditions, or shares at least 80 % sequence identity, with a defined genomic region upstream of the coding region of the G-CSF gene, and a DNA construct containing that DNA molecule as a targeting sequence for homologous recombination.

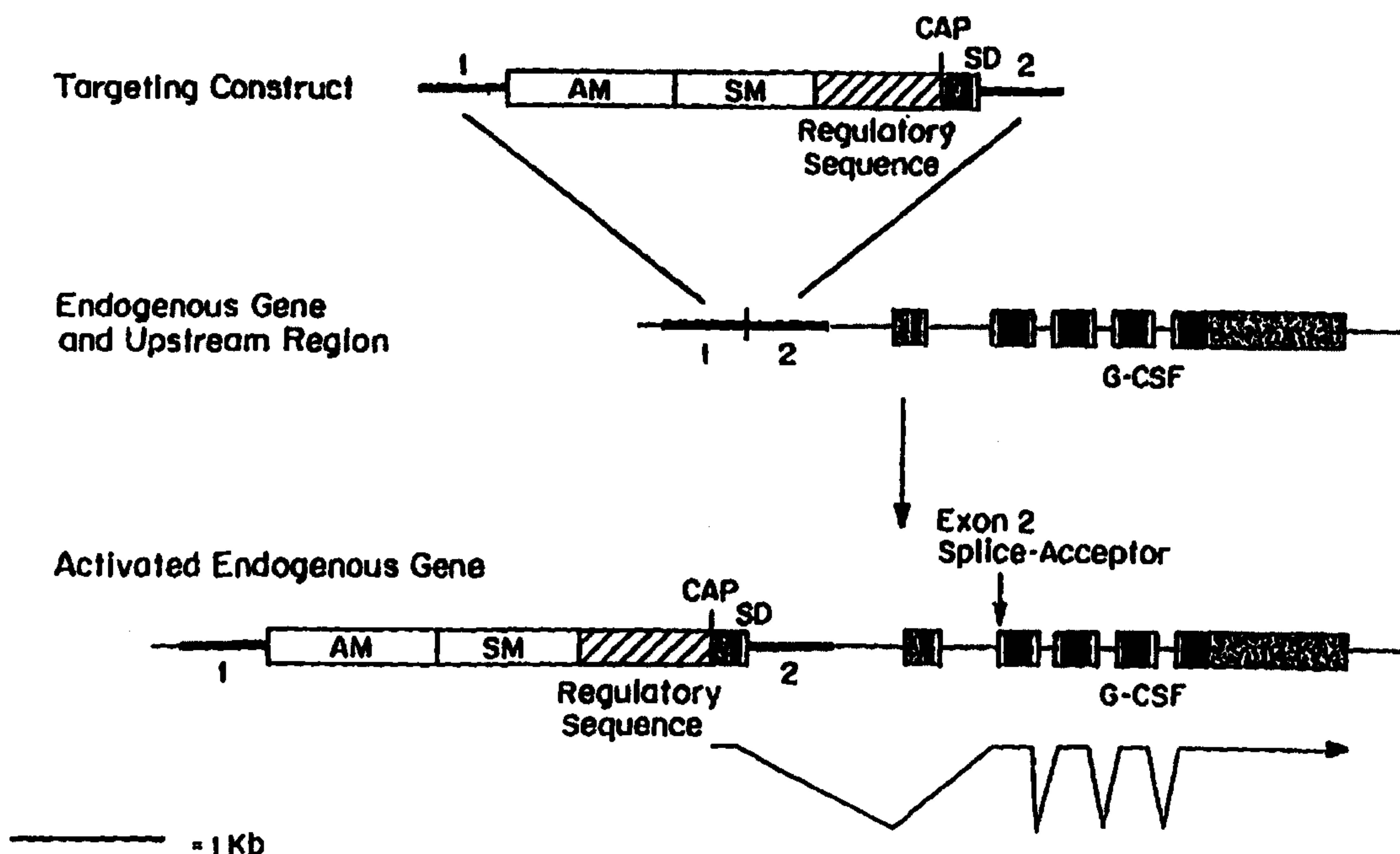


**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/67, 15/90, 15/85, 15/27, C07K 14/535, A61K 48/00, C12N 5/10	A1	(11) International Publication Number: WO 99/57291 (43) International Publication Date: 11 November 1999 (11.11.99)
(21) International Application Number: PCT/US99/09924 (22) International Filing Date: 5 May 1999 (05.05.99) (30) Priority Data: 60/084,649 7 May 1998 (07.05.98) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 60/084,649 (CON) Filed on 7 May 1998 (07.05.98) (71) Applicant (for all designated States except US): TRAN-SKARYOTIC THERAPIES, INC. [US/US]; 195 Albany Street, Cambridge, MA 02139 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): TRECO, Douglas, A. [US/US]; 87 Brantwood Road, Arlington, MA 02174 (US). HEARTLEIN, Michael, W. [US/US]; 167 Reed Farm Road, Boxborough, MA 01719 (US). SELDEN, Richard, F. [US/US]; 106 Bristol Road, Wellesley, MA 02481 (US). (74) Agent: FRASER, Janis, K.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).	(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published With international search report.	

(54) Title: GENOMIC SEQUENCES UPSTREAM OF THE CODING REGION OF THE G-CSF GENE FOR PROTEIN PRODUCTION AND DELIVERY

**(57) Abstract**

An isolated nucleic acid molecule that hybridizes under stringent conditions, or shares at least 80 % sequence identity, with a defined genomic region upstream of the coding region of the G-CSF gene, and a DNA construct containing that DNA molecule as a targeting sequence for homologous recombination.

- 1 -

GENOMIC SEQUENCES UPSTREAM OF THE CODING REGION OF THE G-CSF GENE FOR PROTEIN PRODUCTION AND DELIVERY

Field of the Invention

This invention relates to genomic DNA.

5

Background of the Invention

Current approaches to treating disease with therapeutic proteins include both administration of proteins produced *in vitro* and gene therapy. *In vitro* production of a protein generally involves the introduction of exogenous DNA coding for the protein of interest into appropriate host cells in culture. Gene therapy methods, on the other hand, involve administering to a patient genetically engineered cells, plasmids, viruses that contain a sequence encoding the therapeutic protein of interest.

Certain therapeutic proteins may also be produced by altering the expression of their endogenous genes in a desired manner with gene targeting techniques. See, e.g., U.S. Patent Nos. 5,641,670, 5,733,761, and 5,272,071, WO 91/06666, WO 91/06667, and WO 90/11354, all of which are incorporated by reference in their entirety.

Summary of the Invention

The present invention is based upon the identification and sequencing of genomic DNA 5' to the coding sequence of the human granulocyte colony-stimulating factor ("G-CSF") gene. This DNA can be used, for example, in a DNA construct that alters (e.g., increases) expression of an endogenous G-CSF gene in a mammalian cell upon integration into the genome of the cell via homologous recombination. "Endogenous G-CSF gene" refers to a genomic (i.e., chromosomal) copy of a gene that encodes G-CSF. The construct contains a

- 2 -

targeting sequence including or derived from the newly disclosed 5' noncoding sequence, and a transcriptional regulatory sequence. The transcriptional regulatory sequence preferably differs in sequence from the

5 transcriptional regulatory sequence of the endogenous G-CSF gene. The targeting sequence directs the integration of the regulatory sequence into a region within or upstream of the G-CSF-coding sequences of the target gene such that the regulatory sequence becomes operatively

10 linked to the endogenous coding sequence. By "operatively linked" is meant that the regulatory sequence can direct expression of the endogenous G-CSF-coding sequence. The construct may additionally contain a selectable marker gene to facilitate selection of cells

15 that have stably integrated the construct, and/or another coding sequence operatively linked to a promoter.

In one embodiment, the DNA construct contains: (a) a targeting sequence, (b) a regulatory sequence, (c) an exon, and (d) a splice-donor site. The targeting

20 sequence directs the integration of itself and elements (b) - (d) into a region within or upstream of the G-CSF-coding sequences of the target gene. Once integrated, element (b) can direct transcription of elements (c) and (d) and all downstream coding sequences of the endogenous

25 gene. In the construct, the exon is generally 3' of the regulatory sequence, and the splice-donor site is at the 3' end of the exon.

In another embodiment, the DNA construct comprises: (a) a targeting sequence, (b) a regulatory

30 sequence, (c) an exon, (d) a splice-donor site, (e) an intron, and (f) a splice-acceptor site, wherein the targeting sequence directs the integration of itself and elements (b) - (f) such that elements (b) - (f) are within or upstream of the endogenous gene. The

35 regulatory sequences then directs production of a

- 3 -

transcript that includes not only elements (c) - (f), but also endogenous G-CSF coding sequences. Preferably, the construct-derived intron and splice-acceptor site are situated in the construct downstream from the splice-
5 donor site.

The targeting sequence is homologous to a pre-selected target site in the genome with which homologous recombination is to occur. It contains at least 20 (e.g., at least 50 or 100) contiguous nucleotides from
10 SEQ ID NO:5, which represents nucleotides -6578 to -364 relative to the translation start site of the human G-CSF gene. By "homologous" is meant that the targeting sequence is identical or sufficiently similar to its genomic target site so that the targeting sequence and
15 target site can undergo homologous recombination. A small percentage of basepair mismatches is acceptable, as long as homologous recombination can occur at a useful frequency. To facilitate homologous recombination, the targeting sequence is preferably at least about 20 (e.g.,
20 50, 100, 250, 400, or 1,000) base pairs ("bp") long. The targeting sequence can also include genomic sequences from outside the region covered by SEQ ID NO:5, so long as it includes at least 20 nucleotides from within this region. For example, additional targeting sequence could
25 be derived from the sequence lying between SEQ ID NO:5 and the endogenous transcription initiation sequence of the G-CSF gene.

Due to polymorphism that may exist at the G-CSF genetic locus, minor variations in the nucleotide
30 composition of any given genomic target site may occur in any given mammalian species. Targeting sequences that correspond to such polymorphic variants of SEQ ID NO:5 (particularly human polymorphic variants) are within the scope of this invention.

- 4 -

Upon homologous recombination, the regulatory sequence of the construct is integrated into a pre-selected region upstream of the coding sequence of a G-CSF gene in a chromosome of a cell. The resulting new
5 transcription unit containing the construct-derived regulatory sequence alters the expression of the target G-CSF gene. The G-CSF protein so produced may be identical in sequence to the G-CSF protein encoded by the unaltered, endogenous gene, or may contain additional,
10 substituted, or fewer amino acid residues as compared to the wild type G-CSF protein, due to changes introduced as a result of homologous recombination.

Altering gene expression encompasses activating (or causing to be expressed) a gene which is normally
15 silent (i.e., essentially unexpressed) in the cell as obtained, increasing or decreasing the expression level of a gene, and changing the regulation pattern of a gene such that the pattern is different from that in the cell as obtained. "Cell as obtained" refers to the cell prior
20 to homologous recombination.

Also within the scope of the invention is a method of using the present DNA construct to alter expression of an endogenous G-CSF gene in a mammalian cell. This method includes the steps of (i) introducing the DNA
25 construct into the mammalian cell, (ii) maintaining the cell under conditions that permit homologous recombination to occur between the construct and a genomic target site homologous to the targeting sequence, to produce a homologously recombinant cell; and (iii)
30 maintaining the homologously recombinant cell under conditions that permit expression of the G-CSF coding sequence under the control of the construct-derived regulatory sequence. At least a part of the genomic target site is 5' to the coding sequence of an endogenous

- 5 -

G-CSF gene. That is, the genomic target site can contain coding sequence as well as 5' non-coding sequence.

The invention also features transfected or infected cells in which the construct has undergone homologous recombination with genomic DNA upstream of the endogenous ATG initiation codon in one or both alleles of the endogenous G-CSF gene. Such transfected or infected cells, also called homologously recombinant cells, have an altered G-CSF expression pattern. These cells are particularly useful for *in vitro* G-CSF production and for delivering G-CSF via gene therapy. Methods of making and using such cells are also embraced by the invention. The cells can be of vertebrate origin such as mammalian (e.g., human, non-human primate, cow, pig, horse, goat, sheep, cat, dog, rabbit, mouse, guinea pig, hamster, or rat) origin.

The invention further relates to a method of producing a mammalian G-CSF protein *in vitro* or *in vivo* by introducing the above-described construct into the genome of a host cell via homologous recombination. The homologously recombinant cell is then maintained under conditions that allow transcription, translation, and optionally, secretion of the G-CSF protein.

The invention also features an isolated nucleic acid comprising a sequence of at least 20 (e.g., at least 30, 50, 100, 200, or 1000) contiguous nucleotides of SEQ ID NO:5 or its complement, or of a sequence identical to SEQ ID NO:5 except for polymorphic variations or other minor variations (e.g., less than 5% of the sequence) which does not prevent homologous recombination with the target sequence. In one embodiment, the isolated nucleic acid of the invention includes a contiguous 100 bp block of SEQ ID NO:5. For example, the isolated DNA can contain nucleotides 1 to 100, 101 to 200, 201 to 300, 301 to 400, 401 to 500, 501 to 600, 601 to 700, 701 to 800,

- 6 -

801 to 900, 901 to 1000, 1001 to 1100, 1101 to 1200, 1201 to 1300, 1301 to 1400, 1401 to 1500, 1501 to 1600, 1601 to 1700, 1701 to 1800, 1801 to 1900, 1901 to 2000, 2001 to 2100, 2101 to 2200, 2201 to 2300, 2301 to 2400, 2401 to 2500, 2501 to 2600, 2601 to 2700, 2701 to 2800, 2801 to 2900, 2901 to 3000, 3001 to 3100, 3101 to 3200, 3201 to 3300, 3301 to 3400, 3401 to 3500, 3501 to 3600, 3601 to 3700, 3701 to 3800, 3801 to 3900, 3901 to 4000, 4001 to 4100, 4101 to 4200, 4201 to 4300, 4301 to 4400, 4401 to 4500, 4501 to 4600, 4601 to 4700, 4701 to 4800, 4801 to 4900, 4901 to 5000, 5001 to 5100, 5101 to 5200, 5201 to 5300, 5301 to 5400, 5401 to 5500, 5501 to 5600, 5601 to 5700, 5701 to 5800, 5801 to 5900, 5901 to 6000, 6001 to 6100, 6101 to 6200, or 6136 to 6235 of SEQ ID NO:5 or its complement. These blocks of SEQ ID NO:5 and its complement are also useful as targeting sequences in the constructs of the invention.

In the isolated DNA, the SEQ ID NO:5-derived sequence is not linked to a full-length G-CSF-coding sequence, or at least is not linked in the same configuration (i.e., separated by the same noncoding sequence) as occurs in any native genome. The term "isolated DNA", as used herein, thus does not denote a chromosome or large piece of genomic DNA (as might be incorporated into a cosmid or yeast artificial chromosome) that includes not only part or all of SEQ ID NO:5, but also an intact G-CSF coding sequence and all of the sequence which lies between the G-CSF coding sequence and the sequence corresponding to SEQ ID NO:5 as it exists in the genome of a cell. It does include, but is not limited to, a DNA (i) which is incorporated into a plasmid or virus; or (ii) which exists as a separate molecule independent of other sequences, e.g., a fragment produced by polymerase chain reaction ("PCR") or restriction endonuclease treatment. The isolated DNA

- 7 -

preferably does not contain a sequence which encodes intact G-CSF precursor (i.e., G-CSF complete with its endogenous secretion signal peptide).

The invention also includes isolated DNA
5 comprising a strand which contains a sequence that is at least 100 (e.g., at least 200, 400, or 1000) nucleotides in length and that hybridizes under either highly stringent or moderately stringent conditions with SEQ ID NO:5, or the complement of SEQ ID NO:5. The sequence is
10 not linked to a G-CSF-coding sequence, or at least is not linked in the same configuration as occurs in any native genome. By moderately stringent conditions is meant hybridization at 50°C in Church buffer (7% SDS, 0.5% NaHPO₄, 1 M EDTA, 1% bovine serum albumin) and washing at
15 50°C in 2X SSC. Highly stringent conditions are defined as hybridization at 42°C in the presence of 50% formamide; a first wash at 65°C with 2X SSC containing 1% SDS; followed by a second wash at 65°C with 0.1X SSC.

Also embraced by the invention is isolated DNA
20 comprising a strand which contains a sequence that is at least 100 (e.g., at least 200, 400, or 1000) nucleotides in length and that shares at least 80% (e.g., at least 85%, 90%, 95%, or 98%) sequence identity with a segment of equal length from SEQ ID NO:5 or the complement
25 thereof. The sequence is not linked to a G-CSF-coding sequence, or at least is not linked in the same configuration as occurs in any native genome.

Where a particular polypeptide or nucleic acid molecule is said to have a specific percent identity or
30 conservation to a reference polypeptide or nucleic acid molecule, the percent identity or conservation is determined by the algorithm of Myers and Miller, CABIOS (1989), which is embodied in the ALIGN program (version 2.0), or its equivalent, using a gap length penalty of 12
35 and a gap penalty of 4 where such parameters are

- 8 -

required. All other parameters are set to their default positions. Access to ALIGN is readily available. See, e.g., <http://www2.igh.cnrs.fr/bin/align-guess.cgi> on the Internet.

5 The invention also features a method of delivering G-CSF to an animal (e.g., a mammal such as a human, non-human primate, cow, pig, horse, goat, sheep, cat, dog, rabbit, mouse, guinea pig, hamster, or rat) by providing a cell whose endogenous G-CSF gene has been activated as
10 described herein, and implanting the cell in the animal, where the cell secretes G-CSF. Also included in the invention is a method of producing G-CSF by providing a cell whose endogenous G-CSF gene has been activated as described herein, and culturing the cell in vitro under
15 conditions which permit the cell to express and secrete G-CSF.

 The isolated DNA of the invention can be used, for example, as a source of an upstream PCR primer for use (when combined with a suitable downstream primer) in
20 obtaining the regulatory and/or coding regions of an endogenous G-CSF gene, or as a hybridization probe for indicating the presence of chromosome 17 in a preparation of human chromosomes. It can also be used, as described below, in a method for altering the expression of an
25 endogenous G-CSF gene in a vertebrate cell.

 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Exemplary methods and
30 materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention. All publications, patent applications, patents, and other references mentioned
35 herein are incorporated by reference in their entirety.

- 9 -

In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

5 Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Fig. 1 is a schematic diagram showing the genomic
10 structure of the human G-CSF gene.

Fig. 2 is a schematic diagram showing the human G-CSF genomic regions encompassed by the inserts of plasmids pHGCSF1 and pHGCSF4.

Fig. 3 is a representation of a partial sequence
15 (SEQ ID NO:1) of a human G-CSF gene, including 6,578 nucleotides of the sequence 5' to the ATG initiation codon. Also shown is a partial polypeptide sequence (SEQ ID NO:2) encoded by the coding sequence. Sequences derived from the junction of the genomic insert and the
20 phage arm in the G-CSF/3 phage clone are underlined.

Fig. 4 is a schematic diagram showing a construct of the invention. The construct contains a first targeting sequence (1); an amplifiable marker gene (AM); a selectable marker gene (SM); a regulatory sequence; a
25 CAP site; an exon encoding part of the signal peptide of G-CSF; an unpaired splice-donor site (SD); and a second targeting sequence (2). The black boxes represent coding DNA and the stippled boxes represent transcribed but untranslated sequences.

30 Fig. 5 is a representation of SEQ ID NO:5, a genomic sequence upstream of a human G-CSF transcription start site.

- 10 -

Fig. 6 is a representation of a first targeting sequence (SEQ ID NO:6) used in a construct of the invention.

Fig. 7 is a representation of a second targeting
5 sequence (SEQ ID NO:7) used in a construct of the invention.

Detailed Description

The present invention is based on the discovery of the nucleotide composition of sequence upstream to the
10 coding sequence of a human G-CSF gene.

G-CSF is a cytokine that stimulates the proliferation and differentiation of hematopoietic progenitor cells committed to the neutrophil/granulocyte lineage. G-CSF is routinely used in the prevention of
15 chemotherapy-induced neutropenia and in association with bone marrow transplantation. Chronic idiopathic and congenital neutropenic disorders also show improvement after G-CSF injection.

The human G-CSF gene encodes a 204 or 207 amino
20 acid precursor protein containing a 30 amino acid signal peptide. The genomic map of the human G-CSF gene is shown in Fig. 1. The map is constructed based on a 2,960 bp published sequence (HUMGCSFG, GenBank accession number X03656) which begins at position -363 relative to the
25 translational start site (unless otherwise specified, all positions referred to herein are relative to the translational start site). The gene contains five exons and four introns, with the first exon encoding 13 2/3 amino acids of the signal peptide (i.e., the first exon
30 contains 13 codons and the first two nucleotides of the 14th codon encoding the signal peptide).

- 11 -

Sequence 5' to the Human G-CSF Gene

To obtain genomic DNA containing upstream sequence of a G-CSF gene, a human leukocyte genomic library in lambda EMBL3 (Clontech catalog # HL1006d) was screened with a 729 bp oligonucleotide probe generated by PCR. This probe includes G-CSF exons 1 and 2, and was amplified from human genomic DNA using oligonucleotide primers designated 102 and 105, both of which were designed from the available G-CSF genomic DNA sequence (Fig. 1). The 5' end of primer 102 corresponds to position -345, and the primer's sequence is 5'-TATCAGCGGCTCAGCCTTTG-3' (SEQ ID NO:3). The 5' end of primer 105 corresponds to position +384, and the primer's sequence is 5'-CCACCTCACTCACCAGCTTCTC-3' (SEQ ID NO:4).

Approximately 1.5 million recombinant phage were screened with the radiolabelled 729 bp probe. Four independent phage plaques were isolated. One of them, designated clone G-CSF/3, was used for subsequent studies.

A 6.5 kb HindIII-KpnI fragment from phage G-CSF/3 was subcloned into pBluescript II SK+ (Stratagene, La Jolla, CA) to produce pHGCSF1, which contains the upstream sequences and the entire protein-coding region of the G-CSF gene. An additional upstream subclone, pHGCSF4, was prepared from the 3.3 kb SalI fragment, which overlaps by about 0.4 kb with the insert of pHGCSF1 (Fig. 2).

The pHGCSF1 and pHGCSF4 plasmids were sequenced by the Sanger method. The sequence data sets were aligned to obtain the sequence of a 6.6 kb region immediately upstream of the transcription initiation site of the human G-CSF gene, starting at position -6,578. This sequence is part of SEQ ID NO:1, shown in Fig. 3.

The 19 bp (underlined in Fig. 3) at the 5' end of SEQ ID NO:1 are derived from the junction of the genomic

- 12 -

insert and the phage arm in the G-CSF/3 phage clone. Therefore, the SalI site in this 19 bp region is not present in the human genome from which the phage library is derived. The sequence between positions -6,578 and -
5 364 (SEQ ID NO:5) is human genomic sequence from a region upstream of the previously-published G-CSF genomic sequence, and has not been reported previously.

To alter the expression of an endogenous G-CSF gene, a DNA fragment containing nucleotides 1470 to 4723
10 of SEQ ID NO:5 was cloned into plasmid pGG13 upstream of a CMV promoter and a neomycin resistance gene. Nucleotides 1470 to 4723 (SEQ ID NO:6) represent the first targeting sequence as schematically represented in Fig. 4. For the second targeting sequence of Fig. 4, a
15 DNA fragment containing nucleotides 4728 to 5979 (SEQ ID NO:7), relative to the translation start site, of the G-CSF gene sequence was cloned downstream of the CMV promoter and neomycin resistance gene. The pGG13 plasmid was introduced into human fibroblast cells exhibiting
20 little or no G-CSF gene expression to allow homologous recombination with the endogenous G-CSF gene. Cells resistant to G418 after plasmid introduction were screened to identify cells with increased G-CSF gene expression, as would be expected if a homologous
25 recombination event between pGG13 and the genomic DNA took place in the vicinity of the endogenous G-CSF gene.

General Methodologies

Alteration of Endogenous G-CSF Expression

Using the above-described G-CSF upstream
30 sequences, one can alter the expression of an endogenous human G-CSF gene by a method as generally described in U.S. Patent No. 5,641,670. One strategy is shown in Fig. 4. In this strategy, a targeting construct is designed to include a first targeting sequence homologous to a

- 13 -

first target site upstream of the gene, an amplifiable marker gene, a selectable marker gene, a regulatory region, a CAP site, an exon encoding an amino acid sequence which is identical or functionally equivalent to
5 that of the first 13 2/3 amino acids of the G-CSF signal peptide (i.e., the first exon contains 13 codons and the first two nucleotides of the 14th codon encoding the signal peptide), a splice-donor site, and a second targeting sequence homologous to a second target site
10 downstream of the first target site and terminating either within or upstream of the G-CSF coding sequence. In this strategy, the first and second target sites are immediately adjacent in the chromosome prior to homologous recombination, but such a configuration is not
15 required (see also below). Homologously recombinant cells will produce an mRNA precursor which corresponds to the exogenous exon and splice-donor site, and any sequence between the splice donor site and the transcription termination sequence of the G-CSF gene,
20 including the G-CSF introns, exons, and 3' untranslated region (Fig. 4). Splicing of this transcript results in a mRNA in which the exogenous exon is fused to exon 2 of the endogenous G-CSF gene. Translation of the mRNA produces a precursor G-CSF. The inclusion of a coding
25 exon in the DNA construct allows one to make any desirable modifications to the signal peptide.

Other approaches can also be employed. For example, the first and/or second target sites can be in the first intron of the G-CSF gene. Alternatively, the
30 DNA construct may be designed to include, from 5' to 3', a first targeting sequence, an amplifiable marker gene, a selectable marker gene, a regulatory region, a CAP site, an exon, a splice-donor site, an intron, a splice-acceptor site, and a second targeting sequence. For this
35 strategy, the 5' end of

- 14 -

the second target site is preferably less than 60 bp upstream of the normal G-CSF translational initiation site, in order to avoid undesired ATG start codons. A mRNA precursor produced from the homologously recombined locus will include the exogenous exon, the exogenous splice-donor site, the exogenous intron, the exogenous splice-acceptor site, the second targeting sequence, and any sequences between the second targeting sequence and the 3' untranscribed region of the endogenous gene.

Splicing of this transcript will generate a mRNA which can be translated to produce a precursor of human G-CSF, having either the normal G-CSF secretion signal sequence or a genetically engineered secretion signal sequence. The size of the exogenous intron and thus the position of the exogenous regulatory region relative to the coding region of the gene can be varied to optimize the function of the regulatory region.

In any activation strategy, the first and second target sites need not be immediately adjacent or even be near each other. When they are not immediately adjacent to each other, a portion of the G-CSF gene's normal upstream region and/or a portion of the coding region would be deleted upon homologous recombination.

The DNA Construct

The DNA construct of the invention includes at least a targeting sequence and a regulatory sequence. It may additionally contain an exon; or an exon and an unpaired splice-donor site; or an exon, splice donor site, intron, and splice acceptor site. The exon, if present, is 3' of the regulatory sequence, and the unpaired splice-donor site is at the 3' end of the exon. The intron and splice acceptor site, if present, are 3' of the splice donor site. In addition, there can be multiple exons and introns (with appropriate splice donor and acceptor sites) preceding (i.e., 5' to) the exon

- 15 -

flanked by the unpaired splice-donor site. The DNA in the construct is referred to as exogenous, since the DNA is not an original part of the genome of a host cell. Exogenous DNA may possess sequences identical to or different from portions of the endogenous genomic DNA present in the cell prior to transfection or infection by viral vector. As used herein, "transfection" means introduction of plasmid into a cell by chemical and physical means such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, microinjection, microprojectiles, or biolistic-mediated uptake. As used herein "infection" means introduction of viral nucleic acid into a cell by virus infection. The various elements included in the DNA construct of the invention are described in detail below.

The DNA construct can also include *cis*-acting or *trans*-acting viral sequences (e.g., packaging signals), thereby enabling delivery of the construct into the nucleus of a cell via infection by a viral vector. Where necessary, the DNA construct can be disengaged from various steps of a virus life cycle, such as integrase-mediated integration in retroviruses or episome maintenance. Disengagement can be accomplished by appropriate deletions or mutations of viral sequences, such as a deletion of the integrase coding region in a retrovirus vector. Additional details regarding the construction and use of viral vectors are found in Robbins et al., Pharmacol. Ther. 80:35-47, 1998; and Gunzburg et al., Mol. Med. Today 1:410-417, 1995, herein incorporated by reference.

Targeting Sequences

Targeting sequences permit homologous recombination of a desired sequence into a selected site in the host genome. Targeting sequences are homologous

- 16 -

to (i.e., able to homologously recombine with) their respective target regions in the host genome.

A circular DNA construct can employ a single targeting sequence, or two or more separate targeting
5 sequences. A linear DNA construct may contain two or more separate targeting sequences. The target site to which a given targeting sequence is homologous can reside within an exon and/or intron of the G-CSF gene, upstream of and immediately adjacent to the G-CSF coding region,
10 or upstream of and at a distance from the G-CSF coding region.

The first of the two targeting sequences in the construct (or the entire targeting sequence, if there is only one targeting sequence in the construct) is derived
15 at least in part from the newly disclosed genomic region upstream of the G-CSF-coding sequences. This targeting sequence can contain a portion of SEQ ID NO:1, e.g., at least 20 consecutive nucleotides from the sequence corresponding to positions -6,578 to -364 (SEQ ID NO:5).
20 The second of the two targeting sequences in the construct may target a genomic region upstream of the coding sequence (e.g., also contain a portion of SEQ ID NO:5), or target an exon or intron of the gene.

The targeting sequence(s) may additionally include
25 sequence derived from a previously-disclosed region of the G-CSF gene, including those described herein, as well as a region further upstream which is structurally uncharacterized but can be mapped by one skilled in the art.

30 Genomic fragments that can be used as targeting sequences can be identified by their ability to hybridize to a probe containing all or a portion of SEQ ID NO:5. Such a probe can be generated by PCR using primers derived from SEQ ID NO:1.

35 The Regulatory Sequence

- 17 -

The regulatory sequence of the DNA construct can contain one or more promoters (e.g., a constitutive, tissue-specific or inducible promoter), enhancers, scaffold-attachment regions or matrix attachment sites,
5 negative regulatory elements, transcription factor binding sites, or combinations of these elements.

The regulatory sequence can be derived from a eukaryotic (e.g., mammalian) or viral genome. Useful regulatory sequences include, but are not limited to,
10 those that regulate the expression of SV40 early or late genes, cytomegalovirus genes, and adenovirus major late genes. They also include regulatory regions derived from genes encoding mouse metallothionein-I, elongation factor-1 α , collagen (e.g., collagen I α 1, collagen I α 2, and collagen IV), actin (e.g., γ -actin), immunoglobulin,
15 HMG-CoA reductase, glyceraldehyde phosphate dehydrogenase, 3-phosphoglyceratekinase, collagenase, stromelysin, fibronectin, vimentin, plasminogen activator inhibitor I, thymosin β 4, tissue inhibitors of
20 metalloproteinase, ribosomal proteins, major histocompatibility complex molecules, and human leukocyte antigens.

The regulatory sequence preferably contains a transcription factor binding site such as a TATA Box,
25 CCAAT Box, AP1, Sp1, or a NF- κ B binding site.

Marker Genes

If desired, the construct can include a sequence encoding a desired polypeptide, operatively linked to its own promoter. An example of this would be a selectable
30 marker gene, which can be used to facilitate the identification of a targeting event. An amplifiable marker gene can also be used to facilitate selection of cells having co-amplified flanking DNA sequences. Cells containing amplified copies of the amplifiable marker
35 gene can be identified by growth in the presence of an

- 18 -

agent that selects for the expression of the amplifiable gene. The activated endogenous G-CSF gene will typically be amplified in tandem with the amplified selectable marker gene. Cells containing multiple copies of the
5 activated endogenous gene may produce very high levels of G-CSF, and are thus useful for *in vitro* protein production and gene therapy.

The selectable and amplifiable marker genes do not have to lie immediately adjacent to each other. The
10 amplifiable marker gene and selectable marker gene can be the same gene. One or both of the marker genes can be situated in the intron of the DNA construct. Suitable amplifiable marker genes and selectable marker genes are described in U.S. Patent No. 5,641,670.

15 The Exogenous Exon

The DNA construct may further contain an exon, i.e., a DNA sequence that is copied into RNA and is present in a mature mRNA molecule. The exon in the construct is referred to herein as an exogenous exon.
20 The exogenous exon can be identical to or differ from the first exon of the human G-CSF gene. Alternatively, the exogenous exon encodes one or more amino acid residues, or partially encodes an amino acid residue (i.e., contains one or two nucleotides of a codon). When the
25 exon contains a coding sequence, the DNA construct should be designed such that, upon transcription and splicing, the reading frame of the resulting mRNA is in-frame with the coding region of the target G-CSF gene. That is, the exogenous exon is spliced to an endogenous exon in a
30 manner that does not change the appropriate reading frame of the portion of the mRNA derived from the endogenous exon.

The inclusion of a coding exon in the targeting construct allows the production of a fusion protein that
35 contains both endogenous G-CSF protein sequence and

- 19 -

exogenous protein sequence. Such a hybrid protein may combine the structural, enzymatic, or ligand- or receptor-binding properties from two or more proteins into one polypeptide. For example, the exogenous exon
5 can encode a cell membrane anchor, a signal peptide to improve cellular secretion, a leader sequence, an enzymatic region, a co-factor binding region, or an epitope tag to facilitate purification of the G-CSF hybrid protein produced from the recombined gene locus.

10 The Splice-Donor Site

The exogenous exon is flanked at its 3' end by a splice-donor site. A splice-donor site is a sequence which directs the splicing of one exon of an RNA transcript to the splice-acceptor site of another exon of
15 the RNA transcript. Typically, the first exon lies 5' of the second exon, and the splice-donor site located at the 3' end of the first exon is paired with a splice-acceptor site on the 5' side of the second exon. Splice-donor sites have a characteristic consensus sequence represented as (A/C)AGGURAGU (where R denotes a purine), with
20 the GU in the fourth and fifth positions being required (Jackson, Nucleic Acids Research 19: 3715-3798, 1991). The first three bases of the splice-donor consensus site are the last three bases of the exon: i.e., they are not
25 spliced out. Splice-donor sites are functionally defined by their ability to effect the appropriate reaction within the mRNA splicing pathway.

By way of example, the splice-donor site can be placed immediately adjacent and 3' to an ATG codon when
30 the presence of one or more intervening nucleotides is not required for the exogenous exon to be in-frame with the second exon of the targeted gene. When the exogenous exon encodes one or more amino acids in-frame with the coding sequence of the targeted gene, the splice-donor

- 20 -

site may preferably be placed immediately adjacent to the exogenous coding sequence on its 3' side.

The splice-donor site flanking the exogenous exon is unpaired in the construct, i.e., in the construct
5 itself there is no accompanying splice-acceptor site downstream of the splice-donor site to which the latter can be spliced. Following homologous recombination into the target site upstream of the G-CSF coding sequence, what was the construct's unpaired splice-donor site is
10 functionally paired with an endogenous splice-acceptor site of an endogenous exon of G-CSF. Processing of the transcript produced from the homologously recombined G-CSF gene results in splicing of the exogenous exon to the splice-acceptor site of an endogenous exon.

15 The construct of the invention can also include a splice acceptor site. This site, in conjunction with a splice donor site, directs the splicing of one exon to another exon. Splice-acceptor sites have a characteristic sequence represented as (Y)₁₀NYAG (SEQ ID NO:8), where
20 Y denotes any pyrimidine and N denotes any nucleotide (Jackson, Nucleic Acids Research 19:3715-3798, 1991).

Introns

The DNA construct may optionally contain an intron. An intron is a sequence of one or more
25 nucleotides lying between a splice-donor site and a splice-acceptor site, and is removed, by splicing, from a precursor RNA molecule in the formation of a mature mRNA molecule.

- 21 -

The CAP Site

The DNA construct can optionally contain a CAP site. A CAP site is a specific transcription start site which is associated with and utilized by the regulatory region. This CAP site is located at a position relative to the regulatory sequence in the construct such that following homologous recombination, the regulatory sequence directs synthesis of a transcript that begins at the CAP site. Alternatively, no CAP site is included in the construct, and the transcriptional apparatus will locate by default an appropriate site in the targeted gene to be utilized as a CAP site.

Additional DNA elements

The construct may additionally contain sequences which affect the structure or stability of the RNA or protein produced by homologous recombination. Optionally, the DNA construct can include a bacterial origin of replication and bacterial antibiotic resistance markers or other selectable markers, which allow for large-scale plasmid propagation in bacteria or any other suitable cloning/host system.

All of the above-described elements of the DNA construct are operatively linked or functionally placed with respect to each other. That is, upon homologous recombination between the construct and the targeted genomic DNA, the regulatory sequence can direct the production of a primary RNA transcript which initiates at a CAP site (optionally included in the construct) and includes (i) sequence corresponding to the exon and splice-donor site of the construct, if they are present, and (ii) sequence lying between that splice-donor site and the endogenous gene's transcription stop site. The latter sequence may include the G-CSF gene's endogenous regulatory region as well as sequences neighboring that region that are normally not transcribed. In an opera-

- 22 -

tively linked configuration, the splice-donor site of the targeting construct directs a splicing event to a splice-acceptor site flanking one of the exons of the endogenous G-CSF gene, such that the desired protein can be produced
5 from the fully spliced mature transcript. The splice-acceptor site can be endogenous, such that the splicing event is directed to an endogenous exon. In another embodiment where the splice-acceptor site is included in the targeting construct, the splicing event removes the
10 exogenous intron introduced by the targeting construct.

The order of elements in the DNA construct can vary. Where the construct is a circular plasmid or viral vector, the relative order of elements in the resulting structure can be, for example: a targeting sequence,
15 plasmid DNA (comprised of sequences used for the selection and/or replication of the targeting plasmid in a microbial or other suitable host), selectable marker(s), a regulatory sequence, an exon, and an unpaired splice-donor site.

20 Where the construct is linear, the order can be, for example: a first targeting sequence, a selectable marker gene, a regulatory sequence, an exon, a splice-donor site, and a second targeting sequence; or, in the alternative, a first targeting sequence, a regulatory
25 sequence, an exon, a splice-donor site, a selectable marker gene, and a second targeting sequence. The order of the elements can also be: a first targeting sequence, a selectable marker, a regulatory sequence, an exon, a splice-donor site, an intron, a splice-acceptor site,
30 optionally an internal ribosomal entry site, and a second targeting sequence.

Alternatively, the order can be : a first targeting sequence, a first selectable marker gene, a regulatory sequence, an exon, a splice-donor site, a
35 second targeting sequence, and a second selectable marker

- 23 -

gene; or, a first targeting sequence, a regulatory sequence, an exon, a splice-donor site, a first selectable marker gene, a second targeting sequence, and a second selectable marker gene. Recombination between the

5 targeting sequences flanking the first selectable marker with homologous sequences in the host genome results in the targeted integration of the first selectable marker, while the second selectable marker is not integrated. Desired transfected or infected cells are those that are

10 stably transfected or infected with the first, but not second, selectable marker. Such cells can be selected for by growth in a medium containing an agent which selects for expression of the first marker and another agent which selects against the second marker.

15 Transfected or infected cells that have improperly integrated the targeting construct by a mechanism other than homologous recombination would be expected to express the second marker gene and will thereby be killed in the selection medium.

20 A positively selectable marker is sometimes included in the construct to allow for the selection of cells containing amplified copies of that marker. In this embodiment, the order of construct components can be, for example: a first targeting sequence, an

25 amplifiable positively selectable marker, a second selectable marker (optional), a regulatory sequence, an exon, a splice-donor site, and a second targeting DNA sequence.

The various elements of the construct can be

30 obtained from natural sources (e.g., genomic DNA), or can be produced using genetic engineering techniques or synthetic processes. The regulatory region, CAP site, exon, splice-donor site, intron, and splice acceptor site of the construct can be isolated as a complete unit from,

35 e.g., the human elongation factor-1 α (Genbank sequence

- 24 -

HUMEF1A) gene or the cytomegalovirus (Genbank sequence HEHCMVP1) immediate early region. These components can also be isolated from separate genes.

Transfection or Infection and Homologous Recombination

5 The DNA construct of the invention can be introduced into the cell, such as a primary, secondary, or immortalized cell, as a single DNA construct, or as separate DNA sequences which become incorporated into the chromosomal or nuclear DNA of a transfected or infected
10 cell. The DNA can be introduced as a linear, double-stranded (with or without single-stranded regions at one or both ends), single-stranded, or circular molecule. The DNA construct or its RNA equivalent can also be introduced as a viral nucleic acid.

15 When the construct is introduced into host cells in two separate DNA fragments, the two fragments share DNA sequence homology (overlap) at the 3' end of one fragment and the 5' end of the other, while one carries a first targeting sequence and the other carries a second
20 targeting sequence. Upon introduction into a cell, the two fragments can undergo homologous recombination to form a single molecule with the first and second targeting sequences flanking the region of overlap between the two original fragments. The product molecule
25 is then in a form suitable for homologous recombination with the cellular target sites. More than two fragments can be used, with each of them designed such that they will undergo homologous recombination with each other to ultimately form a product suitable for homologous
30 recombination with the cellular target sites as described above.

 The DNA construct of the invention, if not containing a selectable marker itself, can be co-transfected or co-infected with another construct that
35 contains such a marker. A targeting plasmid may be

- 25 -

cleaved with a restriction enzyme at one or more sites to create a linear or gapped molecule prior to transfection or infection. The resulting free DNA ends increase the frequency of the desired homologous recombination event.

5 In addition, the free DNA ends may be treated with an exonuclease to create overhanging 5' or 3' single-stranded DNA ends (e.g., at least 30 nucleotides in length, and preferably 100-1000 nucleotides in length) to increase the frequency of the desired homologous
10 recombination event. In this embodiment, homologous recombination between the targeting sequence and the genomic target will result in two copies of the targeting sequences, flanking the elements contained within the introduced plasmid.

15 The DNA constructs may be transfected into cells (preferably *in vitro*) by a variety of physical or chemical methods, including electroporation, microinjection, microprojectile bombardment, calcium phosphate precipitation, liposome delivery, or polybrene- or DEAE
20 dextran-mediated transfection.

The transfected or infected cell is maintained under conditions which permit homologous recombination, as described in the art (see, e.g., Capecchi, Science 24:1288-1292, 1989). By "transfected cell" is meant a
25 cell into which (or into an ancestor of which) a DNA molecule has been introduced by a means other than using a viral vector. By "infected cell" is meant a cell into which (or into an ancestor of which) a DNA or RNA molecule has been introduced using a viral vector.

30 Viruses known to be useful as vectors include adenovirus, adeno-associated virus, Herpes virus, mumps virus, poliovirus, lentivirus, retroviruses, Sindbis virus, and vaccinia viruses such as canary pox virus. When the homologously recombinant cell is maintained under
35 conditions sufficient to permit transcription of the DNA,

- 26 -

the regulatory region introduced by the DNA construct will alter transcription of the G-CSF gene.

Homologously recombinant cells (i.e., cells that have undergone the desired homologous recombination) can be identified by phenotypic screening or by analyzing the culture supernatant in enzyme-linked immunosorbent assays (ELISA) for G-CSF. Commercial ELISA kits for detecting G-CSF are available from R&D Systems (Minneapolis, MN). Homologously recombinant cells can also be identified by Southern and Northern analyses or by polymerase chain reaction (PCR) screening.

As used herein, the term "primary cells" includes (i) cells present in a suspension of cells isolated from a vertebrate tissue source (prior to their being plated, i.e., attached to a tissue culture substrate such as a dish or flask), (ii) cells present in an explant derived from tissue, (iii) cells plated for the first time, and (iv) cell suspensions derived from these plated cells. Primary cells can also be cells as they naturally occur within a human or an animal.

Secondary cells are cells at all subsequent steps in culturing. That is, the first time that plated primary cells are removed from the culture substrate and replated (passaged), they are referred to herein as secondary cells, as are all cells in subsequent passages. Secondary cell strains consist of secondary cells which have been passaged one or more times. Secondary cells typically exhibit a finite number of mean population doublings in culture and the property of contact-inhibited, anchorage-dependent growth (anchorage-dependence does not apply to cells that are propagated in suspension culture). Primary and secondary cells are not immortalized.

Immortalized cells are cell lines (as opposed to cell strains, with the designation "strain" reserved for

- 27 -

primary and secondary cells) that exhibit an apparently unlimited lifespan in culture.

Cells selected for transfection or infection can fall into four types or categories: (i) cells which do not, as obtained, make or contain more than trace amounts of the G-CSF protein, (ii) cells which make or contain the protein but in quantities other than those desired (such as, in quantities less than the level which is physiologically normal for the type of cells as obtained), (iii) cells which make the protein at a level which is physiologically normal for the type of cells as obtained, but are to be augmented or enhanced in their content or production, and (iv) cells in which it is desirable to change the pattern of regulation or induction of a gene encoding the protein.

Primary, secondary and immortalized cells to be transfected or infected by the present method can be obtained from a variety of tissues and include all appropriate cell types which can be maintained in culture. For example, suitable primary and secondary cells include fibroblasts, keratinocytes, epithelial cells (e.g., mammary epithelial cells, intestinal epithelial cells), endothelial cells, glial cells, neural cells, formed elements of the blood (e.g., lymphocytes, bone marrow cells), muscle cells, and precursors of these somatic cell types. Where the homologously recombinant cells are to be used in gene therapy, primary cells are preferably obtained from the individual to whom the transfected or infected primary or secondary cells are to be administered. However, primary cells can be obtained from a donor (i.e., an individual other than the recipient) of the same species.

Examples of immortalized human cell lines useful for protein production or gene therapy include, but are not limited to, 2780AD ovarian carcinoma cells (Van der

WO 99/57291

PCT/US99/09924

- 28 -

Blick et al., Cancer Res., 48:5927-5932, 1988), A549 (American Type Culture Collection ("ATCC") CCL 185), BeWo (ATCC CCL 98), Bowes Melanoma cells (ATCC CRL 9607), CCRF-CEM (ATCC CCL 119), CCRF-HSB-2 (ATCC CCL 120.1),

5 COLO201 (ATCC CCL 224), COLO205 (ATCC CCL 222), COLO 320DM (ATCC CCL 220), COLO 320HSR (ATCC CCL 220.1), Daudi cells (ATCC CCL 213), Detroit 562 (ATCC CCL 138), HeLa cells and derivatives of HeLa cells (ATCC CCL 2, 2.1 and 2.2), HCT116 (ATCC CCL 247), HL-60 cells (ATCC CCL 240),

10 HT1080 cells (ATCC CCL 121), IMR-32 (ATCC CCL 127), Jurkat cells (ATCC TIB 152), K-562 leukemia cells (ATCC CCL 243), KB carcinoma cells (ATCC CCL 17), KG-1 (ATCC CCL 246), KG-1a (ATCC CCL 246.1), LS123 (ATCC CCL 255), LS174T (ATCC CCL CL-188), LS180 (ATCC CCL CL-187), MCF-7

15 breast cancer cells (ATCC BTH 22), MOLT-4 cells (ATCC CRL 1582), Namalwa cells (ATCC CRL 1432), NCI-H498 (ATCC CCL 254), NCI-H508 (ATCC CCL 253), NCI-H548 (ATCC CCL 249), NCI-H716 (ATCC CCL 251), NCI-H747 (ATCC CCL 252), NCI-H1688 (ATCC CCL 257), NCI-H2126 (ATCC CCL 256), Raji

20 cells (ATCC CCL 86), RD (ATCC CCL 136), RPMI 2650 (ATCC CCL 30), RPMI 8226 cells (ATCC CCL 155), SNU-C2A (ATCC CCL 250.1), SNU-C2B (ATCC CCL 250), SW-13 (ATCC CCL 105), SW48 (ATCC CCL 231), SW403 (ATCC CCL 230), SW480 (ATCC CCL 227), SW620 (ATCC CCL 227), SW837 (ATCC CCL 235),

25 SW948 (ATCC CCL 237), SW1116 (ATCC CCL 233), SW1417 (ATCC CCL 238), SW1463 (ATCC CCL 234), T84 (ATCC CCL 248), U-937 cells (ATCC CRL 1593), WiDr (ATCC CCL 218), and WI-38VA13 subline 2R4 cells (ATCC CLL 75.1), as well as heterohybridoma cells produced by fusion of human cells

30 and cells of another species. Secondary human fibroblast strains, such as WI-38 (ATCC CCL 75) and MRC-5 (ATCC CCL 171), may be used. In addition, primary, secondary, or immortalized human cells, as well as primary, secondary, or immortalized cells from other species, can be used for

35 in vitro protein production or gene therapy.

- 29 -

G-CSF-expressing Cells

Homologously recombinant cells of the invention express G-CSF at desired levels and are useful for both *in vitro* production of G-CSF and gene therapy.

5 Protein Production

Homologously recombinant cells according to this invention can be used for *in vitro* production of G-CSF. The cells are maintained under conditions, as described in the art, which result in expression of proteins. The
10 G-CSF protein may be purified from cell lysates or cell supernatants. A pharmaceutical composition containing the G-CSF protein can be delivered to a human or an animal by conventional pharmaceutical routes known in the art (e.g., oral, intravenous, intramuscular, intranasal,
15 pulmonary, transmucosal, intradermal, transdermal, rectal, intrathecal, subcutaneous, intraperitoneal, or intralesional). Oral administration may require use of a strategy for protecting the protein from degradation in the gastrointestinal tract: e.g., by encapsulation in
20 polymeric microcapsules.

Gene Therapy

Homologously recombinant cells of the present invention are useful as populations of homologously recombinant cell lines, as populations of homologously
25 recombinant primary or secondary cells, as homologously recombinant clonal cell strains or lines, as homologously recombinant heterogenous cell strains or lines, and as cell mixtures in which at least one representative cell of one of the four preceding categories of homologously
30 recombinant cells is present. Such cells may be used in a delivery system for stimulating the proliferation and differentiation of hematopoietic progenitor cells, or for any other condition treatable with G-CSF. For instance, the cells can be used to prevent chemotherapy-induced
35 neutropenia; to treat patients undergoing, or who have

- 30 -

undergone, bone marrow transplantation; or to treat chronic idiopathic and congenital neutropenic disorders.

Homologously recombinant primary cells, clonal cell strains or heterogenous cell strains are
5 administered to an individual in whom the abnormal or undesirable condition is to be treated or prevented, in sufficient quantity and by an appropriate route, to express or make available the protein or exogenous DNA at physiologically relevant levels. A physiologically rele-
10 vant level is one which either approximates the level at which the product is normally produced in the body or results in improvement of the abnormal or undesirable condition. If the cells are syngeneic with respect to a immunocompetent recipient, the cells can be administered
15 or implanted intravenously, intraarterially, subcutaneously, intraperitoneally, intraomentally, subrenal capsularly, intrathecally, intracranially, or intramuscularly.

If the cells are not syngeneic and the recipient
20 is immunocompetent, the homologously recombinant cells to be administered can be enclosed in one or more semipermeable barrier devices. The permeability properties of the device are such that the cells are prevented from leaving the device upon implantation into
25 a subject, but the therapeutic protein is freely permeable and can leave the barrier device and enter the local space surrounding the implant or enter the systemic circulation. See, e.g., U.S. Patent Nos. 5,641,670, 5,470,731, 5,620,883, 5,487,737, and co-owned U.S. Patent
30 Application entitled "Delivery of Therapeutic Proteins" (inventors: Justin C. Lamsa and Douglas A. Treco), filed April 16, 1999, all herein incorporated by reference. The barrier device can be implanted at any appropriate site: e.g., intraperitoneally, intrathecally,

- 31 -

subcutaneously, intramuscularly, within the kidney capsule, or within the omentum.

Barrier devices are particularly useful and allow homologously recombinant immortalized cells, homologously recombinant cells from another species (homologously recombinant xenogeneic cells), or cells from a nonhistocompatibility-matched donor (homologously recombinant allogeneic cells) to be implanted for treatment of a subject. The devices retain cells in a fixed position in vivo, while protecting the cells from the host's immune system. Barrier devices also allow convenient short-term (i.e., transient) therapy by allowing ready removal of the cells when the treatment regimen is to be halted for any reason. Transfected or infected xenogeneic and allogeneic cells may also be used in the absence of barrier devices for short-term gene therapy. In that case, the G-CSF produced by the cells will be delivered in vivo until the cells are rejected by the host's immune system.

A number of synthetic, semisynthetic, or natural filtration membranes can be used for this purpose, including, but not limited to, cellulose, cellulose acetate, nitrocellulose, polysulfone, polyvinylidene difluoride, polyvinyl chloride polymers and polymers of polyvinyl chloride derivatives. Barrier devices can be utilized to allow primary, secondary, or immortalized cells from another species to be used for gene therapy in humans.

Another type of device useful in the gene therapy of the invention is an implantable collagen matrix in which the cells are embedded. Such a device, which can contain beads to which the cells attach, is described in WO 97/15195, herein incorporated by reference.

The number of cells needed for a given dose or implantation depends on several factors, including the

- 32 -

expression level of the protein, the size and condition of the host animal, and the limitations associated with the implantation procedure. Usually the number of cells implanted in an adult human or other similarly-sized
5 animal is in the range of 1×10^4 to 5×10^{10} , and preferably 1×10^8 to 1×10^9 . If desired, they may be implanted at multiple sites in the patient, either at one time or over a period of months or years. The dosage may be repeated as needed.

10

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not to limit the scope of the
15 invention, which is defined by the scope of the appended claims.

Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

WO 99/57291

PCT/US99/09924

- 1 -

SEQUENCE LISTING

<110> Transkaryotic Therapies Inc.

<120> GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY

<130> 07236/017W01

<150> US 60/084,649

<151> 1998-05-07

<160> 8

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 6679

<212> DNA

<213> Homo sapiens

<400> 1

gtcgacctgc	aggtcaacgg	atcacttgag	gacagtagtt	caagaccagc	ctgggcagca	60
tagggagact	gtctctacga	aaaatcaaaa	aattatggcc	gggcatgggtg	gctcacgtct	120
gtaatccctg	aactttggga	catcaaggca	agtggatcac	ttgagggtcag	gagttcgaga	180
ctagcctggc	caacatgggtg	aaaccctatc	tccactaaaa	aatacaaaaa	ttagccaggc	240
atgggtggcag	gcacctgtaa	tcccggctac	tcaggagggt	gaggcaggag	aatcacttga	300
acccaggagg	cggagggttg	agtgaagtga	gatcacacca	ctgcactcca	gcctgggtga	360
cagagcaaga	ctctatctca	aaaaaaataa	aaaaataaaa	aaattagcca	ggcatggtag	420
tgcacacctc	tagtctcagc	tactcaggag	gctgagggtg	gaggatcact	tgaacctggg	480
gcagtcaagg	ctacagtga	ccaagatcat	gccactacac	tccagcctgg	gcaacagaga	540
gagaccctgt	ctctaaaaaa	ataataataa	taaagaaaaa	aacagctctg	tttatgtctc	600
ctggtccata	catactacta	tgtatatagt	ttgcaaactc	aaagatccag	atagtcaatt	660
ttttaggcct	gtgggccgta	tgggtctctg	cacaatcact	ctgccctgtc	tttctagcac	720
aaaagcagct	ataaacaata	catacatgaa	ttttttatag	acatcgagat	ttgaatttca	780
tatgattttt	acattttata	aaataatctt	tttaaaaatt	ttcccctaac	catttaaaag	840
tgtaaaagcc	ggccagcgcg	ccatcgtcac	gcctgtaatt	ccagcacttt	gggaggctga	900
ggtgggcaga	tcacttgaga	tcaacagttc	gagaccagcc	tggccaacat	agcaaaaccc	960
catttctact	aaaaataaaa	aaattagctg	ggcatagtgg	tgcacacctg	tgatcccagc	1020
tacttgggag	gctgaggcag	gagaatcgct	tgaacctggg	aagcggagggt	tgcagtga	1080
caacatcatg	ccactgcact	ccagcctggg	tgacagagt	agacttcgtc	tcaacgaaaa	1140
aaaaaagtgt	aaaagccatt	cctaattcag	tgtacatcag	tgtacatact	caggtctg	1200
tactcctgct	ctgaggcata	cctgagaagt	agagttgctt	ggtcacagga	catacacatt	1260
tccacattaa	ctagacacta	ccaagttgcc	atccaaggag	gttttttttt	tacaatctac	1320
actcccccca	gcaacaaatg	agagttactc	cagatccttt	acaaagatgc	tctaagccca	1380
gtaccagatg	aaaacaggaa	gtgggagggg	aagctgccag	ccccttctaa	ccatgaagaa	1440
atacctggta	gagccttctg	gatgctggaa	ggatgaataa	cgggggtctc	tggagcctgc	1500
cccctgtcag	atcactgtga	cttctgagcc	tccagtccag	tctcagcccc	atgtgtcatg	1560
gccagtgata	atgagccctc	actctctgtt	tggctctttat	tctcccatg	tggggctgaa	1620
gtctggattg	agccgttatt	caagatgtac	agctttcttg	acaggaaagt	agtgtcacag	1680
aaacagcagg	ggcttggcaa	gatgatctaa	ctgcaaattc	tacctggctc	agccaccagc	1740
tagttctgtg	atcttgaaca	agtttttttca	cttctctgag	gccatccctt	ggctacaaca	1800
caccagttgg	ttgacaggat	gaaatgacga	agtcccttac	acctgtaatc	ccagcacttt	1860
gggaggccaa	ggcgggtgga	tggcttgagc	ctgagagggtg	acagcatgcc	ggcagtcctc	1920
acagccctcg	ttcgctctcg	gcgcctcctc	tgcttgggct	cccacttcgg	tggcacttga	1980
ggagcccttc	agcccaccgc	tgcaactgtg	gagccccttt	ctgggctggc	caaggccaga	2040
gccggctccc	tcagcttgca	gggaggtgtg	gagggagagg	ctcaagcagg	aaccgggggt	2100
gcgcacggcg	cttgccgggc	agctggagtt	ccgggtgggc	gtgggcttgg	cgggccccgc	2160
actcggagca	gcgggccagc	cctgccaggc	cccgggcaat	gagaggctta	gcacccgggc	2220
cagcggctgc	ggagggtgta	ctgggtgccc	cagcagtgcc	agcccgcggg	cgctgtgctc	2280
gctcgatttc	tcactgggcc	ttagcagcct	tcccgcgggg	cagggtctcg	gacctgcagc	2340
ccgccatgcc	tgagcctccc	ctccatgggc	tcctgtgcgg	cccagacctc	cccagcagac	2400
accaccccct	gctccacagc	gcccagtccc	atcgaccacg	caagggtctga	gaagtgcggg	2460
cgcacggcac	cgggactggc	aggcagctac	ccctgcagcc	ctggtgcgga	atccactggg	2520
tgaagccagc	tgggctcctg	agtctgggtg	agacttggag	aacctttatg	tctagctcag	2580
ggatcgtaaa	tacaccaatc	agcaccctgt	gtctagctca	gggtctgtga	atgcaccaat	2640
ccacactctg	tatctagcta	ctctgatggg	gccttggaga	acctttatgt	ctagctcagg	2700

gattgtaa	acaccaatcg	gcactctgta	tctagctcaa	ggtttgtaaa	cacaccaatc	2760
agcaccctgt	gtctagctca	gggtatgtga	atgcaccaat	cgacagtctg	tatctggcta	2820
ctttcatggg	catccgtgtg	aagagaccac	caaacaggct	ttgtgtgagc	aataaagctt	2880
ctatcacctg	ggtgcaggtg	ggctgagtc	gaaaagagag	tcagcgaagg	gagataaggg	2940
tggggccgtt	ttataggatt	tgggtaggta	aaggaaaatt	acagtcaaag	ggggtttgtt	3000
ctctggcggg	caggagtggg	gggtcgcaag	gtgctcagtg	gggggtgctt	ttgagccagg	3060
atgagccagg	aaaaggactt	tcacaaggta	atgtcatcaa	ttaaggcaag	gacccgccat	3120
ttacacctct	tttgtgtgtg	aatgtcatca	gttaagttgg	ggcagggcat	attcacttct	3180
tttgtgattc	ttcagttact	tcaggccatc	tgggcgtata	tgtgcaagtt	acaggggatg	3240
cgatggcctg	gcttgggctc	agaggcttga	cagctactct	ggtggggcct	tggagaatgt	3300
tttgtgtcgac	actctgtatc	tagttaatct	agtggggacg	tggagaacct	tttgtgtctag	3360
ctcagggatt	gtaaacgcac	caatcagcgc	cctgtcaaaa	cagaccactc	ggctctacca	3420
atcagcagga	tgtgggtggg	gccagataag	agaataaaaag	caggctgccc	gagccagcag	3480
tggcaacgcg	cacaggtccc	tatccacaat	atggcagctt	tgttcttttg	ctgtttgcga	3540
taaatcttgc	tactgctcgc	tttttgggtc	cacactgctt	ttatgagctg	taacactcac	3600
cacgaaggtc	tgcagcttca	ctcctgaagc	cactaagacc	acgagcccac	cgggaggaat	3660
gaacaactcc	ggccgcgctg	ccttaagagc	tataacactc	accgcgaagg	tctgcagctt	3720
cactcctcag	ccagcgagac	cacgaaccca	ccagaaggaa	gaaactgcga	acacatctga	3780
acatcagaag	gaacaaactc	cagatgcacc	accttaagag	ctgtaacact	cactgcgagg	3840
gtccgcggct	tccttcttga	agtcagtgag	accaagcact	caccagtttc	ggacacaagc	3900
ccaggagttt	gagatcagcc	tgggcaacat	gatgaaatgc	cctctctgca	aaaaaaaaaa	3960
aaattacaaa	aattggcgga	gcattggtgt	ccgtgcctgt	ggtcccagct	acgcgggagg	4020
ctaaagtggg	aggatcgctt	gagcctggga	ggtgaagact	gcagtgagct	gtgattgtac	4080
cacagccctc	taggctgggg	gacagactga	gaccctgttt	ccctccgca	aaaaaattga	4140
caaaagtgtg	ataagaggtg	cctgatattg	ctaggcgcag	tggctcatgc	ctgtaatccc	4200
agcacttttg	gaagccgagg	cgggcgggtc	acctaaggtc	aggagtgtga	gaccagcctg	4260
gccaacatgg	agaaagccca	tctcttctaa	aaatacaaaa	ttagccggct	gtgggggcag	4320
tgggtggagca	tgcctgtaat	cccagctact	caggaggctg	aggcaggaga	atcacttgaa	4380
cccaggaggc	ggcgggtgca	gtgagccgag	atcgtgccat	tgcactccac	ccactccagc	4440
ctgggcaaca	agagccaaac	tctgtcttaa	aaaaaaaaaa	aaaaagtgcc	tgacatataa	4500
gaggtgtgca	atgcaatagt	tgccaggcaa	catgtttaag	aatgtggagc	tcctgccttc	4560
catggtcctg	ttaaaaaccc	accctcaagg	ccaggtgca	tggctcatgc	ctataatccc	4620
agcacttttg	gaggccgagg	cggttgatc	acctgaggtc	aggagttcga	gaccagcctg	4680
accaccaaca	tgggtgaaatc	ccacctctac	taaaaataca	aaattagatg	agcatggtgg	4740
tgcattgcctg	taatcccacc	tacttgggag	gctgaggcag	gaaaatcact	agaaccaggg	4800
aggcggaggt	tgtagtgagc	cgagatcgtg	ccattgcact	ccagcctgag	caatgagcga	4860
aactccatct	caaaaaaaca	acaacaaaaa	cccactctct	actcccaggg	agctgggtac	4920
agagctgggc	cacatcagtg	caaggtgctg	agccacagag	ctaaggcgga	gctgcaggac	4980
cgcggaaccag	ataacagtgt	gtgagatcag	tgtgtgagat	cagacgtccc	tgccattggt	5040
gaccaccagg	gggcccccaa	gcaccagaga	tggccccatc	cagtcaccac	atccacttct	5100
catccagaga	tgtctgtttc	ttggcacgct	ggggtaaaatt	aggacagaag	gtgacagtct	5160
tgggtgtggt	cagtcaagct	gccccaggca	ggccttgtgg	cctgtagaaa	acgttcaggc	5220
ctaggccggg	cacggtggct	cacgcctgta	atcccagcac	tttgggaggc	cgaggcgggt	5280
ggatcacgag	gtcaggagat	cgtgaccatc	ctggctaaca	cggtgaaacc	ccgtctctac	5340
taaaaataca	aaaaattggc	cgggcatggt	ggcgggcacc	tgtagtcca	gctactcggg	5400
aggctgaggc	aggagaatgg	cgtgaacccg	agaggcagag	tttgagtgga	gccgagatcg	5460
cgccactgca	ctccagcctg	ggcgacagag	caagactcca	tctggaaaag	aaaaagaaaa	5520
cgttcaggtc	tgagccagag	gccagggtg	taattctgtc	acttaccatg	accttgggca	5580
aggcacttcc	ttccctggcc	cagttcacgg	ggttggaatc	gactccaagg	tccttccag	5640
cattaacgct	gcatggttct	aagatgagaa	gatggggcag	tttcccctct	ctcaccaccag	5700
cccgtgtcca	cttcaagggtg	aatgaccagg	gaagtacagt	gtcccaatcc	cgcagttcca	5760
aagcccttgg	ggaccctact	gtcagggtcg	tgcacgagga	ggtgaaggtc	aggtagacca	5820
atcgccctga	agggtcttgc	ctcattcggtg	acagacatcc	ggtttcctct	ggctctaccg	5880
ggattctagg	ggcttttagcc	gaatgagtc	tggggggcgg	gggggtttct	gggggagttc	5940
ccagctaatc	aacttggggac	aggacagcct	ggaactttcg	atggtgccta	tccaagtgtg	6000
gggtgggcac	agcagccaag	acccaatgtc	cttatctcag	gtaggggctc	aggaggtctc	6060
ccagacaggc	agcctccgga	gagtttgggg	gtaggaatgg	gagcaaccag	gcttcttttt	6120
ttctctctta						

WO 99/57291

PCT/US99/09924

- 3 -

<210> 2
 <211> 13
 <212> PRT
 <213> Homo sapiens

<400> 2
 Met Ala Gly Pro Ala Thr Gln Ser Pro Met Lys Leu Met
 1 5 10

<210> 3
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 3
 tatcagcggc tcagcctttg 20

<210> 4
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 4
 ccacctcact caccagcttc tc 22

<210> 5
 <211> 6235
 <212> DNA
 <213> Homo sapiens

<400> 5

gatcacttga	ggacagtagt	tcaagaccag	cctgggacag	atagggagac	tgtctctacg	60
aaaaatcaaa	aaattatggc	cgggcatggg	ggctcacgtc	tgtaatccct	gaactttggg	120
acatcaaggc	aagtggatca	cttgaggtca	ggagttcgag	actagcctgg	ccaacatggg	180
gaaaccctat	ctccactaaa	aaatacaaaa	attagccagg	catgggtggc	ggcacctgta	240
atcccggcta	ctcaggaggc	tgaggcagga	gaatcacttg	aaccaggagg	gcggagggtg	300
cagtgaagctg	agatcacacc	actgcactcc	agcctgggtg	acagagcaag	actctatctc	360
aaaaaaaaata	aaaaaataaa	aaaattagcc	aggcatggta	gtgcacacct	ctagtctcag	420
ctactcagga	ggctgaggtg	ggaggatcac	ttgaacctgg	ggcagtcagg	gctacagtga	480
gccaagatca	tgccactaca	ctccagcctg	ggcaacagag	agagaccctg	tctctaaaaa	540
aataataata	ataaagaaaa	aaacagctct	gtttatgtct	cctgggtccat	acatactact	600
atgtatatag	tttgcaaact	caaagatcca	gatagtcaat	tttttaggct	tgtgggcccgt	660
atgggtctctg	tcacaatcac	tctgcccctgt	ctttctagca	caaaagcagc	tataaacaat	720
acatacatga	atTTTTtata	gacatcgaga	tttgaatttc	atatgatttt	tacattttat	780
aaaataatct	ttttaaaaat	tttcccctaa	ccatttaaaa	gtgtaaaagc	cggccagcgc	840
gccatcgtca	cgcctgtaat	tccagcactt	tgggaggctg	aggtgggcag	atcacttgag	900
atcaacagtt	cgagaccagc	ctggccaaca	tagcaaaacc	ccatttctac	taaaaataaa	960
aaaattagct	gggcatagtg	gtgcacacct	gtgatcccag	ctacttggga	ggctgaggca	1020
ggagaatcgc	ttgaacctgg	gaagcggagg	ttgcagtga	ccaacatcat	gccactgcac	1080
tccagcctgg	gtgacagagt	gagacttcgt	ctcaacgaaa	aaaaaaagtg	taaaagccat	1140
tcctaattca	gtgtacatca	gtgtacatac	tcaggctctgc	gtactcctgc	tctgaggcat	1200
acctgagaag	tagagttgct	tgggtcacagg	acatacacat	ttccacatta	actagacact	1260
accaagttgc	catccaagga	ggTTTTTTTT	ttacaatcta	cactcccccc	agcaacaaat	1320
gagagttact	ccagatcctt	tacaaagatg	ctctaagccc	agtaccagat	gaaaacagga	1380
agtgggaggg	gaagctgcca	gccccttcta	accatgaaga	aatacctggg	agagccttct	1440
ggatgctgga	aggatgaata	acgggggtct	ctggagcctg	ccccctgtca	gatcactgtg	1500
acttctgagc	ctccagtcca	gtctcagccc	catgtgtcat	ggccagtgat	aatgagccct	1560
cactctctgt	ttgggtcttta	ttctccccat	gtggggctga	agtctggatt	gagccgttat	1620
tcaagatgta	cagctttctt	gacaggaaag	tagtgtcaca	gaaacagcag	gggcttggca	1680
agatgatcta	actgcaaata	ctacctggct	cagccaccag	ctagttctgt	gatcttgaac	1740
aagtTTTTtc	acttctctga	ggccatccct	tggctacaac	acaccagttg	gttgacagga	1800
tgaaatgacg	aagtccctta	cacctgtaat	cccagcactt	tgggaggcca	aggcgggtgg	1860
atggccttgag	cctgagaggt	gacagcatgc	cggcagtcct	cacagccctc	gttcgctctc	1920
ggcgccctct	ctgcctgggc	tcccacttcg	gtggcacttg	aggagccctt	cagcccaccg	1980
ctgcactgtg	ggagcccctt	tctgggctgg	ccaaggccag	agccggctcc	ctcagcttgc	2040
agggaggtgt	ggagggagag	gctcaagcag	gaaccggggc	tgcgcacggc	gcttgcgggc	2100
cagctggagt	tccgggtggg	cgtgggcttg	gcggggcccc	cactcggagc	agcggggccag	2160

WO 99/57291

PCT/US99/09924

- 4 -

ccctgccagg	ccccgggcaa	tgagaggctt	agcaccggg	ccagcggctg	cggagggtgt	2220
actgggtgcc	ccagcagtg	cagcccgccg	gcgctgtgct	cgctcgattt	ctcactgggc	2280
cttagcagcc	ttcccgcggg	gcagggtctg	ggacctgcag	cccgccatgc	ctgagcctcc	2340
cctccatggg	ctcctgtgcg	gcccagagct	ccccgacgag	caccaccccc	tgctccacag	2400
cgcccagtc	catcgaccac	gcaagggtctg	agaagtgcgg	gcgcacggca	ccgggactgg	2460
caggcagcta	cccctgcagc	cctgggtgcg	aatccactgg	gtgaagccag	ctgggctcct	2520
gagtctggtg	gagacttgga	gaacctttat	gtctagctca	gggatcgtaa	atacaccaat	2580
cagcacccctg	tgtctagctc	aggggtctgtg	aatgcaccaa	tccacactct	gtatctagct	2640
actctgatgg	ggccttgagg	aacctttatg	tctagctcag	ggattgtaaa	tacaccaatc	2700
ggcactctgt	atctagctca	aggtttgtaa	acacaccaat	cagcacccctg	tgtctagctc	2760
agggatgtgt	aatgcaccaa	tgcacagtct	gtatctggct	actttcatgg	gcacccgtgt	2820
gaagagacca	ccaaacaggc	tttgtgtgag	caataaagct	tctatcacct	gggtgcagg	2880
gggctgagtc	cgaaaagaga	gtcagcgaag	ggagataagg	gtggggccgt	tttataggat	2940
ttgggtaggt	aaaggaaaat	tacagtcaaa	gggggtttgt	tctctggcgg	gcaggagtgg	3000
ggggtcgcaa	ggtgctcagt	gggggtgctt	tttgagccag	gatgagccag	gaaaaggact	3060
ttcacaaggt	aatgtcatca	attaaggcaa	ggaccgcga	tttacacctc	ttttgtggtg	3120
gaatgtcatc	agttaagttg	gggcagggca	tattcacttc	ttttgtgatt	cttcagttac	3180
ttcaggccat	ctgggcgtat	atgtgcaagt	tacaggggat	gcgatggctt	ggcttgggct	3240
cagaggcttg	acagctactc	tggtggggcc	ttggagaatg	tttgtgtcga	cactctgtat	3300
ctagttaatc	tagtggggac	gtggagaacc	tttgtgtcta	gctcagggat	tgtaaacgca	3360
ccaatcagcg	ccctgtcaaa	acagaccact	cggctctacc	aatcagcagg	atgtgggtgg	3420
ggccagataa	gagaataaaa	gcaggctgcc	cgagccagca	gtggcaacgc	gcacaggctc	3480
ctatccacaa	tatggcagct	ttgttctttt	gctgtttgcg	ataaatcttg	ctactgctcg	3540
ctttttgggt	ccacactgct	tttatgagct	gtaacactca	ccacgaagg	ctgcagcttc	3600
actcctgaag	ccactaagac	cacgagccca	ccgggaggaa	tgaacaactc	cggccgcgct	3660
gccttaagag	ctataacact	caccgcgaag	gtctgcagct	tcactcctca	gccagcgaga	3720
ccacgaaccc	accagaagga	agaaactgcg	aacacatctg	aacatcagaa	ggaacaaact	3780
ccagatgcac	caccttaaga	gctgtaacac	tcactgcgag	ggtccgcggc	ttccttcttg	3840
aagtcaagtga	gaccaagcac	tcaccagttt	cggacacaag	cccaggagtt	tgagatcagc	3900
ctgggcaaca	tgatgaaatg	ccctctctgc	aaaaaaaaaa	aaaattacaa	aaattggcgg	3960
agcatgggtg	tccgtgcctg	tggtcccagc	tacgcgggag	gctaaagtgg	gaggatcgct	4020
tgagcctggg	aggtgaagac	tgcaagtgagc	tgtgattgta	ccacagccct	ctaggctggg	4080
ggacagactg	agaccctggt	tcccctccgc	aaaaaaattg	acaaaagtgt	aataagaggt	4140
gcctgatatg	gctaggcgca	gtggctcatg	cctgtaatcc	cagcactttg	ggaagccgag	4200
gcgggcgggt	cacctaaggt	caggagtgtg	agaccagcct	ggccaacatg	gagaaagccc	4260
atctcttcta	aaaatacaaa	attagccggc	tgtgggggca	gtgggtggagc	atgcctgtaa	4320
tcccagctac	tcaggaggct	gaggcaggag	aatcacttga	acccaggagg	cggcgggtgc	4380
agtgaagccga	gatcggtgcca	ttgcactcca	cccactccag	cctgggcaac	aagagccaaa	4440
ctctgtctta	aaaaaaaaaa	aaaaaagtgc	ctgacatata	agaggtgtgc	aatgcaatag	4500
ttgccaggca	acatgtttta	gaatgtggag	ctcctgcctt	ccatggtcct	gttaaaaacc	4560
cacctcaag	gccagggtgca	gtggctcatg	cctataatcc	cagcactttg	ggaggccgag	4620
gcgggtggat	cacctgaggt	caggagtctg	agaccagcct	gaccaccaac	atggtgaaat	4680
cccacctcta	ctaaaaatac	aaaattagat	gagcatgggtg	gtgcatgcct	gtaatcccac	4740
ctacttgga	ggctgaggca	ggaaaatcac	tagaaccagg	gaggcggagg	ttgtagttag	4800
ccgagatcgt	gccattgcac	tccagcctga	gcaatgagcg	aaactccatc	tcaaaaaaac	4860
aacaacaaaa	accactctc	tactcccagg	gagctgggta	cagagctggg	ccacatcagt	4920
gcaagggtgct	gagccacaga	gctaaggcgg	agctgcagga	ccgcggacca	gataacagtg	4980
tgtgagatca	gtgtgtgaga	tcagacgtcc	ctgccattgg	tgaccaccag	ggggcccca	5040
agcaccagag	atggcccat	ccagtcacca	catccacttc	tcacccagag	atgtctgttt	5100
cttggcacgc	tggggtaaat	taggacagaa	ggtgacagtc	ttgggtgtgg	tcagttagac	5160
tgcccaggc	aggccttgtg	gcctgtagaa	aacgttcagg	cctaggccgg	gcacggtggc	5220
tcacgcctgt	aatcccagca	ctttgggagg	ccgaggcggg	tgatcacga	ggtcaggaga	5280
tcgtgaccat	cctggctaac	acggtgaaac	cccgctctcta	ctaaaaatac	aaaaaattgg	5340
ccgggcatgg	tggcggggcac	ctgtagtctc	agctactcgg	gaggctgagg	caggagaatg	5400
gcgtgaaccc	gagaggcaga	gtttgcagtg	agccgagatc	gcgccactgc	actccagcct	5460
gggcgacaga	gcaagactcc	atctggaaaa	gaaaaagaaa	acgttcagg	ctgagccaga	5520
ggcccaggct	gtaattctgt	cacttaccat	gaccttgggc	aaggcacttc	cttccctggc	5580
ccagttcacg	gggttggaat	cgactccaag	gtcccttcca	gcattaacgc	tgcatggttc	5640
taagatgaga	agatggggca	gtttcccttc	tctcacccca	gcccgtgtcc	acttcaaggt	5700
gaatgaccag	ggaagtcacg	tgtcccaatc	ccgcagttcc	aaagcccttg	gggaccctac	5760
tgtcagggtc	gtgcacgagg	aggtgaaggt	caggtgagcc	aatcgccctg	aagggtcttg	5820
cctcattcgg	gacagacatc	cggtttcctc	tggtctacc	gggattctag	gggcttttagc	5880
cgaatgagtc	atggggggcg	gggggggtttc	tgggggagtt	cccagctaat	caacttggga	5940
caggacagcc	tggaaactttc	gatgggtgcct	atccaagtgt	gggggtgggca	cagcagccaa	6000
gacccaatgt	ccttatctca	ggtaggggct	caggaggtct	cccagacagg	cagcctccgg	6060
agagtttggg	ggtaggaatg	ggagcaacca	ggcttctttt	tttctctctt	agaatttggg	6120
ggcttggggg	acaggcttga	gaatcccaaa	ggagaggggc	aaaggacact	ccccacaaag	6180

WO 99/57291

PCT/US99/09924

- 5 -

tctgccagag cgagagaggg agaccccgac tcagctgcca cttccccaca ggcct

6235

<210> 6

<211> 2834

<212> DNA

<213> Homo sapiens

<400> 6

ccggcagtcc	tcacagccct	cgttcgctct	cggcgcctcc	tctgcctggg	ctcccacttc	60
ggtggcactt	gaggagccct	tcagcccacc	gctgcactgt	gggagccctt	ttctgggctg	120
gccaaggcca	gagccggctc	cctcagcttg	cagggaggtg	tggagggaga	ggctcaagca	180
ggaaccgggg	ctgcgcacgg	cgtttgcggg	ccagctggag	ttccgggtgg	gcgtgggctt	240
ggcgggcccc	gcactcggag	cagcggggcca	gccctgccag	gccccgggca	atgagaggct	300
tagcaccggg	gccagcggct	gcggaggggtg	tactgggtgc	cccagcagtg	ccagcccgcc	360
ggcgctgtgc	tcgctcgatt	tctcactggg	ccttagcagc	cttcccgcgg	ggcagggttc	420
gggacctgca	gcccggccatg	cctgagcctc	ccctccatgg	gctcctgtgc	ggcccagacc	480
tccccgacga	gcaccacccc	ctgctccaca	gcgcccagtc	ccatcgacca	cgcaagggtc	540
gagaagtgcg	ggcgcacggc	accgggactg	gcaggcagct	acccctgcag	ccctggtgcg	600
gaatccactg	ggtgaagcca	gctgggctcc	tgagtctggt	ggagacttgg	agaaccttta	660
tgtctagctc	agggatcgta	aatacaccaa	tcagcacctt	gtgtctagct	cagggtctgt	720
gaatgcacca	atccacactc	tgtatctagc	tactctgatg	gggccttggg	gaacctttat	780
gtctagctca	gggattgtaa	atacaccaat	cggcactctg	tatctagctc	aaggtttgta	840
aacacaccaa	tcagcacctt	gtgtctagct	cagggtatgt	gaatgcacca	atcgacagtc	900
tgtatctggc	tactttcatg	ggcatccgtg	tgaagagacc	accaaacagg	ctttgtgtga	960
gcaataaagc	ttctatcacc	tgggtgcagg	tgggctgagt	ccgaaaagag	agtcagcgaa	1020
gggagataag	ggtggggccg	ttttatagga	tttgggtagg	taaaggaaaa	ttacagtcaa	1080
aggggggtttg	ttctctggcg	ggcaggagtg	gggggtcgca	aggtgctcag	tgggggtgct	1140
ttttgagcca	ggatgagcca	ggaaaaggac	tttcacaagg	taatgtcatc	aattaaggca	1200
aggacccgcc	atttacacct	cttttgtggt	ggaatgtcat	cagttaagtt	ggggcagggc	1260
atattcactt	cttttgtgat	tcttcagtta	cttcaggcca	tctgggcgta	tatgtgcaag	1320
ttacagggga	tgcgatggct	tggcttgggc	tcagaggctt	gacagctact	ctgggtggggc	1380
cttgagagaat	gtttgtgtcg	acactctgta	tctagttaat	ctagtgggga	cgtggagaac	1440
ctttgtgtct	agctcaggga	ttgtaaacgc	accaatcagc	gccctgtcaa	aacagaccac	1500
tcggctctac	caatcagcag	gatgtgggtg	gggccagata	agagaataaa	agcaggctgc	1560
ccgagccagc	agtggcaacg	cgcacaggtc	cctatccaca	atatggcagc	tttgttcttt	1620
tgctgtttgc	gataaatctt	gctactgctc	gctttttggg	tccacactgc	ttttatgagc	1680
tgtaacactc	accacgaagg	tctgcagctt	cactcctgaa	gccactaaga	ccacgagccc	1740
accgggagga	atgaacaact	ccggccgcgc	tgccttaaga	gctataacac	tcaccgcgaa	1800
ggtctgcagc	ttcactcctc	agccagcgag	accacgaacc	caccagaagg	aagaaactgc	1860
gaacacatct	gaacatcaga	aggaacaaac	tccagatgca	ccaccttaag	agctgtaaca	1920
ctcactgcga	gggtccgcgg	cttccttctt	gaagtcaagt	agaccaagca	ctcaccagtt	1980
tcggacacaa	gcccaggagt	ttgagatcag	cctgggcaac	atgatgaaat	gccctctctg	2040
caaaaaaaaa	aaaaattaca	aaaattggcg	gagcatggtg	gtccgtgcct	gtgggtcccag	2100
ctacgcggga	ggctaaagtg	ggaggatcgc	ttgagcctgg	gaggtgaaga	ctgcagtgag	2160
ctgtgattgt	accacagccc	tctaggctgg	gggacagact	gagaccctgt	ttccccctccg	2220
caaaaaaatt	gacaaaagtg	taataagagg	tgcctgatat	ggctaggcgc	agtggctcat	2280
gcctgtaatc	ccagcacttt	gggaagccga	ggcgggcggg	tcacctaaagg	tcaggagtgt	2340
gagaccagcc	tggccaacat	ggagaaagcc	catctcttct	aaaaatacaa	aattagccgg	2400
ctgtgggggc	agtgggtggag	catgcctgta	atcccagcta	ctcaggaggc	tgaggcagga	2460
gaatcacttg	aaccagaggag	gcggcggttg	cagtgcagccg	agatcgtgcc	attgcactcc	2520
accactcca	gcctgggcaa	caagagccaa	actctgtctt	aaaaaaaaaa	aaaaaaagtg	2580
cctgacatat	aagagggtgtg	caatgcaata	gttgccaggc	aacatgttta	agaatgtgga	2640
gctcctgcct	tccatgggtcc	tgttaaaaac	ccaccctcaa	ggccagggtgc	agtggctcat	2700
gcctataatc	ccagcacttt	gggaggccga	ggcgggtgga	tcacctgagg	tcaggagttc	2760
gagaccagcc	tgaccaccaa	catggtgaaa	tcccacctct	actaaaaata	caaaattaga	2820
tgagcatggt	ggtg					2834

<210> 7

<211> 1252

<212> DNA

<213> Homo sapiens

<400> 7

cctgtaatcc	cacctacttg	ggaggctgag	gcaggaaaat	cactagaacc	agggaggcgg	60
aggttgtagt	gagccgagat	cgtgccattg	cactccagcc	tgagcaatga	gcgaaactcc	120
atctcaaaaa	aacaacaaca	aaaaccctct	ctctactccc	agggagctgg	gtacagagct	180
gggccacatc	agtgcagggt	gctgagccac	agagctaagg	cggagctgca	ggaccgcgga	240

WO 99/57291

PCT/US99/09924

- 6 -

ccagataaca	gtgtgtgaga	tcagtgtgtg	agatcagacg	tccttgccat	tggtgaccac	300
cagggggccc	ccaagcacca	gagatggccc	catccagtca	ccacatccac	ttctcatcca	360
gagatgtctg	tttcttgcca	cgctggggta	aattaggaca	gaagggtgaca	gtcttgggtg	420
tggtcagtca	gactgcccc	ggcaggcctt	gtggcctgta	gaaaacgttc	aggcctaggc	480
cgggcacggt	ggctcacgcc	tgtaatccca	gcactttggg	aggccgaggc	gggtggatca	540
cgaggtcagg	agatcgtgac	catcctggct	aacacggtga	aaccccgctct	ctactaaaaa	600
tacaaaaaat	tggccgggca	tggtggcggg	cacctgtagt	tccagctact	cgggaggctg	660
aggcaggaga	atggcgtgaa	cccagagaggc	agagtgttga	gtgagccgag	atcgcgccac	720
tgcactccag	cctgggcgac	agagcaagac	tccatctgga	aaagaaaaag	aaaacgttca	780
ggtctgagcc	agaggcccag	gctgtaatc	tgctacttac	catgaccttg	ggcaaggcac	840
ttccttccct	ggcccagttc	acgggggttg	aatcgactcc	aagggtccctt	ccagcattaa	900
cgctgcatgg	ttctaagatg	agaagatggg	gcagtttccc	ctctctcacc	ccagcccgtg	960
tccacttcaa	ggtgaatgac	cagggaagtc	acgtgtccca	atcccgcagt	tccaaagccc	1020
ttggggaccc	tactgtcagg	gtcgtgcacg	aggaggtgaa	ggtcaggtga	gccaatcgcc	1080
tcgaagggtc	ttgcctcatt	cgggacagac	atccggtttc	ctctggctct	accgggattc	1140
taggggcttt	agccgaatga	gtcatggggg	gcgggggggt	ttctggggga	gttcccagct	1200
aatcaacttg	ggacaggaca	gcctggaact	ttcgatggtg	cctatccaag	tg	1252

<210> 8

<211> 14

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(14)

<223> n = A,T,C or G

<400> 8

YYYYYYYYYY nyag

14

- 33 -

1. A DNA construct that alters expression of an endogenous G-CSF gene in a mammalian cell upon integration into the genome of the cell via homologous recombination, the construct comprising (i) a targeting sequence containing at least 20 contiguous nucleotides from SEQ ID NO:5 and (ii) a transcriptional regulatory sequence.
2. The DNA construct of claim 1, wherein the construct further comprises an exon and a splice-donor site.
3. The DNA construct of claim 2, wherein the construct further comprises, downstream from the splice-donor site, an intron and a splice-acceptor site.
4. The DNA construct of claim 1, wherein the construct further comprises a selectable marker gene.
5. The DNA construct of claim 1, wherein the targeting sequence contains at least 50 contiguous nucleotides from SEQ ID NO:5.
6. An isolated nucleic acid molecule comprising at least 100 contiguous nucleotides of SEQ ID NO:5 or its complement, wherein the isolated nucleic acid molecule does not encode full-length G-CSF.

- 34 -

7. The isolated nucleic acid molecule of claim 6, wherein the isolated nucleic acid molecule comprises at least 200 contiguous nucleotides of SEQ ID NO:5 or its complement.

8. The isolated nucleic acid molecule of claim 6, wherein the isolated nucleic acid molecule comprises at least 500 contiguous nucleotides of SEQ ID NO:5 or its complement.

9. The isolated nucleic acid molecule of claim 6, wherein the isolated nucleic acid molecule comprises nucleotides 1470 to 4723 of SEQ ID NO:5, or its complement.

10. The isolated nucleic acid molecule of claim 6, wherein the isolated nucleic acid molecule comprises SEQ ID NO:5 or its complement.

11. An isolated nucleic acid molecule comprising a strand that comprises a nucleotide sequence that (i) is at least 1000 nucleotides in length, and (ii) shares at least 80% sequence identity with a fragment of SEQ ID NO:5 having the same length as the nucleotide sequence.

- 35 -

12. A homologously recombinant cell stably transfected with the DNA construct of claim 1, the DNA construct having undergone homologous recombination with genomic DNA upstream of the ATG initiation codon of an endogenous G-CSF coding sequence.

13. A homologously recombinant cell stably transfected with the DNA construct of claim 2, the DNA construct having undergone homologous recombination with genomic DNA upstream of the ATG initiation codon of an endogenous G-CSF coding sequence.

14. A homologously recombinant cell stably transfected with the DNA construct of claim 3, the DNA construct having undergone homologous recombination with genomic DNA upstream of the ATG initiation codon of an endogenous G-CSF coding sequence.

15. A homologously recombinant cell stably transfected with the DNA construct of claim 4, the DNA construct having undergone homologous recombination with genomic DNA upstream of the ATG initiation codon of an endogenous G-CSF coding sequence.

- 36 -

16. A method of altering expression of an endogenous G-CSF gene in a mammalian cell, the method comprising

- introducing the DNA construct of claim 1 into the cell;
- maintaining the cell under conditions which permit homologous recombination to occur between the construct and a genomic target site homologous to the targeting sequence, to produce a homologously recombinant cell; and
- maintaining the homologously recombinant cell under conditions which permit expression of the G-CSF coding sequence under the control of the transcriptional regulatory sequence.

17. A method of altering expression of an endogenous G-CSF gene in a mammalian cell, the method comprising

- introducing the DNA construct of claim 4 into the cell;
- maintaining the cell under conditions which permit homologous recombination to occur between the construct and a genomic target site homologous to the targeting sequence, to produce a homologously recombinant cell; and
- maintaining the homologously recombinant cell under conditions which permit expression of the G-CSF coding sequence under the control of the transcriptional regulatory sequence.

- 37 -

18. Use of the cell of claim 12 for the manufacture of a medicament for the delivery of G-CSF to an animal by implantation of the cell into the animal.

19. Use of the cell of claim 13 for the manufacture of a medicament for the delivery of G-CSF to an animal by implantation of the cell into the animal.

20. Use of the cell of claim 14 for the manufacture of a medicament for the delivery of G-CSF to an animal by implantation of the cell into the animal.

21. Use of the cell of claim 15 for the manufacture of a medicament for the delivery of G-CSF to an animal by implantation of the cell into the animal.

22. A method of producing G-CSF, comprising
providing the cell of claim 12, and
culturing the cell *in vitro* under conditions which permit the cell to express and secrete G-CSF.

23. A method of producing G-CSF, comprising
providing the cell of claim 13, and
culturing the cell *in vitro* under conditions which permit the cell to express and

- 38 -

secrete G-CSF.

24. A method of producing G-CSF, comprising
providing the cell of claim 14, and
culturing the cell *in vitro* under conditions which permit the cell to express and
secrete G-CSF.

25. A method of producing G-CSF, comprising
providing the cell of claim 15, and
culturing the cell *in vitro* under conditions which permit the cell to express and
secrete G-CSF.

WO 99/57291

PCT/US99/09924

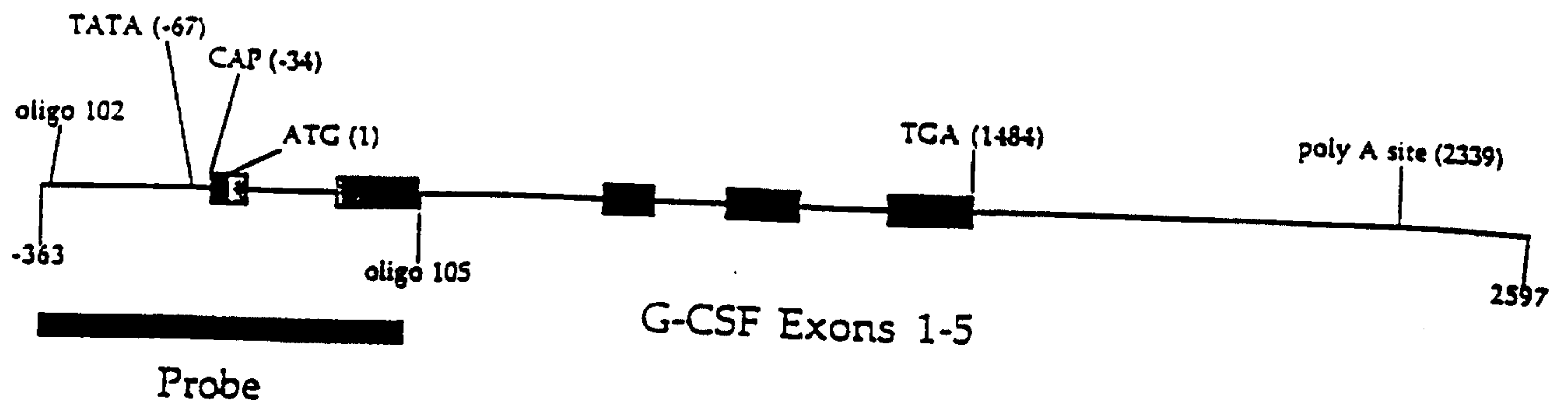


Figure 1

WO 99/57291

PCT/US99/09924

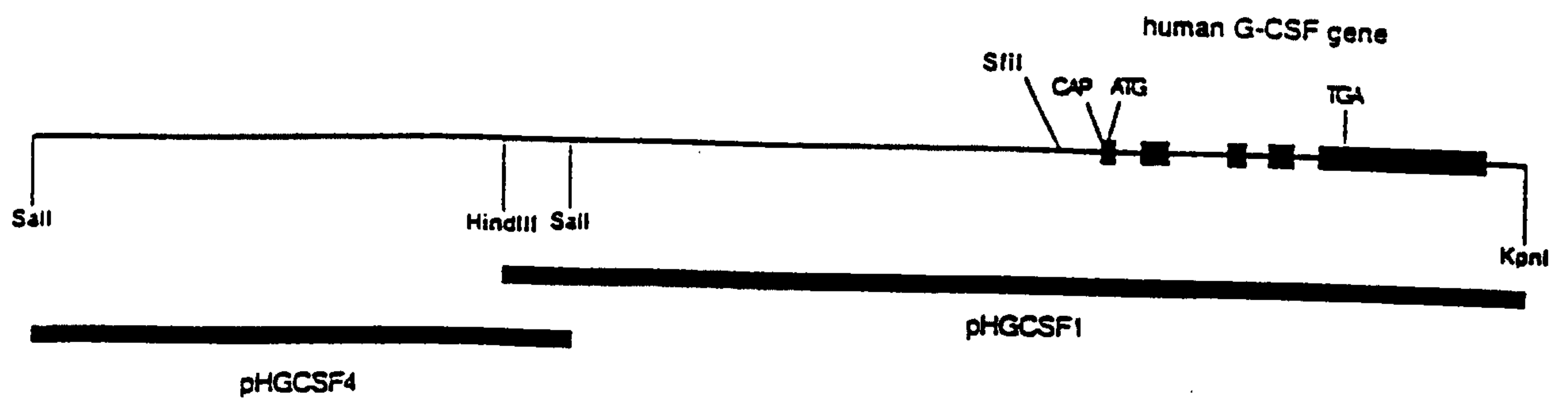


Figure 2

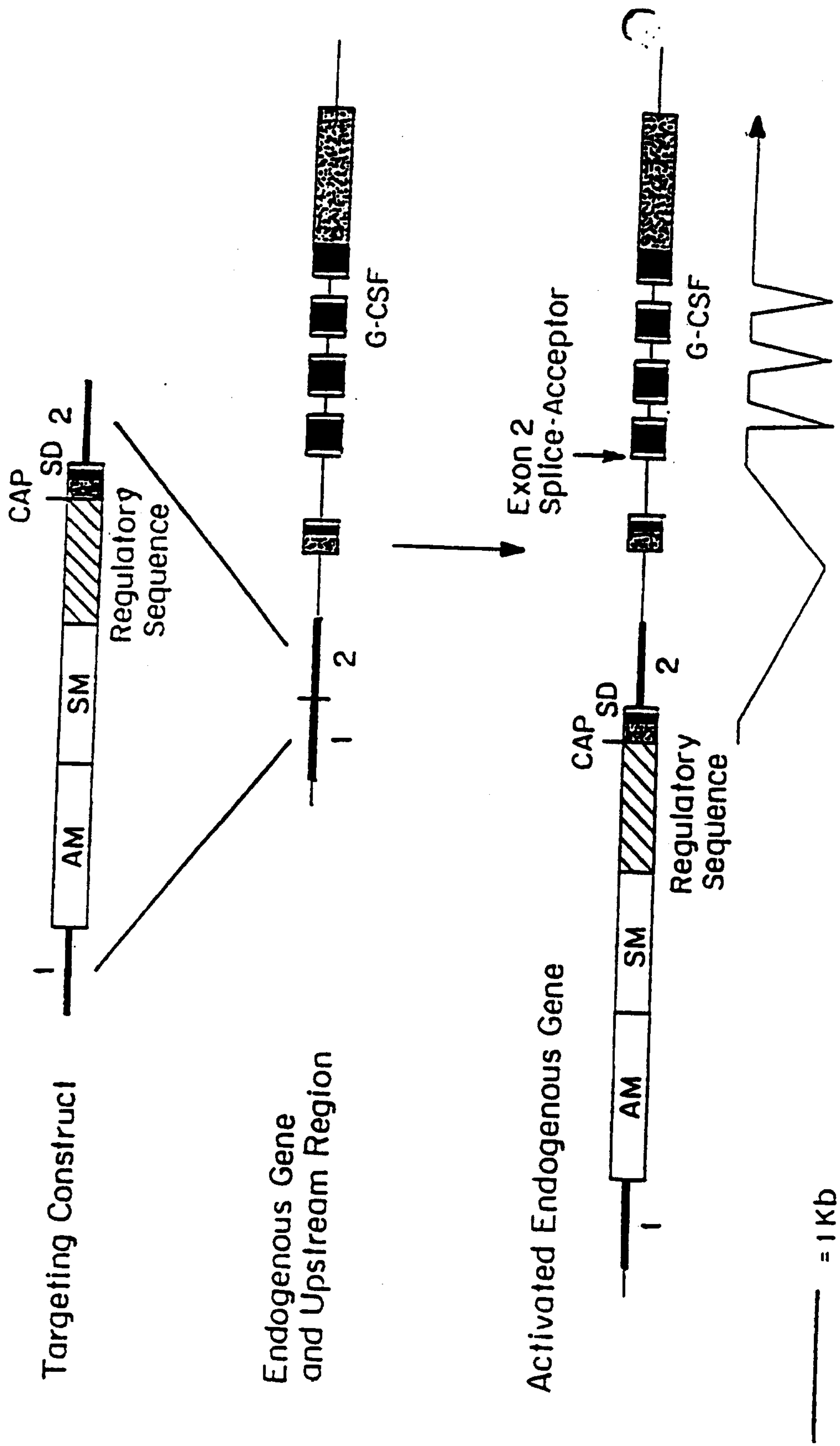


FIG. 4

Fig. 5

GATCACTTGAGGACAGTAGTTCAAGACCAGCCTGGGCAGCATAGGGAGACTGTCTCTACGAAAAA
TCAAAAAATTATGGCCGGGCATGGTGGCTCACGTCTGTAAATCCCTGAACTTTGGGACATCAAGGC
AAGTGGATCACTTGAGGTCAGGAGTTCGAGACTAGCCTGGCCAACATGGTGAAACCCTATCTCCA
CTAAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGCACCTGTAATCCCGGCTACTCAGGAGGC
TGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCACTGC
ACTCCAGCCTGGGTGACAGAGCAAGACTCTATCTCAAAAAAATAAAAAAATAAAAAAATTAGCC
AGGCATGGTAGTGCACACCTCTAGTCTCAGCTACTCAGGAGGCTGAGGTGGGAGGATCACTTGAA
CCTGGGGCAGTCAAGGCTACAGTGAGCCAAGATCATGCCACTACACTCCAGCCTGGGCAACAGAG
AGAGACCCTGTCTCTAAAAAATAATAATAATAAGAAAAAAACAGCTCTGTTTATGTCTCCTGG
TCCATACATACTACTATGTATATAGTTTGCAAACTCAAAGATCCAGATAGTCAATTTTTTTAGGCT
TGTGGGCGGTATGGTCTCTGTCAATCACTCTGCCCTGTCTTTCTAGCACAAAAGCAGCTATAA
ACAATACATACATGAATTTTTTATAGACATCGAGATTTGAATTTTCATATGATTTTTTACATTTTTAT
AAAATAATCTTTTTTAAAAATTTTCCCCTAACCATTTAAAAGTGTAAGCCGGCCAGCGCGCCAT
CGTCACGCCTGTAATTCAGCACTTTGGGAGGCTGAGGTGGGCAGATCACTTGAGATCAACAGTT
CGAGACCAGCCTGGCCAACATAGCAAAACCCCATTTCTACTAAAAATAAAAAAATTAGCTGGGCA
TAGTGGTGCACACCTGTGATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGG
GAAGCGGAGGTTGCAGTGAGCCAACATCATGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGAC
TTCGTCTCAACGAAAAAAGTGTAAGCCATTCTTAATTCAGTGATACATCAGTGATACATAC
TCAGGTCTGCGTACTCCTGCTCTGAGGCATACCTGAGAAGTAGAGTTGCTTGGTTCACAGGACATA
CACATTTCCACATTAAGTAGACACTACCAAGTTGCCATCCAAGGAGGTTTTTTTTTTTACAATCTA
CACTCCCCCAGCAACAAATGAGAGTTACTCCAGATCCTTTACAAAGATGCTCTAAGCCCAGTAC
CAGATGAAAACAGGAAGTGGGAGGGGAAGCTGCCAGCCCCCTTCTAACCATGAAGAAATACCTGGT
AGAGCCTTCTGGATGCTGGAAGGATGAATAACGGGGGTCTCTGGAGCCTGCCCCCTGTGAGATCA
CTGTGACTTCTGAGCCTCCAGTCCAGTCTCAGCCCCATGTGTGATGGCCAGTGATAATGAGCCCT
CACTCTCTGTTTGGTCTTTATTCTCCCCATGTGGGGCTGAAGTCTGGATTGAGCCGTTATTCAAG
ATGTACAGCTTTCTTGACAGGAAAGTAGTGTCACAGAAACAGCAGGGGCTTGGCAAGATGATCTA
ACTGCAAATCCTACCTGGCTCAGCCACCAGCTAGTTCTGTGATCTTGAACAAGTTTTTTTCACTTC
TCTGAGGCCATCCCTTGGCTACAACACACCAGTTGGTTGACAGGATGAAATGACGAAGTCCCTTA
CACCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGATGGCTTGAGCCTGAGAGGTGACAG
CATGCCGGCAGTCCCTCACAGCCCTCGTTCTCGCTCTCGGCGCCTCCTCTGCCTGGGCTCCCCTTCG
GTGGCACTTGAGGAGCCCTTCAGCCCACCGCTGCACTGTGGGAGCCCTTTCTGGGCTGGCCAAG
GCCAGAGCCGGCTCCCTCAGCTTGCAGGGAGGTGTGGAGGGAGAGGCTCAAGCAGGAACCGGGGC
TGCGCACGGCGCTTGCGGGCCAGCTGGAGTTCCGGGTGGGCGTGGGCTTGGCGGGGCCCCGCACTC
GGAGCAGCGGGCCAGCCCTGCCAGGCCCCGGGCAATGAGAGGCTTAGCACCCGGGCCAGCGGCTG
CGGAGGGTGTACTGGGTGCCCCAGCAGTGCCAGCCCGCGGCGCTGTGCTCGCTCGATTCTCAC
TGGGCCTTAGCAGCCTTCCCGCGGGGCAGGGCTCGGGACCTGCAGCCCGCCATGCCTGAGCCTCC
CCTCCATGGGCTCCTGTGCGGGCCGAGCCTCCCCGACGAGCACCACCCCTGCTCCACAGCGCCC
AGTCCCATCGACCACGCAAGGGCTGAGAAGTGCGGGCGCACGGCACCAGGACTGGCAGGCAGCTA
CCCCTGCAGCCCTGGTGCGGAATCCAAGTGGGTGAAGCCAGCTGGGCTCCTGAGTCTGGTGGAGAC
TTGGAGAACCCTTTATGTCTAGCTCAGGGATCGTAAATACACCAATCAGCACCCCTGTGTCTAGCTC
AGGGTCTGTGAATGCACCAATCCACACTCTGTATCTAGCTACTCTGATGGGGCCTTGGAGAACCCT
TTATGTCTAGCTCAGGGATTGTAAATACACCAATCGGCACTCTGTATCTAGCTCAAGGTTTGTAA

WO 99/57291

PCT/US99/09924

Fig. 5 (Continued)

ACACACCAATCAGCACCCCTGTGTCTAGCTCAGGGTATGTGAATGCACCAATCGACAGTCTGTATC
 TGGCTACTTTTCATGGGCATCCGTGTGAAGAGACCACCAAACAGGCTTTGTGTGAGCAATAAAGCT
 TCTATCACCTGGGTGCAGGTGGGCTGAGTCCGAAAAGAGAGTCAGCGAAGGGGAGATAAGGGTGGG
 GCCGTTTTATAGGATTTGGGTAGGTAAAGGAAAATTACAGTCAAAGGGGGTGTGTTCTCTGGCGG
 GCAGGAGTGGGGGGTCGCAAGGTGCTCAGTGGGGGTGCTTTTTTGAGCCAGGATGAGCCAGGAAAA
 GGACTTTTACAAGGTAATGTCATCAATTAAGGCAAGGACCCGCCATTTACACCTCTTTTGTGGTG
 GAATGTCATCAGTTAAGTTGGGGCAGGGCATATTCACTTCTTTTGTGATTCTTCAGTTACTTCAG
 GCCATCTGGGCGTATATGTGCAAGTTACAGGGGATGCGATGGCTTGGCTTGGGCTCAGAGGCTTG
 ACAGCTACTCTGGTGGGGCCTTGGAGAATGTTTGTGTGCGACACTCTGTATCTAGTTAATCTAGTG
 GGGACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCAA
 ACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGG
 CTGCCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAGCTTTGTTCTTTT
 GCTGTTTGGGATAAATCTTGCTACTGCTCGCTTTTTTGGGTCCACACTGCTTTTATGAGCTGTAAC
 ACTCACCAACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAGACCACGAGCCCACCGGGAGGAA
 TGAACAACTCCGGCCGCGCTGCCTTAAGAGCTATAACACTCACCGCGAAGGTCTGCAGCTTCACT
 CCTCAGCCAGCGAGACCACGAACCCACCAGAAGGAAGAACTGCGAACACATCTGAACATCAGAA
 GGAACAACTCCAGATGCACCACCTTAAGAGCTGTAACTCACTGCGAGGGTCCGCGGCTTCCT
 TCTTGAAGTCAGTGAGACCAAGCACTCACCAAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGC
 CTGGGCAACATGATGAAATGCCCTCTCTGCAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCAT
 GGTGGTCCGTGCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCTGGG
 AGGTGAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACTGAGACC
 CTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTGATATGGCTAGGCGCA
 GTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGGCGGGTCACCTAAGGTCAGGA
 GTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTCTTCTAAAAATACAAAATTAGCCGGC
 TGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA
 CTTGAACCCAGGAGGCGGCGGTTGCAGTGAGCCGAGATCGTGCCATTGCACTCCACCCACTCCAG
 CCTGGGCAACAAGAGCCAACTCTGTCTTAAAAAAAAAAAAAAAAAAGTGCCTGACATATAAGAGG
 TGTGCAATGCAATAGTTGCCAGGCAACATGTTTAAGAATGTGGAGCTCCTGCCTTCCATGGTCCT
 GTTAAAAACCCACCCCTCAAGGCCAGGTGCAGTGGCTCATGCCTATAATCCCAGCACTTTGGGAGG
 CCGAGGCGGGTGGATCACCTGAGGTCAGGAGTTTCGAGACCAGCCTGACCACCAACATGGTGAAAT
 CCCACCTCTACTAAAAATACAAAATTAGATGAGCATGGTGGTGCATGCCTGTAATCCCACCTACT
 TGGGAGGCTGAGGCAGGAAATCACTAGAACCAGGGAGGCGGAGGTTGTAGTGAGCCGAGATCGT
 GCCATTGCACTCCAGCCTGAGCAATGAGCGAACTCCATCTCAAAAAACAACAACAAAAACCCA
 CTCTCTACTCCCAGGGAGCTGGGTACAGAGCTGGGCCACATCAGTGCAAGGTGCTGAGCCACAGA
 GCTAAGGCGGAGCTGCAGGACCGCGGACCAGATAACAGTGTGTGAGATCAGTGTGTGAGATCAGA
 CGTCCCTGCCATTGGTGAACACCAGGGGGCCCCCAAGCACCAGAGATGGCCCCATCCAGTCACCA
 CATCCACTTCTCATCCAGAGATGTCTGTTTCTTGGCACGCTGGGGTAAATTAGGACAGAAGGTGA
 CAGTCTTGGGTGTGGTCAGTCAGACTGCCCCAGGCAGGCCTTGTGGCCTGTAGAAAACGTTTCAGG
 CCTAGGCCGGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGAT
 CACGAGGTCAGGAGATCGTGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATAC
 AAAAAATTGGCCGGGCATGGTGGCGGGCACCTGTAGTTCCAGCTACTCGGGAGGCTGAGGCAGGA
 GAATGGCGTGAACCCGAGAGGCAGAGTTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCT
 GGGCGACAGAGCAAGACTCCATCTGGAAAAGAAAAGAAAACGTTTCAGGTCTGAGCCAGAGGCC
 AGGCTGTAATTCTGTCACTTACCATGACCTTGGGCAAGGCACTTCTTCCCTGGCCCCAGTTTCAGG
 GGGTTGGAATCGACTCCAAGGTCCCTTCCAGCATTAACGCTGCATGGTTCTAAGATGAGAAGATG
 GGGCAGTTTCCCCTCTCTCACCCAGCCCGTGTCCACTTCAAGGTGAATGACCAGGGAAGTCAAG
 TGTCCCAATCCCGCAGTTCCAAAGCCCTTGGGGACCCCTACTGTGAGGGTCGTGCACGAGGAGGTG
 AAGGTGAGGTGAGCCAATCGCCTCGAAGGGTCTTGCCTCATTCGGGACAGACATCCGGTTTCCTC
 TGGCTCTACCGGGATTCTAGGGGCTTTAGCCGAATGAGTCATGGGGGGCGGGGGGGTTTCTGGGG
 GAGTTCCCAGCTAATCAACTTGGGACAGGACAGCCTGGAACCTTCGATGGTGCCTATCCAAGTGT
 GGGGTGGGCACAGCAGCCAAGACCCAATGTCTTATCTCAGGTAGGGGCTCAGGAGGTCTCCAG
 ACAGGCAGCCTCCGGAGAGTTTGGGGGTAGGAATGGGAGCAACCAGGCTTCTTTTTTCTCTCTT
 AGAATTTGGGGGCTTGGGGGACAGGCTTGAGAAATCCCAAAGGAGAGGGGGCAAAGGACACTCCCC
 ACAAGTCTGCCAGAGCGAGAGAGGGAGACCCCGACTCAGCTGCCACTTCCCCACAGGCCT

Fig. 6

CC GGCAGTCCTC
ACAGCCCTCG TTCGCTCTCG GCGCCTCCTC TGCCTGGGCT CCCACTTCGG TGGCACTTGA
GGAGCCCTTC AGCCACCGC TGCAGTGTGG GAGCCCTTT CTGGGCTGGC CAAGGCCAGA
GCCGGCTCCC TCAGCTTGCA GGGAGGTGTG GAGGGAGAGG CTCAAGCAGG AACCGGGGCT
GCGCACGGCG CTTGCGGGCC AGCTGGAGTT CCGGGTGGGC GTGGGCTTGG CGGGCCCCGG
ACTCGGAGCA GCGGGCCAGC CCTGCCAGGC CCCGGGCAAT GAGAGGCTTA GCACCCGGGC
CAGCGGCTGC GGAGGGTGTA CTGGGTGCCC CAGCAGTGCC AGCCCGCCGG CGCTGTGCTC
GCTCGATTTC TCACTGGGCC TTAGCAGCCT TCCCGCGGGG CAGGGCTCGG GACCTGCAGC
CCGCCATGCC TGAGCCTCCC CTCCATGGGC TCCTGTGCGG CCCGAGCCTC CCCGACGAGC
ACCACCCCTT GCTCCACAGC GCCCAGTCCC ATCGACCACG CAAGGGCTGA GAAGTGCGGG
CGCACGGCAC CGGGACTGGC AGGCAGCTAC CCCTGCAGCC CTGGTGCGGA ATCCACTGGG
TGAAGCCAGC TGGGCTCCTG AGTCTGGTGG AGACTTGAG AACCTTTATG TCTAGCTCAG
GGATCGTAAA TACACCAATC AGCACCTGT GTCTAGCTCA GGGTCTGTGA ATGCACCAAT
CCCACTCTG TATCTAGCTA CTCTGATGGG GCCTTGAGGA ACCTTTATGT CTAGCTCAGG
GATTGTAAAT ACACCAATCG GCACTCTGTA TCTAGCTCAA GGTTTGTAAT CACACCAATC
AGCACCTGT GTCTAGCTCA GGGTATGTGA ATGCACCAAT CGACAGTCTG TATCTGGCTA
CTTTCATGGG CATCCGTGTG AAGAGACCAC CAAACAGGCT TTGTGTGAGC AATAAAGCTT
CTATCACCTG GGTGCAGGTG GGCTGAGTCC GAAAAGAGAG TCAGCGAAGG GAGATAAGGG
TGGGGCCGTT TTATAGGATT TGGGTAGGTA AAGGAAAATT ACAGTCAAAG GGGGTTTGTT
CTCTGGCGGG CAGGAGTGGG GGGTCGCAAG GTGCTCAGTG GGGGTGCTTT TTGAGCCAGG
ATGAGCCAGG AAAAGGACTT TCACAAGGTA ATGTCATCAA TTAAGGCAAG GACCCGCCAT
TTACACCTCT TTTGTGGTGG AATGTCATCA GTTAAGTTGG GGCAGGGCAT ATCACTTCT
TTTGTGATTC TTCAGTTACT TCAGGCCATC TGGGCGTATA TGTGCAAGTT ACAGGGGATG
CGATGGCTTG GCTTGGGCTC AGAGGCTTGA CAGCTACTCT GGTGGGGCCT TGGAGAATGT
Sall
TTGTGTGAC ACTCTGTATC TAGTTAATCT AGTGGGGACG TGGAGAACCT TTGTGTCTAG
CTCAGGGATT GTAAACGCAC CAATCAGCGC CCTGTCAAAA CAGACCACTC GGCTCTACCA
ATCAGCAGGA TGTGGGTGGG GCCAGATAAG AGAATAAAAG CAGGCTGCCC GAGCCAGCAG
TGGCAACGCG CACAGGTCCC TATCCACAAT ATGGCAGCTT TGTTCTTTTG CTGTTTGCGA
TAAATCTTGC TACTGCTCGC TTTTGGGTC CACTGCTT TTATGAGCTG TAACACTCAO
CACGAAGGTC TGCAGCTTCA CTCCTGAAGC CACTAAGACC ACGAGCCAC CGGGAGGAAT
GAACAACCTC GGCCGCGCTG CETTAAGAGC TATAACACTC ACCGCGAAGG TCTGCAGCTT

Fig. 6 (Continued)

CACTCCTCAG	CCAGCGAGAC	CACGAACCCA	CCAGAAGGAA	GAAACTGCGA	ACACATCTGA
ACATCAGAAG	GAACAAACTC	CAGATGCACC	ACCTTAAGAG	CTGTAACACT	CACTGCGAGG
GTCCGCGGCT	TCCTTCTTGA	AGTCAGTGAG	ACCAAGCACT	CACCAGTTTC	GGACACAAGC
CCAGGAGTTT	GAGATCAGCC	TGGGCAACAT	GATGAAATGC	CCTCTCTGCA	AAAAAAAAAA
AAATTACAAA	AATTGGCGGA	GCATGGTGGT	CCGTGCCTGT	GGTCCCAGCT	ACGCGGGAGG
CTAAAGTGGG	AGGATCGCTT	GAGCCTGGGA	GGTGAAGACT	GCAGTGAGCT	GTGATTGTAC
CACAGCCCTC	TAGGCTGGGG	GACAGACTGA	GACCCTGTTT	CCCCCTCCGA	AAAAAATTGA
CAAAAGTGTA	ATAAGAGGTG	CCTGATATGG	CTAGGCGCAG	TGGCTCATGC	CTGTAATCCG
AGCACTTTGG	GAAGCCGAGG	CGGGCGGGTC	ACCTAAGGTC	AGGAGTGTGA	GACCAGCCTG
GCCAACATGG	AGAAAGCCCA	TCTCTTCTAA	AAATACAAAA	TTAGCCGGCT	GTGGGGGCAG
TGGTGGAGCA	TGCCTGTAAT	CCCAGCTACT	CAGGAGGCTG	AGGCAGGAGA	ATCACTTGAA
CCCAGGAGGC	GGCGGTGCA	GTGAGCCGAG	ATCGTGCCAT	TGCACTCCAC	CCACTCCAGC
CTGGGCAACA	AGAGCCAAAC	TCTGTCTTAA	AAAAAAAAAA	AAAAAGTGCC	TGACATATAA
GAGGTGTGCA	ATGCAATAGT	TGCCAGGCAA	CATGTTTAAG	AATGTGGAGC	TCCTGCCTTC
CATGGTCCTG	TTAAAAACCC	ACCCTCAAGG	CCAGGTGCAG	TGGCTCATGC	CTATAATCCC
AGCACTTTGG	GAGGCCGAGG	CGGGTGGATC	ACCTGAGGTC	AGGAGTTCGA	GACCAGCCTG
ACCACCAACA	TGGTGAAATC	CCACCTCTAC	TAAAAATACA	AAATTAGATG	AGCATGGTGG

TG

Fig. 7

CCTG TAATCCCACC TACTTGGGAG GCTGAGGCAG GAAAATCACT AGAACCAGGG
AGGCGGAGGT TGTAGTGAGC CGAGATCGTG CCATTGCACT CCAGCCTGAG CAATGAGCGA
AACTCCATCT CAAAAAACA ACAACAAAA CCCACTCTCT ACTCCCAGGG AGCTGGGTAC
AGAGCTGGGC CACATCAGTG CAAGGTGCTG AGCCACAGAG CTAAGGCGGA GCTGCAGGAC
CGCGGACCAG ATAACAGTGT GTGAGATCAG TGTGTGAGAT CAGACGTCCC TGCCATTGGT
GACCACCAGG GGGCCCCCAA GCACCAGAGA TGGCCCCATC CAGTCACCAC ATCCACTTCT
CATCCAGAGA TGTCTGTTTC TTGGCACGCT GGGGTAAATT AGGACAGAAG GTGACAGTCT
-1457 TGGGTGTGGT CAGTCAGACT GCCCCAGGCA GGCCTTGTGG CCTGTAGAAA ACGTTCAGGC
-1397 CTAGGCCGGG CACGGTGGCT CACGCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGCGGGT
-1337 GGATCACGAG GTCAGGAGAT CGTGACCATC CTGGCTAACA CGGTGAAACC CCGTCTCTAC
-1277 TAAAAATACA AAAAATTGGC CGGGCATGGT GCGGGGCACC TGTAAGTTCCA GCTACTCGGG
-1217 AGGCTGAGGC AGGAGAATGG CGTGAACCCG AGAGGCAGAG TTTGCAGTGA GCCGAGATCG
-1157 CGCCACTGCA CTCCAGCCTG GCGGACAGAG CAAGACTCCA TCTGGAAAAG AAAAAGAAAA
-1097 CGTTCAGGTC TGAGCCAGAG GCCCAGGCTG TAATTCTGTC ACTTACCATG ACCTTGGGCA
-1037 AGGCACTTCC TTCCCTGGCC CAGTTCACGG GGTGGAATC GACTCCAAGG TCCCTTCCAG
-977 CATTAACGCT GCATGGTTCT AAGATGAGAA GATGGGGCAG TTTCCCTCT CTCACCCAG
-917 CCCGTGTCCA CTTCAAGGTG AATGACCAGG GAAGTCACGT GTCCCAATCC CGCAGTTCCA
-857 AAGCCCTTGG GGACCCTACT GTCAGGGTCG TGCACGAGGA GGTGAAGGTC AGGTGAGCCA
-797 ATCGCCTCGA AGGGTCTTGC CTCATTGGG ACAGACATCC GGTTTCTCT GGCTCTACCG
-737 GGATTCTAGG GGCTTTAGCC GAATGAGTCA TGGGGGGCGG GGGGGTTTCT GGGGGAGTTC
-677 CCAGCTAATC AACTTGGGAC AGGACAGCCT GGAACTTTCG ATGGTGCCTA TCCAAGTG

Xcml

