INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| (51) International Patent Classification 4: | C07K 13/00, C12P 21/02 C12N 16/00, A61K 37/02 |
| A2 | A2 |
| (11) International Publication Number: | WO 89/00582 |
| (43) International Publication Date: | 26 January 1989 (26.01.89) |
| (21) International Application Number: | PCT/US88/02334 |
| (22) International Filing Date: | 15 July 1988 (15.07.88) |
| (31) Priority Application Number: | 074,988 |
| (32) Priority Date: | 17 July 1987 (17.07.87) |
| (33) Priority Country: | US |
| (71) Applicant: | SCHERING BIOTECH CORPORATION [US/US]; 901 California Avenue, Palo Alto, CA 94304 (US). |
| (81) Designated States: | AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), DK, FR (European patent), GB (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent). |

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN GRANULOCYTE-MACROPHAGE COLONY STIMULATING FACTOR AND MUTEINS THEREOF

(57) Abstract

Human granulocyte-macrophage colony stimulating factor (GM-CSF) polypeptides are provided, as well as methods for synthesizing conformationally and antigenically neutral muteins. The disclosed GM-CSFs promote growth and development of various hematopoietic lineages, and may be useful in treating conditions involving depressed blood cell populations and/or depressed blood cell regeneration, such as myeloid hypoplasia, chronic infections, and hematopoiesis after bone marrow transplantation. The invention also includes a novel mammalian expression vector.
### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Austria</td>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
<td>GA</td>
<td>Gabon</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
<td>GB</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
<td>HU</td>
<td>Hungary</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>IT</td>
<td>Italy</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
<td>KR</td>
<td>Republic of Korea</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
<td>LI</td>
<td>Liechtenstein</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
<td>LK</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
<td>LU</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>DE</td>
<td>Germany, Federal Republic of</td>
<td>MC</td>
<td>Monaco</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
<td>MG</td>
<td>Madagascar</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
<td>ML</td>
<td>Mali</td>
</tr>
<tr>
<td>MR</td>
<td>Mauritania</td>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>NL</td>
<td>Netherlands</td>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>RO</td>
<td>Romania</td>
<td>SD</td>
<td>Sudan</td>
</tr>
<tr>
<td>SE</td>
<td>Sweden</td>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>SU</td>
<td>Soviet Union</td>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>TG</td>
<td>Togo</td>
<td>US</td>
<td>United States of America</td>
</tr>
</tbody>
</table>
HUMAN GRANULOCYTE-MACROPHAGE
COLONY STIMULATING FACTOR AND MUTEINS THEREOF

Field of the Invention

The invention relates generally to protein mediators of human hematopoietic system growth and development, and more particularly, to human granulocyte-macrophage colony stimulating factor (GM-CSF) and genetically engineered muteins thereof.

BACKGROUND

Circulating blood cells are constantly replaced by newly developed cells. Replacement blood cells are formed in a process termed hematopoiesis which involves the production of at least eight mature blood cell lineages: red blood cells (erythrocytes), macrophages (monocytes), eosinophilic granulocytes, megakaryocytes (platelets), neutrophilic granulocytes, basophilic granulocytes (mast cells), T lymphocytes, and B lymphocytes (Burgess and Nicola, Growth Factors and Stem Cells (Academic Press, New York, 1983)). Much of the control of blood cell formation is mediated by a group of interacting glycoproteins termed colony stimulating factors (CSFs). These glycoproteins are so named because
of the *in vivo* and *in vitro* assays used to detect their presence. Techniques for the clonal culture of hematopoietic cells in semisolid culture medium have been especially important in the development of *in vitro* assays. In such cultures, individual progenitor cells (i.e., cells developmentally committed to a particular lineage, but still capable of proliferation) are able to proliferate to form a colony of maturing progeny in a manner which is believed to be essentially identical to the comparable process *in vivo*. The role of CSFs in hematopoiesis is the subject of many recent reviews, e.g. Metcalf, *The Hemopoietic Colony Stimulating Factors* (Elsevier, New York, 1984); Metcalf, *Science*, Vol. 229, pgs. 16-22 (1985); Nicola et al., *Immunology Today*, Vol. 5, pgs. 76-80 (1984); Whetton et al., *TIIBS*, Vol. 11, pgs. 207-211 (1986); and Clark and Kamen, *Science*, Vol. 236, pgs. 1229-1237 (1987).

The detection, isolation and purification of these factors is frequently extremely difficult, owing to the complexity of the supernatants they are typically located in, the divergencies and cross-overs of activities of the various components in the mixtures, the sensitivity (or lack thereof) of the assays utilized to ascertain the factors' properties, the frequent similarity in the range of molecular weights and other characteristics of the factors, and the very low concentration of the factors in their natural settings.

As more CSFs become available, primarily through molecular cloning, interest has heightened in finding clinical applications for them. Because of physiological similarities to hormones (e.g., soluble factors, growth mediators, action via cell receptors), potential uses of CSFs have been analogized to the current uses of hormones; e.g. Dexter, *Nature*, Vol. 321, pg. 198 (1986). Their use has been suggested for several
clinical situations where the stimulation of blood cell generation would be desirable, such as for rehabilitative therapy after chemotherapy or radiation therapy of tumors, treatment of myeloid hypoplasias, treatment of neutrophil deficiency, treatment to enhance hematopoietic regeneration following bone marrow transplantation, and treatment to increase host resistance to established infections; e.g. Dexter (cited above), Metcalf, *Science* (cited above), and Clark and Kamen (cited above).

Complementary DNAs (cDNAs) for GM-CSF, a factor which supports growth and development of granulocytes and macrophages, have recently been cloned and sequenced by a number of laboratories for several species; see e.g. Gough et al., *Nature*, Vol. 309, pgs. 763–767 (1984) (mouse); Lee et al., *Proc. Natl. Acad. Sci.*, Vol. 82, pgs. 4360–4364 (1985) (human); Wong et al., *Science*, Vol. 228, pgs. 810–815 (1985) (human and gibbon); and Cantrell et al., *Proc. Natl. Acad. Sci.*, Vol. 82, pgs. 6250–6254 (1985) (human). Moreover, non-recombinant GM-CSF has been purified from culture supernatants of the Mo cell line (described in U.S. Patent 4,438,032), and the first sixteen amino acids from the N-terminus have been sequenced (Gasson et al., *Science*, Vol. 226, pgs. 1339–1342 (1984)). Among the human GM-CSFs, heterogeneity of the nucleotide and amino acid sequences has been observed. For example, at the amino acid level both threonine and isoleucine have been observed at position 100 with respect to the N-terminal alanine of the mature polypeptide, suggesting that several allelic forms, or polymorphs, of GM-CSF may exist within human populations. This is not surprising, for Lewontin, chapter 3, in *The Genetic Basis of Evolutionary Change*, (Columbia University Press, New York, 1974), reviewed data on human and *Drosophila* allozymes which indicated that over a third of all such structural genes encode
polymorphic products distinguishable by electrophoresis. The existence of polymorphic forms of GM-CSF could be an important factor in their use as therapeutic agents, since some polymorphs may give rise to an immunogenic response in members of subpopulations which carry a different allele for GM-CSF than that corresponding to the GM-CSF used in therapy.

In view of the above, it would be advantageous for the medical community to have alternative forms of GM-CSF available, whether naturally occurring or genetically engineered.

SUMMARY OF THE INVENTION

The present invention is directed to human granulocyte-macrophage colony stimulating factor (GM-CSF), and genetically engineered muteins thereof. It includes polypeptides which comprise a sequence of amino acids selected from the set defined below by Formula I, and which possess granulocyte and macrophage colony stimulating factor activity as defined by standard assays. The invention also includes nucleic acids capable of encoding the polypeptides defined by Formula I, methods for making such polypeptides by use of the nucleic acids of the invention, and methods of using the polypeptides of the invention either alone or with other CSFs to treat disorders involving myeloid hypoplasia or to enhance hematopoietic system regeneration following bone marrow transplantation or following certain cancer therapies.

A preferred embodiment of the invention is the set of glycosylated or unglycosylated polypeptides which comprise a sequence of amino acids selected from those defined by the following formula:
wherein the terms X(Xaa) represent groups of synonymous amino acids. Most preferably, the invention includes unglycosylated forms of the polypeptides defined by Formula I.
Synonymous amino acids within a group have sufficiently similar physicochemical properties for substitution between members of the group to preserve the biological function of the molecule; Grantham, *Science*, Vol. 185, pgs. 862-864 (1974), and Dayhoff et al., pgs. 89-99 in *Atlas of Protein Sequence and Structures 1972*, Vol. 5 (National Biomedical Research Foundation, Washington, D.C.). Insertions and deletions of amino acids may also be made in the above-defined sequence without altering biological function, particularly if only a few amino acids are inserted or deleted, e.g. one, two or three, and do not remove or displace amino acids which are critical to a functional conformation; e.g. Shortle and Lin, *Genetics*, Vol. 110, pgs. 539-555 (1985), and Anfinsen, "Principles That Govern The Folding of Protein Chains", *Science*, Vol. 181, pgs. 223-230 (1973). Proteins and muteins produced by such deletions and/or insertions come within the purview of the present invention. Whenever amino acid residues of the protein of Formula I are referred to herein by number, such number or numbers are in reference to the N-terminus of the protein.

Preferably, the synonymous amino acid groups are those defined in Table I. More preferably, the synonymous amino acid groups are those given before the slashes in Table I; and most preferably each synonymous amino acid group consists of a single amino acid, namely Xaa itself.
<table>
<thead>
<tr>
<th>Xaa</th>
<th>X(Xaa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Ser, Ala, Thr,/Gly, Asn</td>
</tr>
<tr>
<td>Arg</td>
<td>Arg,/Lys</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu,/Val</td>
</tr>
<tr>
<td>Pro</td>
<td>Pro,/Ala</td>
</tr>
<tr>
<td>Thr</td>
<td>Thr, Ser, Ala/</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala, Ser, Thr, Gly,/Pro, Val</td>
</tr>
<tr>
<td>Val</td>
<td>Val, Ile,/Ala, Leu</td>
</tr>
<tr>
<td>Gly</td>
<td>Gly, Ala,/Ser</td>
</tr>
<tr>
<td>Ile</td>
<td>Ile, Val,/Leu</td>
</tr>
<tr>
<td>Phe</td>
<td>Phe,/Tyr</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyr,/Phe</td>
</tr>
<tr>
<td>His</td>
<td>His,/Gln, Asn</td>
</tr>
<tr>
<td>Gln</td>
<td>Gln,/Glu, His</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn, Asp,/Ser, Lys</td>
</tr>
<tr>
<td>Lys</td>
<td>Lys,/Arg, Asn</td>
</tr>
<tr>
<td>Asp</td>
<td>Asp, Asn, Glu/</td>
</tr>
<tr>
<td>Glu</td>
<td>Glu, Asp,/Gln</td>
</tr>
<tr>
<td>Met</td>
<td>Met,/Leu, Nle</td>
</tr>
<tr>
<td>Trp</td>
<td>Trp/</td>
</tr>
<tr>
<td>Cys</td>
<td>Cys/</td>
</tr>
</tbody>
</table>

(Nle stands for norleucine.)
The invention includes the polypeptides of Formula I with amino acid substitutions (of a synonymous amino acid for an amino acid of the polypeptide defined by Formula II, the native sequence) at a single position or at multiple positions. The term "N-fold substituted" is used to describe a subset of polypeptides defined by Formula I wherein synonymous amino acids have been substituted for from 0–N amino acids of the native sequence. Thus, for example, the group of 1-fold substituted polypeptides of Formula I consists of 247 polypeptides for the preferred groups of synonymous amino acids, and 105 for the more preferred groups of synonymous amino acids. These figures are the respective sums of the number of amino acids of each kind in the native GM-CSF sequence multiplied by the number of non-native amino acids in the corresponding synonymous amino acid set plus one for the unsubstituted sequence.

In a preferred embodiment, the invention comprises glycosylated or unglycosylated polypeptides comprising an amino acid sequence identical to that of a polypeptide selected from the set of 2-fold substituted polypeptides of Formula I, wherein the substituting amino acids are taken from the respective sets of synonymous amino acids defined by Table I.

In a more preferred embodiment, the invention comprises glycosylated or unglycosylated polypeptides comprising an amino acid sequence identical to that of a polypeptide selected from the set of 1-fold substituted polypeptides of Formula I, wherein the substituting amino acids are taken from the respective sets of synonymous amino acids defined by Table I.

In a still more preferred embodiment, the invention comprises glycosylated or unglycosylated polypeptides comprising an amino acid sequence identical to that of a polypeptide selected from the set of 1-fold
substituted polypeptides of Formula I, wherein the substituting amino acids are taken from the respective sets of synonymous amino acids defined by Table I, and wherein the amino acids in the region defined collectively by residues 1-75 inclusive, 89-95 inclusive, 104-112 inclusive, and 117-127 inclusive do not undergo substitution.

In a yet more preferred embodiment, the invention comprises glycosylated or unglycosylated polypeptides comprising an amino acid sequence identical to that of a polypeptide selected from the set of 2-fold substituted polypeptides of Formula I, wherein the substituting amino acids are taken from the respective sets of synonymous amino acids defined by Table I.

In a particularly preferred embodiment, the invention comprises glycosylated or unglycosylated polypeptides comprising an amino acid sequence identical to that of a polypeptide selected from the set of 1-fold substituted polypeptides of Formula I, wherein the substituting amino acids are taken from the respective sets of more preferred synonymous amino acids defined by Table I.

In a more particularly preferred embodiment, the invention comprises glycosylated or unglycosylated polypeptides comprising an amino acid sequence identical to that of a polypeptide selected from the set of 1-fold substituted polypeptides of Formula I, wherein the substituting amino acids are taken from the respective sets of more preferred synonymous amino acids defined by Table I, and wherein the amino acids in the region defined collectively by residues 1-75 inclusive, 89-95 inclusive, 104-112 inclusive, and 117-127 inclusive do not undergo substitution.
Most preferably, the invention comprises unglycosylated polypeptides comprising the sequence defined by the following formula:

\[
\begin{align*}
H & \rightarrow \text{ Ala } - \text{ Pro } - \text{ Ala } - \text{ Arg } - \text{ Ser } - \text{ Pro } - \text{ Ser } - \text{ Pro } - \\
& & \text{ Ser } - \text{ Thr } - \text{ Gln } - \text{ Pro } - \text{ Trp } - \text{ Glu } - \text{ His } - \text{ Val } - \\
& & \text{ Asn } - \text{ Ala } - \text{ Ile } - \text{ Gln } - \text{ Glu } - \text{ Ala } - \text{ Arg } - \text{ Arg } - \\
& & \text{ Leu } - \text{ Leu } - \text{ Asn } - \text{ Leu } - \text{ Ser } - \text{ Arg } - \text{ Asp } - \text{ Thr } - \\
& & \text{ Ala } - \text{ Ala } - \text{ Glu } - \text{ Met } - \text{ Asn } - \text{ Glu } - \text{ Thr } - \text{ Val } - \\
& & \text{ Glu } - \text{ Val } - \text{ Ile } - \text{ Ser } - \text{ Glu } - \text{ Met } - \text{ Phe } - \text{ Asp } - \\
& & \text{ Leu } - \text{ Gln } - \text{ Glu } - \text{ Pro } - \text{ Thr } - \text{ Cys } - \text{ Leu } - \text{ Gln } - \\
& & \text{ Thr } - \text{ Arg } - \text{ Leu } - \text{ Glu } - \text{ Leu } - \text{ Tyr } - \text{ Lys } - \text{ Gln } - \\
& & \text{ Gly } - \text{ Leu } - \text{ Arg } - \text{ Gly } - \text{ Ser } - \text{ Leu } - \text{ Thr } - \text{ Lys } - \\
& & \text{ Leu } - \text{ Lys } - \text{ Gly } - \text{ Pro } - \text{ Leu } - \text{ Thr } - \text{ Met } - \text{ Met } - \\
& & \text{ Ala } - \text{ Ser } - \text{ His } - \text{ Tyr } - \text{ Lys } - \text{ Gln } - \text{ His } - \text{ Cys } - \\
& & \text{ Pro } - \text{ Pro } - \text{ Thr } - \text{ Pro } - \text{ Glu } - \text{ Thr } - \text{ Ser } - \text{ Cys } - \\
& & \text{ Ala } - \text{ Thr } - \text{ Gln } - \text{ Ile } - \text{ Ile } - \text{ Thr } - \text{ Phe } - \text{ Glu } - \\
& & \text{ Ser } - \text{ Phe } - \text{ Lys } - \text{ Glu } - \text{ Asn } - \text{ Leu } - \text{ Lys } - \text{ Asp } - \\
& & \text{ Phe } - \text{ Leu } - \text{ Leu } - \text{ Val } - \text{ Ile } - \text{ Pro } - \text{ Phe } - \text{ Asp } - \\
& & \text{ Cys } - \text{ Trp } - \text{ Glu } - \text{ Pro } - \text{ Val } - \text{ Gln } - \text{ Glu } - \text{ OH}
\end{align*}
\]

Formula II

The term "N-fold inserted" in reference to the polypeptides of Formula I is used to describe a set of polypeptides wherein from 1 to N amino acids have been inserted into the sequence defined by Formula I. Preferably, the inserted amino acids are selected from the preferred groups of synonymous amino acids (Table I) of the amino acids flanking the insertion; more
preferably they are selected from the more preferred groups of synonymous amino acids (Table I) of the amino acids flanking the insertion. Thus, for example, one subgroup of the group of 1-fold inserted peptides comprises an amino acid inserted between the N-terminal X(Ala) and the adjacent X(Pro). The insertions defining the members of this subgroup are preferably selected from Ala, Pro, Ser, Thr, Gly, and Val; more preferably they are selected from Ala, Pro, Ser, Thr, and Gly; and most preferably they are selected from Ala and Pro. Insertions can be made between any adjacent amino acids of Formula I or at either end of the amino acid sequence. Since there are 127 amino acids and therefore 128 possible insertion locations, and since multiple insertions can be made at the same location, a 2-fold inserted polypeptide of Formula I gives rise to 128 x 128 = 16,384 subgroups of peptides, and the size of each subgroup depends on the sizes of the synonymous amino acid groups of the amino acids flanking the insertions.

The term "N-fold deleted" in reference to the polypeptides of Formula I is used to describe a set of peptides having from 1 to N amino acids deleted from the sequence defined by Formula I. Thus, the set of 1-fold deleted polypeptides of Formula I consists of 127 subgroups of polypeptides each 126 amino acids in length (126-mers). Each of the subgroups in turn consists of all the 126-mers defined by the preferred, more preferred, and most preferred synonymous amino acid groups depending on the multiplicity of substitutions.

Whenever a set of polypeptides is defined in terms of deletions and insertions, both with respect to the sequence of Formula I, first the set of polypeptides defined by the insertions is formed and then for each member of that set a set of polypeptides defined by deletions is formed. The sum of the members of all the
latter sets forms the set of polypeptides defined by both insertions and deletions.

The invention further includes nucleotide sequences capable of encoding the polypeptides of Formula I and nucleic acid cassettes capable of encoding portions of the polypeptides of Formula I for the preferred, more preferred, and most preferred groups of synonymous amino acids. As used herein, "nucleic acid cassette" means a nucleic acid of from 10 to 120 bases (i) which has a nucleotide sequence capable of encoding a portion of a polypeptide defined by Formula I, and (ii) whose ends are defined by restriction endonuclease sites which are unique with respect to the cloning or expression vector containing the nucleic acid and its complementary strand. As explained more fully below, nucleic acid cassettes are used to construct mutant forms of GM-CSF.

Polypeptides of the invention may be part of a longer polypeptide which comprises an additional peptide linked to the N-terminus of the polypeptide of the invention. The additional peptide may be a signal sequence for directing secretion of the protein from the host organism, e.g. M Falphal signal sequence in yeast (Kurjan et al., Cell, Vol. 30, pgs. 933-943 (1982)). Or, the additional peptide may be used to facilitate the purification of the desired polypeptide, e.g. Sassenfeld et al., "A Polypeptide Fusion Designed for the Purification of Recombinant Proteins", Biotechnology, pgs. 76-81 (January 1984). In the latter case, the additional peptide may contain a selective cleavage site adjacent to the N-terminus of the desired polypeptide, so that that part of the amino acid sequence that is superfluous (with respect to biological activity) may be split off by either enzymatic or chemical means.

Polypeptides of the invention may also comprise substitutions by amino acid analogs, such as norleucine.
In particular the invention includes nucleic acids having the sequence of the cDNA insert of clone pcD-human-GM-CSF illustrated in Figure 1, which has been deposited with the American Type Culture Collection, Rockville, Maryland, under accession number 39923.

Furthermore, the invention includes expression and/or cloning vectors for GM-CSF comprising the following covalently linked elements in sequence (as illustrated below): SV40 origin of DNA replication (SV40 ori), an SV40 early region promoter (SV40 early) SRα promoter, splice junction, and a GM-CSF coding region (in either order), and a polyadenylation site (polyA site):

SV40 ori
SV40 early promoter
SRα promoter
splice junction
GM-CSF coding region
poly A site
[vector ring closes onto SV40 ori, above]

Preferably, the expression and/or cloning vector further includes, inserted between the SV40 ori and the polyA site, a bacterial origin of replication (bacterial ori) and a gene for a selectable marker (as illustrated below):

SV40 ori
SV40 early promoter
SRα promoter
splice junction
GM-CSF coding region
poly A site
bacterial ori
selectable marker
[vector ring closes onto SV40 ori, above]
More preferably, the gene for the selectable marker confers drug resistance to a host bacterium, such as resistance to neomycin, ampicillin, tetracycline, streptomycin, kanamycin, hygromycin, or the like. Most preferably, the bacterial ori is the pBR322 ori.


The present invention is addressed to problems associated with the application of immunoregulatory agents to treat medical and/or veterinary disorders. In particular, it provides compounds capable of stimulating the regeneration of blood cells, which may be useful in cancer therapy, the treatment of certain blood diseases, and the treatment of persistent infections.

**BRIEF DESCRIPTION OF THE FIGURES**

Figure 1 illustrates the nucleotide sequence and corresponding amino acid sequence of a cDNA insert which encodes a polypeptide exhibiting human GM-CSF activity.

Figure 2A illustrates pcD-human-GM-CSF, a plasmid carrying a cDNA insert which encodes a polypeptide exhibiting human GM-CSF activity.

Figure 2B is a restriction endonuclease cleavage map of the cDNA insert of Figure 2A.

Figure 3 diagrammatically illustrates restriction sites and major coding regions of plasmid pLL.
Figure 4 diagrammatically illustrates restriction sites and major coding regions of plasmid pcDVI.

DETAILED DESCRIPTION OF THE INVENTION

The present invention includes glycosylated or unglycosylated polypeptides which exhibit human GM-CSF activity, and which are derivable from the GM-CSF polypeptides disclosed herein using standard protein engineering techniques. The invention also includes nucleic acids having sequences capable of coding for the polypeptides of the invention. Finally, the invention includes methods of making the glycosylated or unglycosylated polypeptides of the invention which utilize the nucleotide sequences disclosed herein, and methods of using the polypeptides of the invention.

Techniques for making, using, and identifying the polypeptides of the invention are discussed below in general terms. Afterwards several specific examples are provided wherein the general techniques are applied using specific cell types, vectors, reagents, and the like.

I. De Novo Preparation of GM-CSF cDNA

A variety of methods are now available for de novo preparation and cloning of cDNAs and for the construction of cDNA libraries. A review of relevant literature, together with procedural details, is given at page 29, line 22 to page 32, line 5 from the bottom of Application PCT/US 86/02464, published as WO 87/02990. A preferred source of mRNA encoding the polypeptides of the present invention is discussed in the next paragraph. The disclosure in Application PCT/US 86/02464 must be read here in relation to the production of GM-CSF, especially on page 31 thereof.
A preferred source of mRNA encoding the desired polypeptides is cells whose supernatants contain one of the activities associated with the polypeptides of the present invention, such as T lymphocytes. RNAs for deriving GM-CSF cDNAs can be obtained from the Mo cell line disclosed in U.S. Patent 4,438,032 and deposited in the American Type Culture Collection under accession number CRL 8066. In general, suitable T cells can be obtained from a variety of sources, such as human spleen, tonsils and peripheral blood. T cell clones, such as those isolated from peripheral blood T-lymphocytes, may also be used (see Research Monographs in Immunology, eds. von Doehner and Haaf, Section D: "Human T Cell Clones", Vol. 8, pgs. 243-333; Elsevier Science Publishers, N.Y. (1985)).

II. Preparation of GM-CSF cDNAs via Hybridization Probes Derived from Disclosed cDNA

Very often within a population several closely related forms of a protein exist which perform the same biological function. The best studied examples are (1) the so-called "allozymes", which are allelic forms of an enzyme that can be distinguished by electrophoresis; see e.g. Sensabaugh, "The Utilization of Polymorphic Enzymes in Forensic Science," Isozymes, Vol. 11, pgs. 137-154 (1983), and Lewontin, chapter 3 in The Genetic Basis of Evolutionary Change (Columbia University Press, New York, (1974); (2) the so-called "allootypes", which are polymorphisms of the constant regions of immunoglobulins, e.g. Hood et al., Immunology, pgs. 231-238 (Benjamin/Cummings, Menlo Park, 1978); and (3) hemoglobins, e.g. Dickerson and Geis, Hemoglobin (Benjamin/Cummings, Menlo Park, 1983). Such polymorphisms are also believed to exist among

There is a reasonably high probability that a randomly selected locus is polymorphic. Lewontin (cited above) presents data for human and Drosophila loci which indicate that about a third of all structural genes encode polymorphic products distinguishable by electrophoresis. The number of major variants at each locus does not appear to be large: Vogel et al., Human Genetics, pgs. 373-378 (Springer-Verlag, New York, 1979) lists 33 major human polymorphisms, and the number of allelic variants at each locus ranges from 2 to 6. Herskowitz, Principles of Genetics, 2nd Ed. (Macmillan, New York, 1977) pg. 643, cites examples of polymorphic Drosophila loci having 6, 4, 2, 3, 2, and 33 alleles.

Although sequences of polymorphic variants are not readily predicted, studies such as those of Dayhoff (cited above) strongly suggest that polymorphic variants probably differ from the dominant forms by no more than "conservative" amino acid changes. Also, the nucleotide sequences of the polymorphic variants are highly homologous to those of the dominant forms, e.g. more than 99% homologous.

The cDNAs disclosed herein can be used as probes to identify polymorphic variants and/or homologous sequences in different cell types, as an alternative to de novo isolation and cloning of the GM-CSF coding nucleotides. Standard techniques are employed: e.g. Beltz et al., "Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods," Methods in Enzymology, Vol. 100, pgs. 266-285
(1983); and Callahan et al., "Detection and Cloning of Human DNA Sequences Related to the Mouse Mammary Tumor Virus Genome," Proc. Natl. Acad. Sci., Vol. 79, pgs. 5503-5507 (1982), the former reference being incorporated by reference. Restriction endonuclease fragments of the vector carrying the cDNAs of the invention are used to construct probes (using standard techniques such as nick-translation, e.g. see Maniatis et al., cited above) for screening, at low hybridization stringencies, genomic or cDNA libraries (again, constructed by standard techniques) of a cell type suspected of producing GM-CSF. Standard screening procedures are employed, e.g. Grunstein et al., Proc. Natl. Acad. Sci., Vol. 72, pgs. 3961-3965 (1975), or Benton et al., Science, Vol. 196, pgs. 180-183 (1977), or Woo, Methods in Enzymology, Vol. 68, pgs. 389-396 (1979). Alternatively, libraries can be screened with labeled oligonucleotide probes whose sequences are determined from Formula I or from the corresponding anti-sense sequence. Such probes can be synthesized on commercially available DNA synthesizers, e.g. Applied Biosystems model 381A, using standard techniques, e.g. Gait, Oligonucleotide Synthesis: A Practical Approach, (IRL Press, Washington D.C., 1984). In either case, it is preferable that the probe be at least 18-30 bases long. More preferably, the probe is at least 100-200 bases long. By way of example, restriction fragments formed by digestion with individual or pairs of restriction endonucleases indicated in Figure 2B can be used to construct suitable probes for screening homologous nucleic acids.
III. Preparation of Human GM-CSF Muteins by Protein Engineering

Once nucleic acid sequence and/or amino acid sequence information is available for a native protein, a variety of techniques become available for producing virtually any mutation in the native sequence. Shortly, in Science, Vol. 229, pgs. 1193-1201 (1985), reviews techniques for mutating nucleic acids which are applicable to the present invention. Preferably, mutants of the native GM-CSFs, i.e. GM-CSF muteins, are produced by site-specific oligonucleotide-directed mutagenesis: e.g. Zoller and Smith, Methods in Enzymology, Vol. 100, pgs. 468-500 (1983); Mark et al., U.S. Patent 4,518,584 entitled "Human Recombinant Interleukin-2 Muteins," which are incorporated by reference; direct synthesis by solid phase methods, e.g. Clark-Lewis et al., Science, Vol. 231, pgs. 134-139 (1986); and Doescher, Meth. Enzymol., Vol. 47, pgs. 578-617 (1977); or by so-called "cassette" mutagenesis described by Wells et al., in Gene, Vol. 34, pgs. 315-323 (1985), and by Estell et al., Science, Vol. 233, pgs. 659-663 (1986), and also described essentially by Mullenbach et al., J. Biol. Chem., Vol. 261, pgs. 719-722 (1986), and by Feretti et al., Proc. Natl. Acad. Sci., Vol. 83, pgs. 597-603 (1986).

Muteins of the natural polypeptide may be desirable in a variety of circumstances. For example, undesirable side effects might be shown less by certain muteins, particularly if the side effect is associated with a different part of the polypeptide from that of the desired activity. In some expression systems, the native polypeptide may be susceptible to degradation by proteases. In such systems, selected substitutions and/or deletions of amino acids which change the susceptible sequences can significantly enhance yields;
e.g. British patent application 2173-804-A where Arg at position 275 of human tissue plasminogen activator is replaced by Gly or Glu. Muteins may also increase yields in purification procedures and/or increase shelf lives of proteins by eliminating amino acids susceptible to oxidation, acylation, alkylation, or other chemical modifications. For example, methionine readily undergoes oxidation to a sulfoxide, which in many proteins is associated with loss of biological activity, e.g. Brot and Weissbach, Arch. Biochem. Biophys., Vol. 223, pg. 271 (1983). Often methionines can be replaced by more inert amino acids with little or no loss of biological activity; e.g. Australian patent application AU-A-52451/86. In bacterial expression systems, yields can sometimes be increased by eliminating or replacing conformationally inessential cysteine residues, e.g. Mark et al., U.S. Patent 4,518,584.

In sections below the notation used by Estell et al. (cited above) to identify muteins is followed and generalized. For example, "human GM-CSF mutein Ala\(^9\)" (or simply "Ala\(^9\)" if the native protein is understood from the context) indicates a polypeptide whose amino acid sequence is identical to that of the native protein except for position 9 with respect to the N-terminus. At that position Ala has been substituted for Ser. Where the mutein contains more than one substitution, e.g. Ala for Ser at position 9, and Val for Ile at position 100, the mutein is referred to as human IL-3 mutein (Ala\(^9\), Val\(^{100}\)). Deletions are indicated by "Δ's". For example, a mutein lacking Arg at position 4 is referred to as human GM-CSF mutein Δ\(^4\). Insertions are indicated by "IN(Xaa)'s". For example, a mutein with a Leu inserted after Arg at position 4 is referred to as human GM-CSF mutein IN\(^4\)(Leu). Thus, human GM-CSF mutein (Ser\(^{10}\), Δ\(^4\), IN\(^6\)(Gly)) represents the native human GM-CSF sequence.
which has been modified by replacing Thr by Ser at position 10, deleting Gln at position 4, and inserting Gly immediately after Pro at position 6. Insertion of multiple amino acids at the same site is indicated by \( \text{IN}^i(X\text{aa}_1-X\text{aa}_2-X\text{aa}_3-\ldots) \), where \( X\text{aa}_1-X\text{aa}_2-X\text{aa}_3-\ldots \) is the sequence inserted after position i. N-terminal additions are indicated by superscript "0", e.g. \( \text{IN}^0(X\text{aa}) \), and a sequence of deletions of amino acids 6-10, for example, is designated either as \( \Delta^6-10 \) or as \( (\Delta^6, \Delta^7, \Delta^8, \Delta^9, \Delta^{10}) \).

Most preferably cassette mutagenesis is employed to generate human GM-CSF muteins. As described more fully below, a synthetic human GM-CSF gene is constructed with a sequence of unique restriction endonuclease sites spaced approximately uniformly along the gene. The uniqueness of the restriction sites should be retained when the gene is inserted into an appropriate vector, so that segments of the gene can be conveniently excised and replaced with synthetic oligonucleotides (i.e. "cassettes") which code for desired mutations.

Determination of the number and distribution of unique restriction sites entails the consideration of several factors including (1) preexisting restriction sites in the vector to be employed in expression, (2) whether species or genera-specific codon usage is desired, (3) the number of different non-vector-cutting restriction endonucleases available (and their multiplicities within the synthetic gene), and (4) the convenience and reliability of synthesizing and/or sequencing the segments between the unique restriction sites.

Generally, Formula I defines sets of polypeptides with conservative amino acid substitutions, insertions, and/or deletions with respect to the mature native human GM-CSF sequence. "Conservative" as used
herein means (i) that the alterations are as conformationally neutral as possible, that is, designed to produce minimal changes in the tertiary structure of the mutant polypeptides as compared to the native human GM-CSF, and (ii) that the alterations are as antigenically neutral as possible, that is, designed to produce minimal changes in the antigenic determinants of the mutant polypeptides as compared to the native human GM-CSF. Conformational neutrality is desirable for preserving biological activity, and antigenic neutrality is desirable for avoiding the triggering of immunogenic responses in patients or animals treated with the compounds of the invention. While it is difficult to select with absolute certainty which alternatives will be conformationally and antigenically neutral, rules exist which can guide those skilled in the art to make alterations that have high probabilities of being conformationally and antigenically neutral, e.g. those of Anfiseń (cited above), and Berzofsky, Science, Vol. 229, pgs. 932-940 (1985). Some of the more important rules include (1) substitution of hydrophobic residues is less likely to produce changes in antigenicity because they are likely to be located in the protein's interior, e.g. Berzofsky (cited above); (2) substitution of physiochemically similar, i.e. synonymous, residues is less likely to produce conformational changes because the substituting amino acid can play the same structural role as the replaced amino acid; (3) alteration of evolutionarily conserved sequences is likely to produce deleterious conformational effects because evolutionary conservation suggests sequences may be functionally important, and (4) negatively charged residues, e.g. Asp and Glu, tend to be more immunogenic than neutral or positively charged residues: Geysen et al., Science, Vol. 235, pgs. 1184-1190 (1987).
In addition to such basic rules for selecting mutein sequences, assays are available to confirm the biological activity and conformation of the engineered molecules. Biological assays for the polypeptides of the invention are described more fully below. Changes in conformation can be tested by at least two well known assays: the microcomplement fixation method, e.g. Wasserman et al., J. Immunol., Vol. 87, pgs. 290-295 (1961), or Levine et al. Methods in Enzymology, Vol. 11, pgs. 928-936 (1967), used widely in evolutionary studies of the tertiary structures of proteins; and affinities to sets of conformation-specific monoclonal antibodies, e.g. Lewis et al., Biochemistry, Vol. 22, pgs. 948-954 (1983).

IV. Assays for GM-CSF Activity

To determine GM-CSF activity, hemopoietic cells, e.g. bone marrow cells or fetal cord blood cells, are made into a single cell suspension. The individual cells are then "immobilized" in a semi-solid (agar) or viscous (methylcellulose) medium containing nutrients and usually fetal calf serum. In the presence of an appropriate stimulating factor, individual cells will proliferate and differentiate. Since the initial cells are immobilized, colonies develop as the cells proliferate and mature. These colonies can be scored after 7-14 days. Detailed instructions for conducting GM-CSF assays are given by Burgess, A., Growth Factors and Stem Cells, pgs. 52-55 (Academic Press, New York 1984), and Metcalf, The Hemopoietic Colony Stimulating Factors, pgs. 103-125 (Elsevier, New York, 1984), both sections of these references being incorporated by reference. If desired, individual colonies can be extracted, placed on microscope slides, fixed and stained with Wright/Geimsa (Todd-Sanford, Clinical Diagnosis by
Laboratory Methods, 15th Edition, Eds. Davidson and Henry [1974]). Morphological analysis of cell types present per single colony can then be determined.

Bone marrow cells collected from patients with nonhematologic disease are layered over Ficoll (type 400, Sigma Chemical Co., St. Louis, MO), centrifuged (600 x g, 20 min), and the cells at the interface removed. These cells are washed twice in Iscove's Modified Dulbecco's Medium containing 10% fetal calf serum (FCS) and resuspended in the same medium, and the adherent cells removed by adherence to plastic Petri dishes (adherent cells are frequently GM-CSF producing cells: Metcalf (cited above)). The nonadherent cells are added at 10^5 cells/ml to Iscove's Medium containing 20% FCS, 50 μM 2-mercaptoethanol, 0.9% methylcellulose and various concentrations of either supernatants known to contain colony stimulating activity or test supernatants. One ml aliquots are plated in 35 mm petri dishes and cultured at 37°C in a fully humidified atmosphere of 6% CO₂ in air. Three days after the initiation of the culture, 1 unit of erythropoietin is added to each plate. Granulocyte-macrophage colonies and erythroid bursts are scored at 10-14 days using an inverted microscope.

Cord blood cells collected in heparin are spun at 600 x g for 6 min. The white blood cells at the interface between the plasma and red blood cell peak are transferred to a tube containing 0.17 N ammonium chloride and 6% FCS. After 5 min on ice, the suspension is underlaid with 4 ml FCS and centrifuged for 6 minutes at 600 x g. The cell pellet is washed with Dulbecco's phosphate buffered saline and put through the Ficoll and plastic adherence steps as described above for bone marrow cells. The low-density nonadherent cells are collected and placed at 10^5 cells/culture in the semi-solid culture medium as described above.
At the end of the assays, the cellular composition is determined after applying the individual colonies to glass slides and staining with Wright-Geimsa. Eosinophils are determined by staining with Luxol Fast Blue, e.g. Johnson, G. and Metcalf, D., *Exp. Hematol.*, Vol. 8, pgs. 549-561 (1980).

V. **Purification and Pharmaceutical Compositions**

The polypeptides of the present invention expressed in *E. coli*, in yeast or in other cells can be purified according to standard procedures of the art, including ammonium sulfate precipitation, fractionation column chromatography (e.g., ion exchange, gel filtration, electrophoresis, affinity chromatography, etc.) and ultimately crystallization (see generally "Enzyme Purification and Related Techniques," Methods in Enzymology, 22:233-577 [1977] and Scopes, R., *Protein Purification: Principles and Practice*, Springer-Verlag, New York [1982]). The precise purification protocol selected depends, to a large extent, on the expression host used. For example, mammalian expression hosts frequently require the presence of serum in their growth medium, which necessitates additional steps for removing the serum proteins. Yeast and bacteria, on the other hand, grow on defined media from which it is usually simpler to separate secreted products. Some expression hosts may not secrete the expressed product, and then the product must be separated from a homogenate or extract of expression host. The purification problems encountered in all of these situations are readily solvable by those with ordinary skill in the art of biochemical purifications.

Once purified, partially or to homogeneity, the polypeptides of the invention may be used for research
purposes, e.g., as a supplement to cell growth media (e.g., minimum essential medium Eagle, Iscove's modified Dulbecco Medium or RPMI 1640, available from Sigma Chemical Company (St. Louis, MO) and GIBCO Division (Chagrin Falls, OH)) and as an antigenic substance for eliciting specific immunoglobulins useful in immunoassays, immunofluorescent stainings, etc. (See generally Immunological Methods, Vols. I & II, Eds. Lefkovits, I. and Pernis, B., Academic Press, New York, N.Y. [1979 & 1981], and Handbook of Experimental Immunology, ed. Weir, D., Blackwell Scientific Publications, St. Louis, MO [1978].)

The polypeptides of the present invention may also be used in pharmaceutical compositions, e.g., to enhance natural defenses against persistent infections or to promote blood cell regeneration. Thus, patients with rheumatoid arthritis, in need of a transplant, or with immunodeficiency caused by cancer chemotherapy, advanced age, immunosuppressive agents, etc., may be treated directly with such polypeptides. Or populations of hematopoietic cells of one person may be maintained and/or expanded or caused to differentiate ex vivo for eventual reintroduction into the same or another person for a beneficial effect. The compositions can selectively stimulate various components of the immune system, either alone or with other agents well known to those skilled in the art. In particular, the compositions may include other immune-reactive agents, such as lymphokines (e.g. IL-1, IL-2, IL-3, IL-4, G-CSF, M-CSF, or the like).

Pharmaceutical compositions containing the polypeptides of this invention can be prepared and used as described in Application PCT/US 86/02464 at page 44, line 27 to page 45, line 21. In particular, the quantity of active compound in a unit dose of preparation may be
varied or adjusted from 1 µg to 100 mg, according to the particular application and the potency of the active ingredient.

VI. **Expression Systems**

Once the cDNA of the invention has been cloned, a wide range of expression systems (i.e. host-expression vector combinations) can be used to produce the proteins of the invention. Possible types of host cells include, but are not limited to, bacterial, yeast, insect, mammalian, and the like. Selecting an expression system, and optimizing protein production thereby, requires the consideration and balancing of many factors, including (1) the nature of the protein to be expressed, e.g. the protein may be poisonous to some host organisms, it may be susceptible to degradation by host proteases, or it may be expressed in inactive conformations or in insoluble form in some hosts, (2) the nature of the messenger RNA (mRNA) corresponding to the protein of interest, e.g. the mRNA may have sequences particularly susceptible to host endonucleases, which drastically reduce the functional lifetime of the mRNA, or the mRNA may form secondary structures that mask the start codon or ribosome binding site, thereby inhibiting translation initiation in some hosts, (3) the selection, availability, and arrangement of host-compatible expression control sequences in the 3'- and 5'-regions flanking the coding region -- these include promoters, 5'- and 3'-protector sequences, ribosome binding sites, transcription terminators, enhancers, polyadenylate addition sites, cap sites, intron-splice sites, and the like, (4) whether the protein has a secretion signal sequence which can be processed by the host, or whether an expression control sequence encoding a signal sequence
endogenous to the host must be spliced onto the region encoding the mature protein, (5) the available modes and efficiencies of transfection or transformation of the host, and whether transient or stable expression is desired, (6) the scale and cost of the host culture system desired for expressing the protein, (7) whether, and what type of, posttranslational modifications are desired, e.g. the extent and kind of glycosylation desired may affect the choice of host, e.g. Uy and Wold, Science, Vol. 198, pgs. 890-896 (1977), (8) the ease with which the expressed protein can be separated from proteins and other materials of the host cells and/or culture medium e.g. in some cases it may be desirable to express a fusion protein with a specialized signal sequence to aid in later purification steps, (9) the stability and copy number of a particular vector in a selected host, e.g. Hofschneider et al., eds. Gene Cloning in Organisms Other than E. Coli (Springer Verlag, Berlin, 1982), and (10) like factors known to those skilled in the art.

Many reviews are available which provide guidance for making choices and/or modifications of specific expression systems in light of the recited factors: e.g., de Boer and Shepard, "Strategies for Optimizing Foreign Gene Expression in Escherichia coli," pgs. 205-247, in Kroon, ed. Genes: Structure and Expression (John Wiley & Sons, New York, 1983), review several E. coli expression systems; Kucherlapati et al., Critical Reviews in Biochemistry, Vol. 16, Issue 4, pgs. 349-379 (1984), and Banerji et al., Genetic Engineering, Vol. 5, pgs. 19-31 (1983) review methods for transfecting and transforming mammalian cells; Reznikoff and Gold, eds. Maximizing Gene Expression (Butterworths, Boston, 1986) review selected topics in gene expression in E. coli, yeast, and mammalian cells; and Thilly Mammalian
Cell Technology (Butterworths, Boston, 1986) reviews mammalian expression systems.

Likewise, many reviews are available which describe techniques and conditions for linking and/or manipulating specific cDNAs and expression control sequences to create and/or modify expression vectors suitable for use with the present invention: e.g. Maniatis et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, N.Y., 1982); Glover, DNA Cloning: A Practical Approach, Vol. I and II (IRL Press, Oxford, 1985); and Perbal, A Practical Guide to Molecular Cloning (John Wiley & Sons, N.Y., 1984). Generally, within an expression vector various sites may be selected for insertion of the cDNA of the invention. These sites are usually designated by the restriction endonuclease which cuts them and are well recognized by those of skill in the art. Various methods for inserting DNA sequences into these sites to form recombinant DNA molecules are also well known. These include, for example, dG-dC or dA-dT tailing, direct ligation, synthetic linkers, exonuclease and polymerase-linked repair reactions followed by ligation, or extension of the DNA strand with DNA polymerase and an appropriate single-stranded template followed by ligation.

Often a vector containing the cDNA of the invention must be obtained in large quantities before transfection and/or transformation of cells in a host culture can take place. For this purpose the vector is often replicated without significant expression in an organism (the cloning host) other than the one finally used for expression. Then, after propagation, the vectors are separated from the cloning host using standard techniques, e.g. as disclosed by Maniatis et al. (cited above).
"Digestion" of DNA refers to catalytic cleavage of the DNA usually with an enzyme that acts only at certain locations in the DNA. For the most part such enzymes are restriction endonucleases, and the sites on the DNA for which each is specific are called a restriction site. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors, and other requirements as established by the enzyme suppliers are used. Generally, about 1 microgram of plasmid or DNA fragment is used with about 1 unit of enzyme in about 20 microliters of buffer solution. Appropriate buffer and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After incubation, protein is removed by extraction with phenol and chloroform, and the digested nucleic acid is recovered from the aqueous fraction by precipitation with ethanol.

"Isolation" or "purification" of a given DNA fragment from a restriction (or other) digest means separation of the digest by polyacrylamide gel electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. Isolation procedures are well known, e.g. Lawn et al., Nucleic Acids Research, Vol. 9, pgs. 6103-6114 (1981); Goeddel et al. Nucleic Acids Research, Vol. 8, pg. 4057 (1980), and Maniatis et al. (cited above).

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments. Ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase per
0.5 micrograms of approximately equimolar amounts of the DNA fragments to be ligated.

"Amplification" of a plasmid means transforming a suitable cloning host and propagating the host to increase the total number of plasmids.

"Kinased" DNA fragments refer to fragments which have been phosphorylated with polynucleotide kinase. Such treatment enhances the efficiency of ligation of DNA fragments lacking 5'-phosphates.

Suitable prokaryote expression vectors include plasmids from E. coli, e.g. Col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, e.g. RP4, phage DNAs, such as phage lambda and its various derivatives, M13, and the like. Additional E. coli vectors are described in Chapter 12 of Maniatis et al. (cited above) for expressing eukaryotic proteins, either fused or unfused with prokaryotic peptides. And Riggs discloses still further E. coli expression systems in U.S. Patent 4,431,739, which is incorporated by reference.

Commonly used prokaryotic promoters include the β-lactamase (penicillinase) and lactose promoter systems (Chang et al., Nature, Vol. 275, pg. 615 (1978); Itakura, et al., Science, Vol. 198, pg. 1056 (1977); Goeddel, et al. Nature Vol. 281, pg. 544 (1979)); and the tryptophan (trp) promoter system (Goeddel, et al., Nucleic Acids Res., Vol. 8, pg. 4057 (1980); EPO Appl Publ No. 0036776). Whereas these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling a skilled worker to ligate them functionally with plasmid vectors, e.g. Siebenlist et al., Cell, Vol. 20, pg. 269 (1980).

A particularly useful prokaryotic promoter for high expression in E. coli is the tac promoter, disclosed
by de Boer in U.S. Patent 4,551,433, which is incorporated herein by reference. Secretion expression vectors are also available for E. coli hosts. Particularly useful are the pIN-III-ompA vectors, disclosed by Ghayeb et al. in EMBO J., Vol. 3, pgs. 2437-2442 (1984), in which the cDNA to be transcribed is fused to the portion of the E. coli ompA gene encoding the signal peptide of the ompA protein which, in turn, causes the mature protein to be secreted into the periplasmic space of the bacteria. Likewise, U.S. Patents 4,336,336 and 4,338,397 disclose secretion expression vectors for prokaryotes. Accordingly, these references are incorporated by reference.

Numerous strains of bacteria are suitable hosts for prokaryotic expression vectors including strains of E. coli, such as W3110 (ATCC No. 27325), JA221, C600, ED767, DH1, LE392, HB101, X1776 (ATCC No. 31244), X2282, RRL (ATCC No. 31343) MRC1; strains of Bacillus subtilis; and other enterobacteriaceae such as Salmonella typhimurium or Serratia marcescens, and various species of Pseudomonas. General methods for deriving bacterial strains, such as E. coli K12 X1776, useful in the expression of eukaryotic proteins is disclosed by Curtis in U.S. Patent 4,190,495. Accordingly this patent is incorporated by reference.

Eukaryotic microbes, such as yeast cultures, can also be used to express proteins of the invention. Saccharomyces cerevisiae, or common baker's yeast, is the most commonly used among eukaryotic microorganisms, although a number of other strains are commonly available. For expression in Saccharomyces, the plasmid YRp7 can be used, e.g. Stinchcomb, et al, Nature, Vol. 282, pg. 39 (1979), Kingsman et al, Gene, Vol. 7, pg. 141 (1979), and Tschemper, et al., Gene, Vol. 10, pg. 157 (1980). This plasmid already contains the trpL gene.
which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1 (Jones, Genetics, Vol. 85, pg. 12 (1977)). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Additional vectors for yeast include the pGAL plasmids disclosed by Miyajima et al., Nucleic Acids Research, Vol. 12, pgs. 6397-6414 (1984), and the 2μm plasmid disclosed by Beggs, Nature, Vol. 275, pgs. 104-109 (1978).

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman, et al., J. Biol. Chem., Vol. 255, pg. 2073 (1980)) or other glycolytic enzymes (Hess, et al, J. Adv. Enzyme Reg., Vol. 7, pg. 149 (1968); Holland, et al, Biochemistry, Vol. 17, pg. 4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose-phosphate isomerase, phosphoglucone isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector adjacent to the 3'-end of the sequence to be expressed to provide polyadenylation and termination. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization (Holland, cited above). Yet another
regulatable promoter is the metallothionein promoter system disclosed by Fogel et al., in U.S. Patent 4,511,652, incorporated herein by reference. Virtually any plasmid vector containing a promoter, origin of replication and termination sequences that are all yeast-compatible is suitable.


In addition to prokaryotic and eukaryotic microorganisms, expression systems comprising cells derived from multicellular organism may also be used to produce proteins of the invention. Of particular interest are mammalian expression systems because their posttranslational processing machinery is more likely to produce biologically active mammalian proteins. Several DNA tumor viruses have been used as vectors for mammalian hosts: e.g. see Tooze, ed., *DNA Tumor Viruses*, 2nd Ed. (Cold Spring Harbor Laboratory, N.Y., 1981) for a review of their biology. Particularly important are the numerous vectors which comprise SV40 replication, transcription, and/or translation control sequences coupled to bacterial replication control sequences, e.g. the pCD vectors developed by Okayama and Berg, disclosed in *Mol. Cell. Biol.*, Vol. 2, pgs. 161-170 (1982) and *Mol. Cell. Biol.*, Vol. 3, pgs. 280-289 (1983), both of which are incorporated herein by reference; the SV40 vectors disclosed by Hamer in *Genetic Engineering*, Vol. 2, pgs.
83-100 (1980), and U.S. Patent 4,599,308, both of which are incorporated herein by reference; and the vectors additionally containing adenovirus regulatory elements, disclosed by Kaufman and Sharp, in Mol. Cell. Biol., Vol. 2, pgs. 1304-1319 (1982), and Kaufman et al., in European Patent Application 8406107, both of which are incorporated herein by reference. Monkey cells are usually the preferred hosts for the above vectors. Such vectors containing the SV40 ori sequences and an intact A gene can replicate autonomously in monkey cells (to give higher copy numbers and/or more stable copy numbers than nonautonomously replicating plasmids). Moreover, vectors that contain the SV40 ori sequences without an intact A gene and can replicate autonomously to high copy numbers (but not stably) in COS7 monkey cells are described by Gluzman, Cell, Vol. 23, pgs. 175-182 (1981) and are available from the ATCC (accession no. CRL 1651). The above SV40-based vectors are also capable of transforming other mammalian cells, such as mouse L cells, by integration into the host cell DNA.

Cell. Biol., Vol. 5, pgs. 410-413 (1985), for stable transformation of various mammalian cells including human and monkey cells; (iii) vectors comprising murine polyoma virus sequences, e.g. O'Hara, J. Mol. Biol., Vol. 151, pg. 203 (1981), for transforming mouse and hamster cells; (iv) vectors carrying the dihydrofolate reductase (dhfr) gene, e.g. Alt et al., J. Biol. Chem., Vol. 253, pgs. 1357-1370 (1978), which in response to methotrexate treatment duplicates together with adjacent coding regions cointegrated into murine genomes (e.g., that of a Chinese hamster ovary (CHO) cell line deficient in dhfr activity) e.g. as described by Urlamb et al., Proc. Natl. Acad. Sci., Vol. 77, pg. 4216 (1980); and (v) cotransformation systems such as that disclosed by Axel et al. in U.S. Patent 4,399,216. Additional mammalian expression vectors can be constructed, or existing ones modified, by using various elements from available DNA tumor viruses and retroviruses, e.g. origins of replication, enhancer sequences (such as the long terminal repeat sequence from Rous sarcoma virus (RSV-LTR) disclosed by Gorman et al., Proc. Natl. Acad. Sci., Vol. 79, pgs. 6777-6781 (1982)) intron-splice sites, polyadenylation sites, and the like.

Invertebrate expression systems can also be constructed for use with the invention, e.g. the larvae of silk worm, Bombyx mori, infected by a baculovirus vector, BmNPV, described by Maeda et al. in Nature, Vol. 315, pgs. 892-894 (1985), and in Saibo Koguku, Vol. 4, pgs. 767-779 (1985).
EXAMPLES

The following examples serve to illustrate the present invention. Selection of cDNA libraries, vectors, and hosts as well as the concentration of reagents, temperatures, and the values of other variables are only to exemplify application of the present invention and are not to be considered limitations thereof.

Examples I-IV and VII describe the production of mature GM-CSF (127 amino acids long).

Example I. Construction of cDNA Libraries from Peripheral Blood Lymphocytes and T Cell Clones, Isolation of cDNA Clones, and Expression in COS7 Monkey Cells

cDNA libraries were constructed from mRNA isolated from a human clonal T cell line, designated T-7, and from human peripheral blood lymphocytes (PBLs). The libraries were constructed in the pCD plasmid according to the method of Okayama and Berg (discussed above). After a single clone was isolated from the T-7 library, the existence of an identical clone in the PBL library was confirmed.

A. Cloned Helper T Cells.

A human T cell clone designated T-7 was isolated according to the method described in Chapters 36 and 37 of Isolation, Characterization, and Utilization of T Lymphocyte Clones, Eds. Fathman and Fitch, Academic Press, New York (1982). The cell line was continuously maintained at about $0.5 \times 10^5$ cells/ml in Dulbecco's Modified Eagle (DME) medium with 10% heat-inactivated fetal calf serum, $5 \times 10^{-5}$ M 2-ME, 2 mM glutamine, non-essential amino acids, and essential vitamins conditioned
with 30% supernatants from phytohemagglutinin (PHA) stimulated human peripheral leukocytes.

B. Induction of GM-CSF Production.

T-7 cells were cultured at 5 x 10^5/ml in DME with 4% heat-inactivated fetal calf serum, 5 x 10^{-5} M 2-ME, 2mM glutamine, non-essential amino acids, essential vitamins and 4 µg/ml concanavalin A (ConA). After 4-6 hours' incubation at 37°C in 10% CO_2, the cell suspension was centrifuged at 1500 rpm for 10 minutes. The cell pellets were collected and frozen immediately at -70°C. The supernatants were filtered (Nalgene-0.22 microns) and stored at -80°C as a source of growth factors. Aliquots of the supernatant were assayed for CSF activity (see below) to verify the induction of the line by the ConA treatment.

PBLs were induced under similar conditions, except that 7 µg/ml ConA was used.

C. Isolation of mRNA.

Total cellular RNA was isolated from cells using the guanidine isothiocyanate procedure of Chirgwin, J. et al., (Biochemistry, 18:5294-5299 [1979]). Frozen cell pellets from ConA-induced T-7 cells or PBLs (4 hrs after stimulation) were suspended in guanidine isothiocyanate lysis solution. Twenty ml of lysis solution was used for 1.5 x 10^8 cells. Pellets were resuspended by pipetting, then DNA was sheared by 4 passes through a 16 gauge needle using a syringe. The lysate was layered on top of 20 ml of 5.7 M CsCl, 10 mM EDTA in 40 ml polyallomer centrifuge tube. This solution was centrifuged at 25,000 rpm in Beckman SW28 rotor (Beckman Instruments, Inc., Palo Alto, CA) for 40 hrs at 15°C. The guanidine isothiocyanate phase containing DNA was pipetted off from the top, down to the interface.
The walls of the tube and interface were washed with 2-3 ml of guanidine isothiocyanate lysis solution. The tube was cut below the interface with scissors, and the CsCl solution was decanted. RNA pellets were washed twice with cold 70% ethanol. Pellets were then resuspended in 500 µl of 10 mM Tris.HCl pH 7.4, 1 mM EDTA, 0.05% SDS. 50 µl of 3M sodium acetate was added and RNA was precipitated with 1 ml ethanol. About 0.3 mg total RNA was collected by centrifuging and the pellets washed once with cold ethanol.

Washed and dried total RNA pellet was resuspended in 900 µl of oligo (dT) elution buffer (10 mM Tris.HCl, pH 7.4, 1 mM EDTA, 0.5% SDS). RNA was heated for 3 min. at 68°C and then chilled on ice. 100 µl of 5 M NaCl was added. The RNA sample was loaded onto a 1.0 ml oligo (dT) cellulose column (Type 3, Collaborative Research, Waltham, MA) equilibrated with binding buffer (10 mM Tris.HCl pH 7.4, 1 mM EDTA, 0.5 M NaCl, 0.5% SDS.) Flow-through from the column was passed over the column twice more. The column was then washed with 20 ml binding buffer. PolyA+ mRNA was collected by washing with elution buffer. RNA usually eluted in the first 2 ml of elution buffer. RNA was precipitated with 0.1 volume 3 M sodium acetate (pH 6) and two volumes of ethanol. The RNA pellet was collected by centrifugation, washed twice with cold ethanol, and dried. The pellet was then resuspended in water. Aliquots were diluted, and absorbance at 260 nm was determined.

D. Construction of pcD cDNA Library

Parts 1) and 2) steps 1) to 5) were carried out as disclosed in PCT/US 85/02464 at pages 56 to 60, except that the pcDVI DNA (page 57, lines 3-4) was digested with NsiI endonuclease.
Step 6 (of cDNA Library Preparation):
Transformation of *E. coli*. Transformation was carried out using minor modifications of the procedure described by Cohen et al., *Proc. Natl. Acad. Sci.*, Vol. 69, pgs. 2110-2114 (1972). *E. coli* K-12 strain MC1061, described by Casabdan, M. and Cohen, S. in *J. Mol. Biol.*, Vol. 138, pgs. 179-207 (1980), was grown to 0.5 absorbancy unit at 600 nm at 37°C in 20 ml of L-broth. The cells were collected by centrifugation, suspended in 10 ml of 10 mM Tris-HCl (pH 7.3) containing 50 mM CaCl₂, and centrifuged at 0°C for 5 min. The cells were resuspended in 2 ml of the above buffer and incubated again at 0°C for 5 min.; then, 0.2 ml of the cell suspensions was mixed with 0.1 ml of the DNA solution from step 5 and the mixture was incubated at 0°C for 15 min. Next the cells were kept at 37°C for 2 min. and thereafter at room temperature for 10 min.; then 0.5 ml of L-broth was added, and the culture was incubated at 37°C for 30 min., mixed with 2.5 ml of L-broth soft agar at 42°C, and spread over L-broth agar containing 50 μg of ampicillin per ml. After incubation at 37°C for 12 to 24 hr., individual colonies were picked with sterile toothpicks. In all, approximately 5 x 10⁴ independent cDNA clones were generated.

E. Screening of the Human T-cell cDNA Library by DNA Transfections

A collection of 10⁴ independent clones were picked at random from the T-cell cDNA library and propagated individually in wells of microtiter dishes containing 200 μl L broth with ampicillin at 50 μg/ml and dimethyl sulfoxide at 7%. Pools containing 48 cDNA clones were prepared from the microtiter cultures. 40 such pools were grown up in 1 liter cultures of L-broth containing 100 μg/ml ampicillin. Plasmid DNA was
isolated from each culture and purified by twice banding through CsCl gradients. The DNA representing each pool was transfected into COS7 monkey cells as follows.

One day prior to transfection, approximately 10^6 COS7 monkey cells were seeded onto individual 60 mm plates in DME containing 10% fetal calf serum and 2 M glutamine. To perform the transfection, the medium was aspirated from each plate and replaced with 1.5 ml of DME containing 50 mM Tris-HCl (pH 7.4), 400 µg/ml DEAE-Dextran and 15 µg of the plasmid DNAs to be tested. The plates were incubated for four hours at 37°C, then the DNA-containing medium was removed, and the plates were washed twice with 2 ml of serum-free DME. DME containing 150 µM chloroquine was added back to the plates which were then incubated for an additional 3 hours at 37°C. The plates were washed once with DME, and then DME containing 4% fetal calf serum, 2 mM glutamine, penicillin and streptomycin was added. The cells were then incubated for 72 hours at 37°C. The growth medium was collected and assayed for GM-CSF activity as described above.

Four pools (groups 1A, 3B, 7A, and 14A) yielded human GM-CSF activity (see Table I below). Each group was then subdivided into 6 pools, each containing 8 of the original pooled clones. One subpool from each initial pool was positive in the transfection assay, with the exception of 7A which yielded two positive subpools. Each of the plasmids in four of the five subpools was transfected individually into COS7 cells. Four single clones, designated 3-8a, 7-1a, 7-4d, and 14-1e were active in producing GM-CSF activity. Restriction endonuclease analysis showed that all of these clones share essentially the same structure.
Table II presents the number of hemopoietic colonies stimulated by each transfection sample in duplicate cord blood assays. A cluster represents 20 to 50 cells, a small colony was 51 to 150 cells, and a colony was more than 150 cells.
TABLE II
Assay of Colony Stimulating Activity from Plasmid DNA Pools Transfections

<table>
<thead>
<tr>
<th>First Screening</th>
<th>40 pools of 48 clones (1-20, A + B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mock-infected COS7:</td>
<td>7 + 12 clusters</td>
</tr>
<tr>
<td>Pool 1A:</td>
<td>29 + 25 clusters</td>
</tr>
<tr>
<td>Pool 3B:</td>
<td>38 + 20 clusters</td>
</tr>
<tr>
<td>Pool 7A:</td>
<td>22 + 19 clusters</td>
</tr>
<tr>
<td>Pool 14A:</td>
<td>26 + 32 clusters</td>
</tr>
<tr>
<td>All other pools:</td>
<td>fewer than 20 clusters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second Screening</th>
<th>Sub-pools of 8 clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mock-infected COS7:</td>
<td>9 + 15 clusters</td>
</tr>
<tr>
<td>Sub-pool 1-5:</td>
<td>56 + 54 clusters</td>
</tr>
<tr>
<td>Sub-pool 3-8:</td>
<td>98 + 52 small colonies</td>
</tr>
<tr>
<td>Sub-pool 7-1:</td>
<td>29 + 41 small colonies</td>
</tr>
<tr>
<td>Sub-pool 7-4:</td>
<td>100 + 93 small colonies</td>
</tr>
<tr>
<td>Sub-pool 14-1:</td>
<td>40 + 73 small colonies</td>
</tr>
<tr>
<td>All other sub-pools:</td>
<td>fewer than 20 clusters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Third Screening</th>
<th>Individual clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 3 - 8a:</td>
<td>120 + 127 colonies</td>
</tr>
<tr>
<td>Clone 7 - 1a:</td>
<td>198 + 164 colonies</td>
</tr>
<tr>
<td>Clone 7 - 4a:</td>
<td>176 + 160 colonies</td>
</tr>
<tr>
<td>Clone 14 - 1e:</td>
<td>62 + 67 colonies</td>
</tr>
<tr>
<td>All other clones</td>
<td>fewer than 20 clusters</td>
</tr>
</tbody>
</table>
A plasmid (pcD-human-GM-CSF) carrying a substantially full-length cDNA insert is shown in Figure 2, and an E. coli bacterium (MC1061) carrying the plasmid has been deposited with the ATCC (accession number 39923). In Figure 2, transcription of the 776 bp cDNA insert contained in the pcD expression vector from the SV40 early promoter is indicated by the arrow. The locations of the splice donor and acceptor sites are shown. A polyadenylation signal derived from SV40 is located at the 3'-end of the cDNA insert. The GM-CSF coding region in the cDNA insert is heavily shaded, while the non-coding regions are lightly shaded. The remainder of the vector sequences are derived from pBR322, including the β-lactamase gene (AmpR) and the origin of replication.

Utilizing both the M13 dideoxy chain termination method (Sanger, F., et al., Proc. Natl. Acad. Sci., Vol. 74, pgs. 5463-5467 (1977) and modified Maxam/Gilbert technique (Rubin, C. and Schmid, C., Nucleic Acid Res., Vol. 8, pgs. 4613-4619 (1981) the 3'-8a sequence was determined. The cDNA insert contains a single open reading frame. The first ATG is found 33-35 nucleotides from the 5'-end, and is followed by 144 codons before the termination triplet (TGA) at nucleotide positions 465-467.

Table III shows a percentage breakdown of the cellular composition of about 60 human bone marrow and cord blood colonies grown under the influence of the individual clones 3-8a, 7-1a, 7-4d, and 14-1e. The existence of mixed colonies of eosinophils and the other cell types could be due to colonies growing together.
TABLE III

Cellular Composition of Human Bone Marrow Colonies

<table>
<thead>
<tr>
<th></th>
<th>Neu</th>
<th>Mφ</th>
<th>Eos</th>
<th>Neu/Mφ</th>
<th>Mφ/Eos</th>
<th>Neu/Mφ/Eos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15%</td>
<td>30%</td>
<td>7%</td>
<td>37%</td>
<td>2%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Example II. **Enhanced Expression of GM-CSF in Mammalian Cells Using the SRα Promoter**

Enhanced expression of GM-CSF in a variety of mammalian cells is obtained by the use of a promoter (designated SRα) which is constructed by inserting a portion of the long terminal repeat (LTR) of an HTLV(I) retrovirus at the Xho I site of the SV40 early origin of replication. The presence of the LTR fragment increases expression by a factor of 20-50 in COS monkey cells (e.g. available from the ATCC under accession numbers CRL 1650 or CRL1651) and CV1 monkey cells (e.g. available from the ATCC under accession number CCL 70), and in mouse L cells (e.g. L-M(TK-)) available from the ATCC under accession number CCL 1.3). Directions are provided below for constructing the SRα promoter from synthetic DNA fragments. Alternatively, the SRα promoter can be obtained from plasmid pcD-huIL-3-22-1 deposited with the ATCC under accession number 67318.


The portion of the HTLV(I) LTR inserted at the XhoI site of pLI consists of a 267 base segment including the entire R region and part of the U5 region (designated U5',...
in Figure 3) extending from the R/U5 boundary to the first downstream TagI site (39 bases in all). The sequence of the HTLV(I) LTR is disclosed by Yoshida et al. in Current Topics in Microbiology and Immunology, Vol. 115, pgs. 157-175 (1985).

The HTLV(I) LTR segment is inserted into the plasmid pLL in four steps utilizing procedures disclosed in Example IV, to form the modified plasmid, pLL', illustrated in Figure 3. First, pLL is digested with BglI and XhoI, and the large fragment is isolated. The large fragment is then mixed with synthetics 1A/B and 2A/B (defined below) in a standard ligation mixture. Synthetics 1A/B and 2A/B together consist of, in sequence, 5'-->3', the BglI/XhoI fragment of the SV40 ori, the 267-base 5'-piece of the LTR R region, and a polylinker consisting of SmaI, SacI, NaeI, and XhoI sites in the stated order.

(BglI)
TCGGCCCTCTGAGCTATTTCCAGAAGTAGTGGAGGGCTTTTGGAGGGCTCTGAGAGACACTGATAAGGCTCTTCATCACTCTCTCCGAAAAACCTTGG
GGCCTAGGCTTTT
CCG

**Synthetics 1A/B**

<table>
<thead>
<tr>
<th>SV40 ori</th>
<th>R</th>
<th>R</th>
<th>SmaI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAAAAAGCTCGGCTCGCATCTCCCTTTCAGCGCGCGGGAGGAGACACTCCGAAAACGTCTTTCGCGAGCGGAGTAGAGAGGGAGGTGCAGCCGCCCCCTC-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NaeI** | **XhoI** |
| CTGCGCGGCCC |
| GAGCGGCCGGAGCT |

**Synthetics 2A/B**
After ligation, the resulting plasmid of step 1 is propagated, isolated, digested with SmaI and SacI, and the large fragment is isolated. The large fragment is then mixed with synthetics 3A/B and 4A/B (defined below in a standard ligation mixture).

```
GCGGCCCTACCTGAGGGCCGCTGACCCGCTGTTGATCGGCGTTCT
CGGCCGGATTTGACTCCGCGGGTAGGTGCGGGCAACTC
```

**Synthetics 3A/B**

```
GCGGCCTCCGCTGGTCTGCTCTGGATCTGACGCTGGC-
AGCGCAAGAGGGGAGGCGAGACCACCGGAGACTTGTACGCAGGGC-
```

SacI
CGTCTAGGTAAGTTAGAGCT
GCAGATCCATTCAAATC

**Synthetics 4A/B**

After ligation, the resulting plasmid of step 2 is propagated, isolated, digested with SacI and NaeI, and the large fragment is isolated. The large fragment is then mixed with synthetics 5A/B (defined below) in a standard ligation mixture.

```
CAGGTAGACACGGCGCTTGTCCGGCTCCCTGAGCCTACCT-
TCGAGTCAGCTCTGGCCGAAACGAGGCGGAGGGAAACTTGGAGGA-
```

NaeI
AGACTCAGCC
TCTGAGTCGG

**Synthetics 5A/B**

After ligation, the resulting plasmid of step 3 is propagated, isolated, digested with NaeI and XhoI, and the large fragment is isolated. The large fragment is
then mixed with synthetics 6A/B and 7A/B (defined below) in a standard ligation mixture to form the modified pLI plasmid, designated pLI'.

NaeI
GGCTCTCCACGCTTTGCTGACCCTGCTTGCTCAACTCTACG
CCGAGAGGTGCCAAACCGAAGACTGGGACGAAGGAG

**Synthetics 6A/B**

XhoI
TCTTCTGTTTCTGTTTCTGTCGCCGTTACAGATC
TTGAGATGCAGAAACAAGCAAGACAGCAGCAGCAATGCTTAGAGCT

**Synthetics 7A/B**

After amplification pLI' is purified, digested with HindIII and XhoI, and the small fragment containing the SV40 ori and R-US' is isolated. Separately, plasmid pcDV1 (illustrated in Figure 4) is purified, digested with HindIII and XhoI, and the large fragment containing the pBR322 ori and AmpR gene is isolated. The small fragment of pLI' and the large fragment of pcDV1 are ligated, and the resultant plasmid pcD (SRα) is amplified. Separately plasmid pcD-human-GM-CSF is digested with XhoI and the small fragment containing the coding region for GM-CSF is isolated, and ligated with XhoI-digested pcD(SRα). The resulting recombinants molecules are transfected into E. coli, e.g. strain HB101 or MC1061, and plated. Bacteria from ampicillin-resistant colonies are randomly selected and separately amplified. Plasmids are extracted, purified, and then used to transfect COS7 monkey cells as described above. The COS7 cultures whose supernatants assay positive for GM-CSF activity carry the desired pcD (SRα)-hGM-CSF plasmids.
Example III. Expression of GM-CSF in Saccharomyces cerevisiae

The native signal peptide coding region of human GM-CSF was removed and replaced by the signal peptide coding region of the yeast α-factor mating pheromone in plasmid pMFω8. The yeast, Saccharomyces cerevisiae, was transformed by pMFω8, the mating factor/GM-CSF fusion protein was expressed, and mature GM-CSF was secreted into the culture medium. The results of this example are also described by Miyajima et al., in EMBO Journal, Vol. 5, pgs. 1193-1197 (1986).

S. cerevisiae secretes mating-type specific oligopeptide pheromones. MATα cells secrete α-factor, which induces the growth arrest of MATα cells at G1 phase of the cell cycle (Thorner, J., The Molecular Biology of the Yeast Saccharomyces, Cold Spring Harbor Laboratory, New York (1981); see particularly pages 143-180. The α-factor is initially synthesized as a larger precursor molecule consisting of an NH₂-terminal signal sequence of about 20 amino acids, followed by a leader sequence of an additional 60 amino acids, and ending with four identical tandem repeats of the mature α-factor sequence. The repeats are separated from each other by six or eight amino acids spacers (Lys-Arg-Glu-Ala-Glu-Ala and Lys-Arg-Glu-Ala-Glu[or Asp]-Ala-Glu-Ala). This prepro-α-factor is cleaved at several specific sites. The first processing is the cleavage of the COOH-terminal side of the Lys-Arg pair of the spacer sequence catalysed by the KEX2 product: Julius et al., Cell, Vol. 37, pgs. 1075-1089 (1984). A carboxypeptidase-B like enzyme cleaves at the NH₂-terminal side of the Lys-Arg pair. The final step is the removal of Glu-Ala or Asp-Ala pairs by diaminopeptidase, which is encoded by the STE13.
Brake et al., Proc. Natl. Acad. Sci., Vol. 81, pgs. 4642-4646 (1984), have shown that the fusion of the sequence encoding mature human proteins to the first processing site allowed secretion of such proteins.

A general yeast expression vector, designated pMF\(\alpha 8\), containing the alpha factor promoter and downstream leader sequence in conjunction with other elements, has been deposited with the ATCC (accession number 40140). It can be constructed as follows:

A 1.7 kb EcoRI fragment carrying the MF\(\alpha 1\) gene (Kurjan, J. and Hershowitz, I., Cell, Vol. 30, pgs. 933-943 (1982) is cloned into the EcoRI restriction site of M13mp8 (Viera, J. and Messing, J., Gene, Vol. 19, pgs. 259-268 (1982). In order to introduce a HindIII site after the lysine codon of the first spacer region, the synthetic oligonucleotide TCTTTATCCAAAGATACCC is hybridized to the single stranded M13-MF\(\alpha 1\) DNA and the oligonucleotide primer extended by DNA polymerase I Klenow fragment. After S1 nuclease treatment, the DNA is cleaved with EcoRI and the fragment carrying the MF\(\alpha 1\) promoter and leader sequence cloned into the EcoRI and filled-in HindIII restriction sites of pUC8 (Viera, J. and Messing, J., cited above). One plasmid with the desired structure can be isolated (designated pMF\(\alpha 4\Delta 1\)). The pMF\(\alpha 4\Delta 1\) is cleaved with HindIII and partially filled in with DNA polymerase I Klenow fragment in the presence of dATP and dGTP. The DNA is treated with mung bean nuclease, and the oligonucleotide linker GCCTCGAGGC attached. The resultant plasmid (designated pMF\(\alpha 5\)) will have a StuI cleavage site immediately after an arginine codon, followed by the XhoI restriction site. An S. cerevisiae-E. coli shuttle vector (pTRP584) can be constructed as follows: the PstI-XbaI fragment carrying 2 \(\mu\)m plasmid replication origin (J. Broach, cited above), is cloned into the ClaI restriction site of pTRP56
(Miyajima et al., Mol. Cell. Biol., Vol. 4, pgs. 407-414 (1984)), and the StuI restriction site within the TRP1-ARSl fragment converted to PvuII restriction site by PvuII linker insertion. The KpnI restriction site in the original pTRP56 is converted to XhoI by the XhoI linker insertion. The general secretion vector pMFα8 is then obtained by insertion of the BglII-XhoI fragment of pMFα5 into the BamHI-XhoI restriction sites of pTRP584.

Before the cloning into pMFα8, an FspI restriction site was introduced into the GM-CSF cDNA by oligonucleotide-directed mutagenesis in M13mp8 according to the procedure disclosed by Zoller and Smith in Methods in Enzymology, Vol. 100, pgs. 468-500 (1983). The introduced FspI site permits the cDNA insert to be cleaved at the junction between the endogenous signal peptide coding region and the region coding for the mature polypeptide prior to insertion into pMFα8.

Briefly, the BamHI fragment of pcD-human-GM-CSF containing the complete cDNA insert (see Figure 2A) was cloned into the replicative form of M13mp8, which was then used to transform E. coli JM101. M13mp8s carrying the insert were selected from clear plaques on agar plates containing isopropyl-beta-D-thiogalactosidase (IPTG) and 5-bromo-4-chloro-3-indolyl-beta-D-galactoside (X-gal). Single stranded M13mp8 DNA was separately prepared from eight E. coli JM101 cultures corresponding to selected plaques using standard techniques (centrifugation, removal of phage particles from supernatant with polyethylene glycol, phenol extraction to separate the DNA from the phage coats, and ethanol precipitation).

The eight sets of DNAs were grouped into two subsets according to the orientation of the cDNA insert, as determined by cross hybridization and gel electrophoresis. Oligonucleotide-directed mutagenesis
was carried out separately on single stranded DNAs from each subset. Alternatively, members from each of the two subsets could have been sequenced to determine the single stranded DNAs containing the cDNA insert in the correct orientation.

A 20-mer oligonucleotide primer was synthesized (Applied Biosystems, Inc., model 380A, Foster City, CA) having the following sequence (the mismatches are underlined):

\[ \text{GTCGTAAGACGCCGTGGGC} \]

After 3-phosphorylation, the 20-mer oligonucleotide primer was mixed in approximately 30-fold excess with the single stranded DNA (approximately 1 µg), and heated to 65°C for 5 min in 0.5 M NaCl in a 5 µl volume and allowed to stand at room temperature for 1 hour. Buffer and salt concentrations were adjusted to 100mM NaCl, 20mM Tris·HCl at pH 7.5, 10mM dithiothreitol, 10mM MgCl₂, each of the four deoxynucleoside triphosphates to about 400 µM, and ATP at about 500 µM. DNA polymerase I (Klenow fragment) and T4 DNA ligase were then added and the reaction mixture was allowed to incubate at 12°C. Double stranded closed-loop DNAs were isolated by alkaline sucrose gradient centrifugation, and used to transform CaCl₂-treated \text{E. coli} JM101. \text{E. coli} JM101 colonies were screened for the presence of the mutant cDNA by hybridization to a \(^{32}\)P-labeled oligonucleotide probe having the same sequence as that used for the primer.

After amplification, the mutant-carrying M13 phage was digested with \text{FspI} and \text{BamHI}, treated with DNA polymerase I Klenow fragment (to blunt the protruding end of the \text{BamHI} cut), and subjected to gel electrophoresis. Finally, the isolated GM-CSF-containing fragment was inserted into the \text{SuI} site of pMF8.
S. cerevisiae 20B-12 (MATα trpl-289 pep 4-3) was transformed with the pMF68 by the lithium acetate method (Ito et al., J. Bacteriol., Vol. 153, pgs. 163-168 (1983)) and grown in medium containing 0.67% yeast nitrogen base without amino acids, 0.5% casamino acids, and 2% glucose.

Colony assays with umbilical cord blood confirmed the presence of secreted GM-CSF in the yeast medium.

Example IV. **Construction of a Synthetic GM-CSF Gene for Cassette Mutagenesis in pcD and Expression in COS7 Monkey Cells**


Briefly, the synthetic human GM-CSF gene is assembled from a plurality of chemically synthesized double stranded DNA fragments. Base sequences of the synthetic gene are selected so that the assembled synthetic gene contains a series of unique restriction sites with respect to the vector carrying the gene.

The series of unique restriction sites defines a series of segments which can be readily excised and replaced with segments having altered base sequences. The synthetic fragments are inserted either directly or
after ligation with other fragments into a suitable vector, such as a pcD plasmid, or the like. The above-mentioned segments correspond to the "cassettes" of Wells et al. (cited above). The synthetic fragments are synthesized using standard techniques, e.g. Gait, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, UK, 1984). Preferably an automated synthesizer is employed, such as an Applied Biosystems, Inc. (Foster City, CA) model 380A. pcD plasmids and other suitable vectors are commercially available, e.g. from Pharmacia. For pcD, cloning can be carried out in *E. coli* MC1061 or HB101, and expression can be carried out in COS or CV1 monkey cells, Chinese hamster ovary cells, or mouse L cells.

In this example a synthetic human GM-CSF gene is constructed for insertion into a pcD plasmid. Briefly, the pcD vector is constructed by first ligating four fragments together with T4 DNA ligase: the large HindIII/BamHI fragment from pcDV1 containing the pBR322 ori and Amp<sup<r> regions, the HindIII/PstI fragment from pLl containing the SV40 early region promoter and late region introns, and synthetics 11A/B described below. Next, in a series of amplification, purification, and insertion steps the rest of the synthetics 12A/B through 19A/B are added until a complete synthetic GM-CSF gene is present in the pcD vector. The synthetic gene is made up of the sequence cassettes defined by the restriction sites SacI, NarI, BstEII, NaeI, XmaIII, BclI, MluI, SauI, ApaI, NheI, BspMII, SpeI, and SacII.

Restriction endonuclease digestions and ligase reactions are performed using standard protocols, e.g. Maniatis et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

The alkaline method (Maniatis et al., cited above) is used for small scale plasmid preparations. For
large scale preparations a modification of the alkaline method is used in which an equal volume of isopropanol is used to precipitate nucleic acids from the cleared lysate. Precipitation with cold 2.5 M ammonium acetate is used to remove RNA prior to cesium chloride equilibrium density centrifugation and detection with ethidium bromide.

Complementary strands of synthetic DNAs to be cloned (about 400ng each) are mixed and phosphorylated with polynucleotide kinase in a reaction volume of 50 μl. This DNA is ligated with 1μg of vector DNA digested with appropriate restriction enzymes, and ligations are in a volume of 50 μl at room temperature for 4 to 12 hours. Conditions for phosphorylation, restriction enzyme digestions, polymerase reactions, ligations, and bacterial transfections have been described (Maniatis et al., cited above).

In step 1, pcDV1 is amplified, isolated, and digested with HindIII and BamHI. The fragment containing the pBR322 ori and AmpR regions is isolated by gel electrophoresis using standard techniques, e.g. Maniatis et al. (cited above). pL1 is amplified, isolated and digested with HindIII and PstI. The fragment containing the SV40 early region promoter is isolated by gel electrophoresis. The two isolated pcDV1 and pL1 fragments are then mixed with synthetics 11A/B (illustrated below) in a standard T4 DNA ligase solution. Synthetics 11A/B form an insert to the newly created pcD vector, which is amplified in E. coli and then isolated.

In step 2, the isolated pcD vector of step 1 is digested with SacI and NaeI. The large fragment is separated and mixed with synthetics 12A/B and 13A/B in a standard T4 DNA ligase solution to form the pcD vector of step 2, which is amplified in E. coli MC1061 and isolated.
In step 3, the isolated pcD vector of step 2 is digested with NaeI and MluI. The large fragment is separated and mixed with synthetics 14A/B and 15A/B in a standard T4 DNA ligase solution to form the pcD vector of step 3, which is amplified in E. coli MC1061 and isolated.

In step 4, the isolated pcD vector of step 3 is digested with MluI and BspMII. The large fragment is separated and mixed with synthetics 16A/B and 17A/B in a standard T4 DNA ligase solution to form the pcD vector of step 4 which is amplified in E. coli MC1061 and isolated.

In step 5, the isolated pcD vector of step 4 is digested with BspMII and SacII. The large fragment is separated and mixed with synthetics 18A/B and 19A/B in a standard T4 DNA ligase solution to form the complete synthetic GM-CSF gene in pcD. After amplification, the isolated vector is used to transfecT COS7 monkey cells as described above.

The sequences of synthetics 11A/B through 19A/B are listed below. Locations and types of unique restriction sites are indicated above the sequences.

(PstI) SacI  NaeI  MluI  BspMII  SacII (BamHI)
GGAGCTCGCCGGCAGCCGACGTCCGGCAGCCGCGG
ACGTTCAGCCGGCGCCGTCGGAGCCCTGGCCCTAG

Synthetics 11A/B

(SacI)
CATGTGGCTGAGACCGCTGTGGCTTGGCCACTGTGGCTGCAG-
TCGAGTACACCCGACGTCTGCCGACGAGCCAAACCCGTGACACCAGGACGTGG-

NarI
ATCTCGCCGCGCGCC
TAGAGC

Synthetics 12A/B
BstEII
CGGTCAAGGAGGGGCTATGATGGG
CGGCGGGGGGAGGTGGGTCTGGGCTCGGAGGACCCCTCGTACAC
(NaeI)
AATGCCATCCAGGAGGCCC
TTACGGTAGGCTCTCGGGGCGG

Synhetics 13A/B

(NaeI) XmaIII BclI
GGCTCCTGAACCTGCTGAACGCAGATGCTGAAATGAAACAGTAGAGGTGCCAGGACTTTGGACAGTTCTCTGTGCTCGGGCTCTACTTTGTCATCTT

GATCAGA

Synhetics 14A/B

(MluI)
GAAATGTTTGACCTCCAGGAGCGACCTGCCCTACAGA
CACTAGTCATTTACAAACTGGGAGTTCTCCGGCTGGACGGATGTCTGGGC

Synhetics 15A/B

(MluI) SauI
CGGTTTTGAACTGTACAGCAGGCCCTGGGATGGCGCTCCAAGCTCAAAACCTTGACATGTTCTCCGGACTCCCCGTCGGAGTGATGTTGA

ApaI
AGGGCCCC

Synhetics 16A/B
-58-

NheI
CTTGACCATGATGCGGTAGCCACTACAAGCAGCAGCTCCCT-
TTCCGGGGAAGCTTACTACGATCGGTGTGTTCTCGTGACGGGA-

(BspMII)
CCAAGT
GGTTAGGGCC

Synthetics 17A/B

(BspMII)
CCGGAACCTCCTGTGCAACCCAGATTATCATCCTAGTTGAAAGTTTCAA-
TTGAGGAAGACGTTGGCCTGAATAGTGGAAACTTTCAAAAGTTT-

GAGAACCTGAGGAC
CTCTTG

Synthetics 18A/B

SpeI
TTTCTACTATGCTCATCCCTTTGACTTGGAGCCAGTC-
GACTTCCGAAAGATGATGACGATAGGGGAAACTGACGCCCTCGTACG-

(SacII)
CAGGAGGTGACCAGC
GTCCCTCAGCTGG

Synthetics 19A/B

Example V.  Construction of GM-CSF Mutein Leu\textsuperscript{79} and Expression in COS7 Monkey Cells

Met at amino acid position 79 is changed to Leu to form human GM-CSF mutein Leu\textsuperscript{79}. The pCD plasmid of Example IV containing the complete synthetic GM-CSF gene is digested with ApaI and NheI. The large fragment is isolated and combined with the following synthetic in a standard T4 DNA ligase solution:
The modified codon is indicated by underlining. The resulting pcD vector is amplified in E. coli MC1061, isolated, and transfected into COS7 monkey cells as described in Example I. After incubation for 3-4 days, the COS7 culture supernatants are harvested and assayed for GM-CSF activity. GM-CSF mutein Leu^79 is purified using standard procedures.

Example VI. Construction of GM-CSF Mutein Ile^116 and Expression in COS7 Monkey Cells.

Val at amino acid position 116 is changed to Ile to form human GM-CSF mutein Ile^116. The pcD plasmid of Example IV containing the complete synthetic GM-CSF gene is digested with BspMII and SacII. The large fragment is isolated and combined with the following synthetic fragments 1 and 2 in a standard T4 DNA ligase solution:

CCGGAAACTCTCCTGTGCAACCAGATTATCACCCTTTTGAAGTTTCCAAGAGAAC
TTTGAAGGCAACGTTGGGTCTAATAGTGGAACACTTTCAAAG

Fragment 1

CTGAAGGACTTTTCTGCTTATCATCCCCTTGGACTGCTGG-
TTTCTCTGGACTTCCGTAAAGACGAATAGTAGGGGAACGTGACGCC-
AGCCAGTCAGGAGTGACCGC
TCGGTCAGGTCTCTAAG

Fragment 2

The modified codon is indicated by underlining. The resulting pcD vector is amplified in E. coli MC1061, isolated, and transfected into COS7 monkey cells as
described in Example I. After incubation for 3-4 days, the COS7 culture supernatants are harvested and assayed for GM-CSF activity. GM-CSF mutein Ile116 is purified using standard procedures.

Example VII. Expression of GM-CSF in Escherichia coli


The insertion was carried out in three steps. First, the *PstI/BamHI* fragment of pcD-human-GM-CSF containing the mature GM-CSF coding region was inserted into *PstI/BamHI*-digested M13mp10 replicative form (RF), and an *EcoRI* site was introduced adjacent to, and 5'- of, the initial codon of the mature GM-CSF by site-directed mutagenesis. Secondly, the *EcoRI/BamHI* fragment of the mutated M13mp10 was inserted into *EcoRI/BamHI*-digested pIN-III-ompA2. Finally, a superfluous nine-base-pair sequence between the OmpA signal peptide coding region and the mature GM-CSF coding region in pIN-III-OmpA2 was removed by site-directed mutagenesis. In all cases, mutant forms of M13 were detected by using the synthetic primer as a probe.

All restriction enzyme digestions, DNA polymerase, kinase and ligase reactions were performed essentially as described by Maniatis et al. (cited above). Enzymes were purchased from New England Biolabs and Boehringer Mannheim. Plasmid isolations were done according to the alkaline method (small scale: Maniatis et al., cited above) or a modification of this method
(large scale: Zurawski et al., J. Immunol., Vol. 137, pgs. 3354-3360 (1986). Site-directed mutagenesis was performed on DNA fragments subcloned into a Ml3mp10 vector. Ten ng of kinased primer DNA was mixed with 50-100 ng single stranded template DNA in ligase buffer in a volume of 40 μl. To this mix 100 ng of linearized Ml3 RF was added. The reaction mix was heated for 10 min. at 95°C. Annealing occurred at room temperature for 30 min. followed by 15 min. at 4°C. dNTPs (50 μM), ligase (400 U) and Klenow (5 U) were added (total volume 50 μl) and the mixture incubated for 30 min. at 4°C followed by 1 hr at 12°C. The mixture was transformed into JM101 cells and plated on L broth (LB) plates in the presence of IPTG and X-gal. White plaques were picked into microtiter dishes in 150 μl L broth. M13-infected cells were transferred to LB plates overlayed with a lawn of JM101 cells with the use of a stamp. Following incubation for 16 hrs at 37°C, pre-wet nitrocellulose paper was laid on plaques for 1 min. Filters were soaked in 0.05 M NaOH for 3 min, followed by 15 min in neutralization buffer (3 M NaCl, 0.5 M Tris–HCl pH 7.5). After transfer to fresh neutralization buffer for another 15 min, the filters were submerged in 2xSSC. Filters were dried and baked at 80°C for 1.5 hrs. Pre-hybridization of filters was for 3 hrs at room temperature in 0.09M Tris–HCl pH 7.5, lx Denhardt's, 0.9M NaCl, 0.1mM ATP, 1mM Pi, 1mM PP, 0.5% NP40, 6mM EDTA and 0.2 mg/ml E. coli tRNA. 32p labelled probe was added and incubation continued for 16 hrs at room temperature. Filters were washed in 6xSSC, 0.1% SDS, at increasing temperatures until background was washed off as monitored by autoradiography. If necessary the SSC concentration was lowered. Single stranded DNA from positive plaques was isolated and sequenced.

DNA sequence analysis was performed using standard dideoxy procedures (Sanger et al., Proc. Natl.
DNA oligonucleotides were synthesized by phosphoramidite chemistry using an Applied Biosystems 380A synthesizer.

The PstI/BamHI fragment of pcD-human-GM-CSF was cloned into PstI- and BamHI digested Ml3mp10 RF. Single stranded DNA of this construction was prepared using standard techniques (Messing, cited above), and site-directed mutagenesis was used to introduce an EcoRI site (GAATTC) between base pairs 83 and 84 of the GM-CSF cDNA sequence illustrated in Figure 1. The following synthetic primer was used:

5'-CTGCAGCATCTCTGAATTCGACCGCCGCT-3'

The EcoRI site is indicated by underlining. After sequence confirmation of the insertion mutants, the EcoRI/BamHI fragment of Ml3 containing GM-CSF cDNA was inserted into EcoRI/BamHI digested pIN-III-OmpA2, which was amplified in E. coli JM101, isolated, and digested with XbaI and BamHI. The XbaI/BamHI fragment containing GM-CSF cDNA was inserted into XbaI/BamHI-digested Ml3mp10 RF. Single stranded DNA was isolated using standard techniques and site-directed mutagenesis was carried out using the following synthetic primer:

5'-GTACCAGGCACGCCGACCGCCTC-3'

GCTGAATTC

The gap in the primer sequence and the underlined sequence below indicate the location of the 9-base-pair deletion and the deleted sequence itself. After sequence confirmation of the deletion mutant, an XbaI/BamHI
fragment thereof was isolated and ligated to XbaI/SalI-digested pIN-III-OmpA2 to give the final construction, which was used to transform *E. coli* JM101.

Example VIII. **Purification of GM-CSF Expressed in \*Escherichia coli\***

Purification of GM-CSF from a soluble *E. coli* extract was accomplished by sequential anion-exchange, dye-ligand affinity, gel filtration, and reverse-phase chromatographies. Preferably, the sequence of chromatographies consists of QAE (quaternary aminoethyl) column chromatography, Matrex Gel Red A column chromatography, concentration by ultrafiltration and/or ammonium sulfate precipitation, Sephadex G-100 gel filtration, and either FPLC or HPLC reverse-phase chromatography. The final product by this procedure demonstrated high specific biological activity, 95-100% electrophoretic purity, a low level of pyrogens, and a low level of either *E. coli*-derived contaminants or derivatives of GM-CSF.

All operations were performed at 2-15°C, unless otherwise indicated. Protein concentration was determined at each stage by a Coomassie blue binding assay. pH measurements may vary by + 0.2 pH units and conductivity measurements may vary by + 3μS. If the expected degree of purification was not achieved in any chromatographic procedure, eluted fractions were rechromatographed on the same column, or alternatively, recycled through a previous step or a previous series of steps. An ammonium sulfate precipitation and/or ultrafiltration step may be performed to concentrate and/or store the protein; as e.g., in step C. All solvents, including those that are employed for either equilibration or elution of columns, were filtered prior
to use through a 0.2 μ membrane. USP XIX grade water was employed in the preparation of all solvents after step D.

A. Cell Killing And Protein Extractions

Cells were killed by adjusting the pH to 2.0 by first adding 85% phosphoric acid to bring the pH to 4.5 and then by adding 50% trichloroacetic acid to bring the pH to 2.0. After filtration through a 0.2 μ membrane, the conductivity was adjusted to 15-20 mS by addition of water. A low conductivity was required to minimize the amount of dilution required to allow GM-CSF to bind to the QAE column.

B. Quaternary Aminoethyl (QAE) Column Chromatography.

If required, the neutralized extract was adjusted to pH 7.5 with sodium hydroxide or hydrochloric acid, as appropriate. The conductivity of the neutralized extract was adjusted to 5-10 mS by addition of deionized water. The QAE column was equilibrated with at least 2-3 column volumes of 20 mM Tris-HCl, pH 7.5. The extract containing GM-CSF was applied to the column at a loading of not greater than 20 mg per ml of resin. The column was eluted with 10-20 column volumes of a gradient of sodium chloride in the range of 0-0.5 M dissolved in the same buffer in which the column was equilibrated. Fractions were combined based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and chromatographed in Step C.

C. Dye Affinity Chromatography

The combined fractions from the QAE column were diluted with deionized water to a conductivity of 5-10 mS and loaded onto a Red affinity column (e.g., Procion Red HE3B attached to an agarose support) that had been previously equilibrated with 2-3 column volumes of 20 mM
-65-

Tris-HCl, pH 7.5. The amount of protein loaded onto the column should not exceed 10 mg per ml of gel. The column was washed with 3-4 column volumes of the equilibration buffer. Elution was performed with between 5 and 15 volumes of a gradient of sodium chloride from 0 to 0.75 M dissolved in 20 mM Tris-HCl, pH 7.5. The fractions to be employed in Step D are combined on the basis of SDS-PAGE.

D. Concentration By Ultrafiltration and/or Ammonium Sulfate Precipitation

If the protein concentration of the combined fractions from the previous stage was less than 1.0 mg/ml, the solution was concentrated by ultrafiltration employing a 3,000-5,000 molecular weight cut-off membrane. The final protein concentration should be 1.0 mg/ml or greater. Ammonium sulfate was added to a final concentration of 60-70%. The precipitate was collected by centrifugation and then washed once with 20 mM Tris-HCl (pH 7.5) containing the same concentration of ammonium sulfate employed for precipitation. The precipitate was collected by centrifugation and dissolved in 20 mM Tris-HCl, pH 7.5.

E. Sephadex G-100 Column Chromatography

The redissolved ammonium sulfate precipitate was clarified by centrifugation, if necessary. The solution was filtered through a 0.2 μ filter and charged onto a Sephadex G-100 column equilibrated with 20 mM Tris-HCl, pH 7.5. The amount of protein loaded onto the column should not exceed 3 mg per ml of gel. The column was eluted with the same buffer. Fractions for Step F were combined based on SDS-PAGE.
F. Reversed-Phase Column Chromatography

The combined Sephadex G-100 fractions were passed through a 0.2 μ filter and applied to a reversed-phase column; as, e.g., FPLC (fast protein liquid chromatography) and HPLC (high performance liquid chromatography) columns. The column was equilibrated with 0.1% trifluoroacetic acid (TFA) and elution was performed with a gradient of 0-100% acetonitrile dissolved in 0.1% TFA. Fractions were combined based on SDS-PAGE. The solution was lyophilized and the dried powder was dissolved in 20 mM sodium phosphate, pH 7.2.

G. Dialysis Of The Purified GM-CSF

The solution from Step F was dialyzed against 20 mM sodium phosphate, pH 7.2. Two changes of buffer were performed with a minimum of 4-5 hours between changes. If required, the dialyzed solution was concentrated by ultrafiltration using a 3,000-5,000 molecular weight cut-off membrane to a protein concentration of at least 1.0 mg/ml. The solution of purified GM-CSF was passed through a 0.2 μ filter and stored frozen at about -20°C or below.

The descriptions of the foregoing embodiments of the invention have been presented for purpose of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that
the scope of the invention be defined by the claims appended hereto.

Applicants have deposited pcD-human-GM-CSF with the American Type Culture Collection, Rockville, MD, USA (ATCC), under accession number 39923. This deposit satisfies the requirements of the Budapest Treaty.
CLAIMS

1. A protein exhibiting human granulocyte-macrophage colony stimulating factor activity comprising a glycosylated or unglycosylated 2-fold substituted polypeptide selected from the set defined by the formula:

H - X(Ala) - X(Pro) - X(Ala) - X(Arg) - X(Ser) - X(Pro) -
  10
X(Ser) - X(Pro) - X(Ser) - X(Thr) - X(Gln) - X(Pro) -
Trp - X(Glu) - X(His) - X(Val) - X(Asn) - X(Ala) -
  20
X(Ile) - X(Gln) - X(Glu) - X(Ala) - X(Arg) - X(Arg) -
  30
X(Leu) - X(Leu) - X(Asn) - X(Leu) - X(Ser) - X(Arg) -
  40
X(Asp) - X(Thr) - X(Ala) - X(Ala) - X(Glu) - X(Met) -
   X(Asn) - X(Glu) - X(Thr) - X(Val) - X(Glu) - X(Val) -
X(Ile) - X(Ser) - X(Glu) - X(Met) - X(Phe) - X(Asp) -
  50
X(Leu) - X(Gln) - X(Glu) - X(Pro) - X(Thr) - Cys -
   X(Leu) - X(Gln) - X(Thr) - X(Arg) - X(Leu) - X(Glu) -
X(Leu) - X(Tyr) - X(Lys) - X(Gln) - X(Gly) - X(Leu) -
  60
X(Arg) - X(Gly) - X(Ser) - X(Leu) - X(Thr) - X(Lys) -
   X(Leu) - X(Lys) - X(Gly) - X(Pro) - X(Leu) - X(Thr) -
X(Met) - X(Met) - X(Ala) - X(Ser) - X(His) - X(Tyr) -
   80
X(Lys) - X(Gln) - X(His) - Cys - X(Pro) - X(Pro) -
X(Thr) - X(Pro) - X(Glu) - X(Thr) - X(Ser) - Cys -
   X(Ala) - X(Thr) - X(Gln) - X(Ile) - Cys - X(Ile) -
X(Phe) - X(Glu) - X(Ser) - X(Phe) - X(Lys) - X(Glu) -
   100
X(Asn) - X(Leu) - X(Lys) - X(Asp) - X(Phe) - X(Leu) -
   110
X(Leu) - X(Val) - X(Ile) - X(Pro) - X(Phe) - X(Asp) -
   120
Cys - Trp - X(Glu) - X(Pro) - X(Val) - X(Gln) -
   130
X(Glu) - OH
wherein:

X(Ser) represents Ser, Ala, Thr, Gly, or Asn;
X(Arg) represents Lys or Arg;
X(Leu) represents Leu or Val;
X(Pro) represents Ala or Pro;
X(Thr) represents Thr, Ser, or Ala;
X(Ala) represents Pro, Ala, Ser, Thr, Gly, or Val;

X(Val) represents Val, Ile, Ala, or Leu;
X(Gly) represents Gly, Ala, or Ser;
X(Ile) represents Ile, Val, or Leu;
X(Phe) represents Phe or Tyr;
X(Tyr) represents Tyr or Phe;
X(His) represents His, Gln, or Asn;
X(Gln) represents Gln, Glu, or His;
X(Asn) represents Asn, Asp, Ser, or Lys;
X(Lys) represents Lys, Arg, or Asn;
X(Asp) represents Asp, Asn, or Glu;
X(Glu) represents Glu, Gln, or Asp; and
X(Met) represents Nle, Met or Leu.

2. The protein of claim 1 wherein said polypeptide is not more than 1-fold substituted in the region defined collectively by amino acids 1-75 inclusive, 89-95 inclusive, 104-112 inclusive, and 117-127 inclusive.

3. The protein of claim 2 wherein said polypeptide is not more than 1-fold substituted.

4. The protein of claim 3 wherein said polypeptide is not substituted in the region defined collectively by amino acids 1-75 inclusive, 89-95 inclusive, 104-112 inclusive, and 117-127 inclusive.
5. The protein of claim 1 wherein:
X(Ser) represents Ser, Ala, or Thr;
X(Arg) is Arg;
X(Leu) is Leu;
X(Pro) is Pro;
X(Thr) represents Thr, Ser, or Ala;
X(Ala) represents Ala, Ser, Thr, or Gly;
X(Val) represents Val or Ile;
X(Gly) represents Gly or Ala;
X(Ile) represents Ile or Val;
X(Phe) is Phe;
X(Tyr) is Tyr;
X(His) is His;
X(Gln) is Gln;
X(Asn) represents Asn or Asp;
X(Lys) is Lys;
X(Asp) represents Asp, Glu, or Asn;
X(Glu) represents Glu or Asp; and
X(Met) is Met.

6. The protein of claim 5 wherein said polypeptide
is not more than 1-fold substituted in the region defined
collectively by amino acids 1-75 inclusive, 89-95
inclusive, 104-112 inclusive, and 117-127 inclusive.

7. The protein of claim 2 wherein said polypeptide
is not more than 1-fold substituted.

8. The protein of claim 7 wherein said polypeptide
is not substituted in the region defined collectively by
amino acids 1-75 inclusive, 89-95 inclusive, 104-112
inclusive, and 117-127 inclusive.
9. The protein of claim 8 wherein said polypeptide is unglycosylated and comprises the amino acid sequence defined by the formula:

Ala - Pro - Ala - Arg - Ser - Pro - Ser - Pro -
10
Ser - Thr - Gln - Pro - Trp - Glu - His - Val -
20
Asn - Ala - Ile - Gln - Glu - Ala - Arg - Arg -
30
Leu - Leu - Asn - Leu - Ser - Arg - Asp - Thr -
40
Ala - Ala - Glu - Met - Asn - Glu - Thr - Val -
Glu - Val - Ile - Ser - Glu - Met - Phe - Asp -
50
Leu - Gln - Glu - Pro - Thr - Cys - Leu - Gln -
60
Thr - Arg - Leu - Glu - Leu - Tyr - Lys - Gln -
70
Gly - Leu - Arg - Gly - Ser - Leu - Thr - Lys -
80
Leu - Lys - Gly - Pro - Leu - Thr - Met - Met -
Ala - Ser - His - Tyr - Lys - Gln - His - Cys -
90
Pro - Pro - Thr - Pro - Glu - Thr - Ser - Cys -
100
Ala - Thr - Gln - Ile - Ile - Thr - Phe - Glu -
110
Ser - Phe - Lys - Glu - Asn - Leu - Lys - Asp -
120
Phe - Leu - Leu - Val - Ile - Pro - Phe - Asp -
Cys - Trp - Glu - Pro - Val - Gln - Glu .

10. A protein exhibiting human granulocyte-macrophage colony stimulating factor activity comprising a glycosylated or unglycosylated 1-fold substituted, 1-fold deleted, and/or 1-fold inserted polypeptide selected from the set defined by the formula:
-72-

X(Ala) - X(Pro) - X(Ala) - X(Arg) - X(Ser) - X(Pro) -
X(Ser) - X(Pro) - X(Ser) - X(Thr) - X(Gln) - X(Pro) -
Trp - X(Glu) - X(His) - X(Val) - X(Asn) - X(Ala) -
X(Ile) - X(Gln) - X(Glu) - X(Ala) - X(Arg) - X(Arg) -
X(Leu) - X(Leu) - X(Asn) - X(Leu) - X(Ser) - X(Arg) -
X(Asn) - X(Thr) - X(Ala) - X(Ala) - X(Glu) - X(Met) -
X(Asn) - X(Thr) - X(Glu) - X(Val) - X(Glu) - X(Val) -
X(Ile) - X(Ser) - X(Glu) - X(Met) - X(Phe) - X(Asp) -
X(Leu) - X(Gln) - X(Glu) - X(Pro) - X(Thr) - Cys -
X(Leu) - X(Gln) - X(Thr) - X(Arg) - X(Leu) - X(Glu) -
X(Leu) - X(Tyr) - X(Lys) - X(Gln) - X(Gly) - X(Leu) -
X(Arg) - X(Gly) - X(Ser) - X(Leu) - X(Thr) - X(Lys) -
X(Leu) - X(Lys) - X(Gly) - X(Pro) - X(Leu) - X(Thr) -
X(Met) - X(Met) - X(Ala) - X(Ser) - X(His) - X(Tyr) -
X(Lys) - X(Gln) - X(His) - Cys - X(Pro) - X(Pro) -
X(Thr) - X(Pro) - X(Glu) - X(Thr) - X(Ser) - Cys -
X(Ala) - X(Thr) - X(Gln) - X(Ile) - X(Ile) - X(Thr) -
X(Phe) - X(Glu) - X(Ser) - X(Phe) - X(Lys) - X(Glu) -
X(Asn) - X(Leu) - X(Lys) - X(Asp) - X(Phe) - X(Leu) -
X(Leu) - X(Val) - X(Ile) - X(Pro) - X(Phe) - X(Asp) -
Cys - Trp - X(Glu) - X(Pro) - X(Val) - X(Gln) -
X(Glu)

wherein the symbols of the type X(Xaa) are as defined in claim 1.
11. The protein of claim 10 wherein said polypeptide is neither substituted, deleted, nor inserted in the region defined collectively by amino acids 1-75 inclusive, 88-96 inclusive, 104-112 inclusive, and 117-127 inclusive.

12. The protein of claim 11 wherein said polypeptide is neither substituted nor inserted.

13. A pharmaceutical composition for in vitro or in vivo stimulation of blood cell growth and development comprising a therapeutically compatible carrier and an effective amount of a protein of claim 1.

14. The pharmaceutical composition of claim 13 wherein said protein comprises a sequence of amino acids defined by the formula:

| Ala - Pro - Ala - Arg - Ser - Pro - Ser - Pro - 10 |
| Ser - Thr - Gln - Pro - Trp - Glu - His - Val - 20 |
| Asn - Ala - Ile - Gln - Glu - Ala - Arg - Arg - 30 |
| Leu - Leu - Asn - Leu - Ser - Arg - Asp - Thr - 40 |
| Ala - Ala - Glu - Met - Asn - Glu - Thr - Val - 50 |
| Glu - Val - Ile - Ser - Glu - Met - Phe - Asp - | 60 |
| Leu - Gln - Glu - Pro - Thr - Cys - Leu - Gln - | 70 |
| Thr - Arg - Leu - Glu - Leu - Tyr - Lys - Gln - | 80 |
| Gly - Leu - Arg - Gly - Ser - Leu - Thr - Lys - | 90 |
| Leu - Lys - Gly - Pro - Leu - Thr - Met - Met - | 100 |
| Ala - Ser - His - Tyr - Lys - Gln - His - Cys - | 110 |
| Pro - Pro - Thr - Pro - Glu - Thr - Ser - Cys - | 120 |
| Ala - Thr - Gln - Ile - Ile - Thr - Phe - Glu - | 130 |
15. The pharmaceutical composition of claim 14 wherein said protein has been glycosylated by a mammalian cell expression host.

16. The pharmaceutical composition of claim 14 wherein said protein is unglycosylated.

17. An expression vector for the production of human granulocyte-macrophage colony stimulating factor in a mammalian cell host, the expression vector comprising in sequence:

- an SV40 origin of DNA replication;
- an SV40 early region promoter;
- an SRα promoter;
- a splice junction and a nucleotide sequence capable of encoding human granulocyte-macrophage colony stimulating factor; and
- a polyadenylation site.

18. The expression vector of claim 17 further including in sequence after said SV40 polyadenylation site:

- a bacterial origin of replication permitting said expression vector to be cloned in a bacterial host; and
- a selectable marker permitting the identification of bacterial hosts transformed by said expression vector.
19. The expression vector of claim 18 wherein said bacterial origin of replication is the pBR322 origin of replication, said bacterial host is *Escherichia coli*, and said selectable marker confers antibiotic resistance to *Escherichia coli*.

20. The expression vector of claim 18 wherein said mammalian cell host is selected from COS monkey cells, CV1 monkey cells, mouse L cells, and Chinese hamster ovary cells.

21. A nucleic acid cassette for constructing human granulocyte-macrophage colony stimulating factor muteins in a pcD expression vector, the nucleic acid cassette being selected from double-stranded DNA fragments having a first end, a second end, and a nucleotide sequence, wherein the first end and second end are each defined by nucleotide sequences characteristic of any two restriction endonucleases selected from *SacI*, *NarI*, *BstEII*, *NaeI*, *XmaIII*, *BclI*, *MluI*, *SauI*, *Apal*, *NheI*, *BspMII*, *SpeI*, and *SacII*; and wherein the nucleotide sequence is capable of encoding a 3 to 40 amino acid segment of the 2-fold substituted polypeptide defined by the formula:

\[
\begin{align*}
H & \rightarrow X(Ala) - X(Pro) - X(Ala) - X(Arg) - X(Ser) - X(Pro) - \\
& \quad 10 \\
X(Ser) & \rightarrow X(Pro) - X(Ser) - X(Thr) - X(Gln) - X(Pro) - \\
\quad 20 \\
Trp & \rightarrow X(Glu) - X(His) - X(Val) - X(Asn) - X(Ala) - \\
\quad 30 \\
X(Ile) & \rightarrow X(Gln) - X(Glu) - X(Ala) - X(Arg) - X(Arg) - \\
X(Leu) & \rightarrow X(Leu) - X(Asn) - X(Leu) - X(Ser) - X(Arg) - \\
X(Asp) & \rightarrow X(Thr) - X(Ala) - X(Ala) - X(Glu) - X(Met) - \\
\quad 40 \\
X(Asn) & \rightarrow X(Glu) - X(Thr) - X(Val) - X(Glu) - X(Val) - 
\end{align*}
\]
-76-

X(Ile)  X(Ser)  X(Glu)  X(Met)  X(Phe)  X(Asp)  50
X(Leu)  X(Gln)  X(Glu)  X(Pro)  X(Thr)  Cys  60
X(Leu)  X(Gln)  X(Thr)  X(Arg)  X(Leu)  X(Glu)  70
X(Leu)  X(Tyr)  X(Lys)  X(Gln)  X(Gly)  X(Leu)  80
X(Arg)  X(Gly)  X(Ser)  X(Leu)  X(Thr)  X(Lys)  90
X(Leu)  X(Lys)  X(Gly)  X(Pro)  X(Leu)  X(Thr)  100
X(Met)  X(Met)  X(Ala)  X(Ser)  X(His)  X(Tyr)  110
X(Lys)  X(Gln)  X(His)  Cys  X(Pro)  X(Pro)  120
X(Thr)  X(Pro)  X(Glu)  X(Thr)  X(Ser)  Cys  130
X(Ala)  X(Thr)  X(Gln)  X(Ile)  X(Ile)  X(Thr)  140
X(Phe)  X(Glu)  X(Ser)  X(Phe)  X(Lys)  X(Glu)  150
X(Asn)  X(Leu)  X(Lys)  X(Asp)  X(Phe)  X(Leu)  160
X(Leu)  X(Val)  X(Ile)  X(Pro)  X(Phe)  X(Asp)  170
Cys  Trp  X(Glu)  X(Pro)  X(Val)  X(Gln)  180
X(Glu)  OH

wherein the symbols of the type X(Xaa) are as defined in claim 1.

22. The nucleic acid cassette of claim 21 wherein said nucleotide sequence is capable of encoding a 3 to 30 amino acid segment of the 1-fold substituted polypeptide defined by the formula:

H  X(Ala)  X(Pro)  X(Ala)  X(Arg)  X(Ser)  X(Pro)  10
X(Ser)  X(Pro)  X(Ser)  X(Thr)  X(Gln)  X(Pro)  20
Trp  X(Glu)  X(His)  X(Val)  X(Asn)  X(Ala)  30
X(Ile)  X(Gln)  X(Glu)  X(Ala)  X(Arg)  X(Arg)  40
X(Leu)  X(Leu)  X(Asn)  X(Leu)  X(Ser)  X(Arg)
X(Asp) - X(Thr) - X(Ala) - X(Ala) - X(Glu) - X(Met) -
40
X(Asn) - X(Glu) - X(Thr) - X(Val) - X(Glu) - X(Val) -
50
X(Ile) - X(Ser) - X(Glu) - X(Met) - X(Phe) - X(Asp) -
X(Leu) - X(Gln) - X(Glu) - X(Pro) - X(Thr) - Cys -
60
X(Leu) - X(Gln) - X(Thr) - X(Arg) - X(Leu) - X(Glu) -
X(Leu) - X(Tyr) - X(Lys) - X(Gln) - X(Gly) - X(Leu) -
70
X(Arg) - X(Gly) - X(Ser) - X(Leu) - X(Thr) - X(Lys) -
X(Leu) - X(Lys) - X(Gly) - X(Pro) - X(Leu) - X(Thr) -
80
X(Met) - X(Met) - X(Ala) - X(Ser) - X(His) - X(Tyr) -
90
X(Lys) - X(Gln) - X(His) - Cys - X(Pro) - X(Pro) -
X(Thr) - X(Pro) - X(Glu) - X(Thr) - X(Ser) - Cys -
100
X(Ala) - X(Thr) - X(Gln) - X(Ile) - X(Ile) - X(Thr) -
X(Phe) - X(Glu) - X(Ser) - X(Phe) - X(Lys) - X(Glu) -
110
X(Asn) - X(Leu) - X(Lys) - X(Asp) - X(Phe) - X(Leu) -
120
X(Leu) - X(Val) - X(Ile) - X(Pro) - X(Phe) - X(Asp) -
Cys - Trp - X(Glu) - X(Pro) - X(Val) - X(Gln) -
X(Glu) - OH

wherein the symbols of the type X(Xaa) are as defined in claim 5.

23. The nucleic acid cassette of claim 22 wherein said polypeptide has no substitutions in the region defined collectively by amino acids 1-75 inclusive, 89-95 inclusive, 104-112 inclusive, and 117-127 inclusive.
ACACAGAGG AAAGGCTAAA GTTCTCTGGA GAG ATG TGG CTG CAG AGC CTG CTC TTG
MET Trp Leu Gin Ser Leu Leu Leu Leu

GGC ACT GTG GCC TGC AGC ATC TCT GCA CCC GGC GCG TCG CCC AGC AGC ACG
Gly Thr Val Ala Cys Ser Ile Ser Ala Pro Ala Arg Ser Pro Ser Pro Ser Thr

CAG CCC TGG GAG CAT GTG AAT GCC ATC CAG GAG GCC CGT CTC GTG AAC CTG
Gln Pro Trp Glu His Val Asn Ala Ile Gln Glu Ala Arg Arg Leu Leu Asn Leu

AGT AGA GAC ACT GCT GGT GATG AAT GAA ACA GTA GAA GTC ATC GAA ATG
Ser Arg Asp Thr Ala Ala Glu Met Asn Glu Thr Val Glu Val Ile Ser Glu Met

TTT GAC CTC CAG GAG CCG ACC TGC CTA CAG ACC CTC TCA GAG CTC TAC AAG CAG
Phe Asp Leu Gln Glu Pro Thr Cys Leu Gln Thr Arg Leu Glu Tyr Lys Gln

GGC CTG CGG GAG AGC CTC ACC AAG CTC AAG GGC CCC TGG ACC ATG ATG GCC AGC
Gly Leu Arg Gly Ser Leu Thr Lys Leu Tyr Pro Leu Thr Met Met Ala Ser

CAC TAC AAG CAG CAC TGC CCT CCA ACC CCG GAA ACT TCC TTG GCA ACC CAG ATT
His Tyr Lys Gln His Cys Pro Thr Pro Thr Pro Glu Ser Thr Pro Ala Thr Gln Ile

ATC ACC TTT GAA AGT TTC AAA GAG AAC CTG AAG GAC TGT CTG CTT GTC ATC CCC
Ile Thr Phe Glu Ser Phe Lys Glu Asn Leu Lys Asp Phe Leu Leu Val Ile Pro

TTT GAC TGC TGG GAG GAC ACC TGG CAG GAG GAC ATGAGGTCTGG CCAAGCCCGG
Phe Asp Cys Trp Glu Val Gin Glu

GAGC GTCTCCT CTCATGAAC AAGAGCTAGA AACTCAGGAT GGTACCTTG GAGGGCAAA

GGGTTGGGCC ACAGCCATGG TGGAAGTGGC CAGGACCTGC CCTGGGCAAC CTGACCCCTGA

TACAGCCATG GCAGAAGAT GGGAATATTT TATACTGACA GAAATCAGTA ATATTTATAT

ATTTATATT TTAAAAATATT TATTTATTTA TTTATTTAAG TTTATATCC ATATTTATTC

AAGATGTTTT ACCGTAATAA TTATTTAA AATATGGCTT CTAaaaaaa FIG 1
FIG 3

HTLV(I) LTR

AmpR

EcoRI

(0.95)
PstI

RII
(0.93)

ClaI

BamHI

XhoI

(0.75)
(0.65)

BgIII

SV40 ori

AvaII
(0.77)

HindIII
(0.71)

pBR322 ori

pL1

16S splice junction

3/4
FIG 4