PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/19, C07K 14/52, A61K 38/19, C12N 15/11, C12Q 1/68, C07K 16/24

A1

(11) International Publication Number: **WO 98/11227**

(43) International Publication Date: 19 March 1998 (19.03.98)

(21) International Application Number: PCT/US97/16034

(22) International Filing Date: 10 September 1997 (10.09.97)

(30) Priority Data:
08/713,288 12 September 1996 (12.09.96) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/713,288 (CIP)
Filed on 12 September 1996 (12.09.96)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

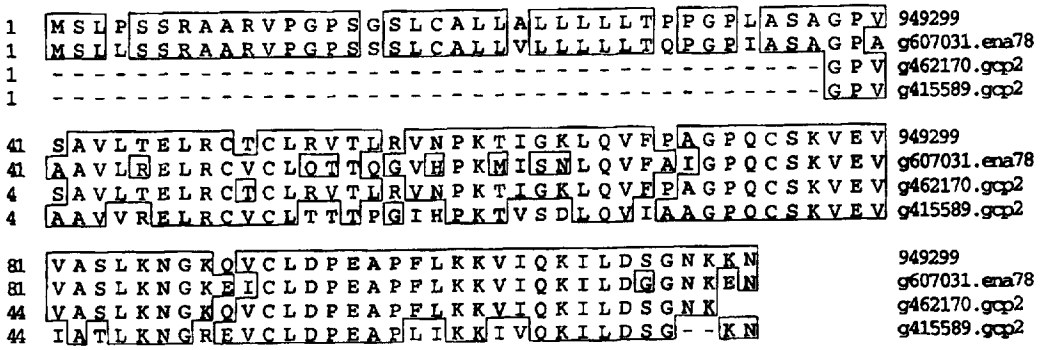
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). COLEMAN, Roger [US/US]; 260 Mariposa #2, Mountain View, CA 94041 (US). MURRY, Lynn, E. [US/US]; 1124 Los Trancos Road, Portola Valley, CA 94028 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GRANULOCYTE CHEMOTACTIC PROTEIN 2 VARIANT



(57) Abstract

The present invention provides a polynucleotide which identifies and encodes a novel human granulocyte chemotactic protein 2 variant (NGCP). The invention provides for genetically engineered expression vectors and host cells comprising the nucleic acid sequence encoding NGCP. The invention also provides for the production and use of substantially purified NGCP in pharmaceutical compositions to stop cell division in cancerous cells. The invention also describes diagnostic assays which utilize the polynucleotide to hybridize with the transcripts encoding NGCP and anti-NGCP antibodies which specifically bind to NGCP.

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.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GRANULOCYTE CHEMOTACTIC PROTEIN 2 VARIANT

The present invention relates to nucleic acid and amino acid sequences of a novel granulocyte chemotactic protein 2 variant which shares features with alpha intercrines and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

BACKGROUND ART

Chemokines are small polypeptides, generally about 70-100 amino acids in length, 6-11 kD in molecular weight, and active over a 1-100 ng/ml concentration range. Initially, they were isolated and purified from inflamed tissues and characterized relative to their bioactivity. More recently, chemokines have been discovered and produced through molecular cloning techniques and characterized by structural as well as functional analysis.

Chemokines participate in gradients of chemoattractant factors which are involved in leukocyte trafficking in different inflammatory situations. They mediate the expression of particular adhesion molecules on endothelial cells, stimulate proliferation of specific cell types, and regulate the activation of cells which bear specific receptors.

The chemokines are related through cysteine motifs and generally fall into three families, CXC chemokines (α), CC chemokines (ß) and C chemokines (g; Graves DT and Y Jiang (1995) Crit Rev Oral Biol Med 6:109-118 and Kelner G et al (1994) Science 266:1395-99). CXC chemokines include the well described platelet factor-4 as well as the nonplatelet chemokines epithelial-derived neutrophil attractant-78 (ENA-78) and granulocyte chemotactic protein-2 (GCP-2). Chemokines are generally expressed and secreted in response to proinflammatory cytokines, such as interleukin-1B and tumor necrosis factor, endotoxin, mitogens, particulates, bacteria or viruses.

ENA-78 consists of 76 amino acids and has a molecular weight of 8,357 (Walz A et al (1991) J Exp Med 174:1355-62). When endothelial cells express and secrete ENA-78, it stimulates neutrophils, chemotaxis, increased intracellular calcium, and exocytosis. The genomic DNA for ENA-78 was cloned by Corbett et al (1994; Biochem Biophys Res Commun 205:612-17) from human chromosome 4 as a 2.2 kb fragment containing four exons and three introns. The open reading frame of 342 nucleotides encodes a protein of 114 amino acids. The upstream flanking region is contains a promoter binding site for nuclear factor kappa B (Chang M et al (1994) J Biol Chem 269:41:25277-82).

GCP-2 is a 6 kD protein isolated from the supernatants of human MG63 osteosarcoma

cells. Proost et al (1993) J Immunol 150:1000-10) identified this cxc chemokine by amino acid sequencing and report that using HPLC separation, GCP-2 appears to exists in four different N-terminal forms. At concentrations of 3-10 nM, GCP-2 attracts and activates neutrophils in vitro, causes granulocyte accumulation in vivo, and has no effect on monocytes.

Chemokines and their chemoattractant activities are being studied to provide a better understanding of and the means for intervening at the molecular level in various pathogenic processes. Hirose et al (1995; Br J Cancer 72:708-714) showed that the transfection of chemokine genes into tumor cells of nude mice reduced tumorigenicity. This anti-tumor activity was attributed to the recruitment and activity of neutrophilic granulocytes to the site of the cancer.

Broxmeyer HE et al (1995; Ann Hematol 71:235-24) evaluated the myelosuppressive effects of chemokines on bone marrow and cord blood progenitor cell populations. Preincubation of the effective chemokines in acetonitril enhanced suppression of granulocyte-macrophage, erythroid and multipotential progenitor cells. These results suggested daily chemokine treatment may be useful for patients with acute or chronic leukemia.

In contrast, the suppression of chemokine expression and secretion may be useful in preventing pathogenesis in myocardial tissues. Seino Y et al (1995; Cytokine 7:301-304) used PCR to show the expression of leukocyte CC and CXC chemotactic cytokines in endomyocardial biopsies from patients with idiopathic dilated cardiomyopathy. This suggests that cytotoxic action of leukocytes recruited to the site might be contributing to inflammatory heart muscle disorders.

The discovery of a new CXC chemokines provides the opportunity to extend the knowledge of the various steps in these pathologic processes and to develop effective therapies to intervene in various forms of cancer and heart disease.

DISCLOSURE OF THE INVENTION

The present invention discloses a novel granulocyte chemotactic protein 2 variant, hereinafter referred to as NGCP, which shares features with other alpha intercrines involved in the chemoattraction and activation of leukocytes, particularly neutrophils and granulocytes. Accordingly, the invention features a substantially purified NGCP, as shown in the amino acid sequence of SEQ ID NO:1.

One aspect of the invention features isolated and substantially purified polynucleotides which encode NGCP. In a particular aspect, the polynucleotide is the nucleotide sequence of

SEQ ID NO:2. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:2.

The invention further relates to the nucleic acid sequence encoding NGCP, oligonucleotides, peptide nucleic acids, fragments, portions or antisense molecules thereof. The present invention relates, in part, to the inclusion of the nucleic acid sequence encoding NGCP in an expression vector which can be used to transform host cells or organisms. The invention also provides therapeutic transformation of cells or tissues involved in the cancers, particularly melanomas and cytokine sensitive tumors, immune deficiencies, or excessive immune responses.

The present invention also relates to a method for producing NGCP or a fragment thereof. It contemplates the delivery of purified NGCP, alone or in a pharmaceutically acceptable excipient, to cancerous cells or tissues. It also encompasses antibodies which bind specifically to NGCP and can be used to monitor testing of cytokine sensitive tumors and immune system disorders.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1A and 1B shows the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of the novel granulocyte chemotactic protein 2 variant. The alignment was produced using MacDNAsis™ software (Hitachi Software Engineering Co Ltd, San Bruno CA).

Figure 2 shows the amino acid sequence alignments among granulocyte chemotactic protein 2 variant (SEQ ID NO:1),ENA-78 (GI 607031; SEQ ID NO:3; Walz et al, supra), GCP-2 (GI 462170; SEQ ID NO:4; Proost et al, supra) and GCP-2 (GI 415589; SEQ ID NO:5; Proost et al (1993) Biochemistry 32:10170-77). These alignments were produced using the multisequence alignment program of DNAStar™ software (DNAStar Inc, Madison WI).

Figure 3 shows the hydrophobicity plot for granulocyte chemotactic protein 2 variant, SEQ ID NO:1 (MacDNAsis software); the X axis reflects amino acid position, and the negative Y axis, hydrophobicity.

Figure 4 shows the hydrophobicity plot for ENA-78, SEQ ID NO:3.

Figure 5 shows the isoelectric plot for granulocyte chemotactic protein 2 variant, SEQ ID NO:1 (MacDNAsis software).

Figure 6 shows the isoelectric plot for ENA-78, SEQ ID NO:4.

Figures 7A and 7B shows the alignment between the nucleic acid sequences of granulocyte chemotactic protein 2 variant and a human EST annotated as 5" similar to

granulocyte chemotactic protein-2 (GI 973875; Hillier L et al (1995) Unpublished; DNAStar™ software).

MODES FOR CARRYING OUT THE INVENTION

Definitions

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. Similarly, amino acid sequence as used herein refers to peptide or protein sequence.

"Consensus" as used herein may refer to a nucleic acid sequence 1) which has been resequenced to resolve uncalled bases, 2) which has been extended using XL-PCR (Perkin Elmer) in the 5' or the 3' direction and resequenced, 3) which has been assembled from the overlapping sequences of more than one Incyte clone GCG Fragment Assembly System, (GCG, Madison WI), or 4) which has been both extended and assembled.

"Peptide nucleic acid" as used herein refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen PE et al (1993) Anticancer Drug Des 8:53-63).

A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

An "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring NGCP.

A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

As used herein, NGCP refers to the amino acid sequence of substantially purified NGCP obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic or recombinant.

A "variant" of NGCP is defined as an amino acid sequence differs by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, eg, replacement of leucine with isoleucine. More rarely,

a variant may have "nonconservative" changes, eg, replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNAStar software.

The term "biologically active" refers to NGCP having structural, regulatory or biochemical functions of a naturally occurring NGCP. Likewise, "immunologically active" defines the capability of the natural, recombinant or synthetic NGCP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "derivative" as used herein refers to the chemical modification of a nucleic acid encoding NGCP or the encoded NGCP. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of natural NGCP.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

The term "hybridization" as used herein shall include "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY). Amplification is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art (Dieffenbach CW and GS Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

"Stringency" typically occurs in a range from about Tm-5°C (5°C below the Tm of the probe)to about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a stringency hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

Description

The consensus nucleotide sequence, disclosed herein, encodes a novel granulocyte chemotactic protein-2 variant of 114 amino acid residues (SEQ ID NO:1). The consensus sequence is based on the extension and assembly of Incyte clones 949299 (PANCNOT05) and

1321776 (BLADNOT04).

The nucleotide and amino acid alignments of NGCP are shown in Figures 1A and 1B. Figure 2 presents the alignment and shows the conserved cysteine residues, C_{19}, C_{49}, C_{51}, C_{75}, and C_{91}, among NGCP and the related chemokines, ENA-78 (GI 607031; SEQ ID NO:3); and GCP-2 (GI 462170; SEQ ID NO:4 and GI 415589; SEQ ID NO:5). The hydrophobicity plots of NGCP (Figure 3) and ENA-78 (Figure 4) are similar; however as shown in Figure 2, there are significant differences between the actual residues that comprise the leader and mature protein sequences of these molecules. In addition the isoelectric points of NGCP and ENA-78 are 10.34 and 9.10, respectively. Figures 7A and 7B shows the alignment and 70% identity between the nucleic acid sequences of NGCP and the human EST (GI 973875; SEQ ID NO:6) annotated as 5" similar to granulocyte chemotactic protein-2. A three frame translation of the first 180 nucleotides of this sequence reveals several frame shifts and predicts the exact leader sequence for the published ENA-78 peptide. NGCP displays the general has no obvious N-glycosylation sites.

The NGCP Coding Sequences

The nucleic acid and deduced amino acid sequences of NGCP are shown in Figures 1A and 1B. In accordance with the invention, any nucleic acid sequence which encodes NGCP can be used to generate recombinant molecules which express NGCP. In a specific embodiment described herein, a partial sequence encoding NGCP was first isolated as Incyte Clone 949299 from a pancreas cDNA library (PANCNOT05).

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of NGCP-encoding nucleotide sequences, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene may be produced. The invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence encoding naturally occurring NGCP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode NGCP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring sequence under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding NGCP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a

particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NGCP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

It is now possible to produce a DNA sequence, or portions thereof, encoding NGCP and its derivatives entirely by synthetic chemistry, after which the synthetic gene may be inserted into any of the many available DNA vectors and cell systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding NGCP or any portion thereof.

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridizing to the nucleotide sequence of Figure 1 under various conditions of stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and may be used at a defined "stringency".

Altered nucleic acid sequences encoding NGCP which may be used in accordance with the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent NGCP. The protein may also show deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NGCP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of NGCP is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine phenylalanine, and tyrosine.

Included within the scope of the present invention are alleles encoding NGCP. As used herein, an "allele" or "allelic sequence" is an alternative form of the nucleic acid sequence encoding NGCP. Alleles result from a mutation, ie, a change in the nucleic acid sequence, and generally produce altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common mutational

changes which give rise to alleles are generally ascribed to natural deletions, additions or substitutions of amino acids. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

Methods for DNA sequencing are well known in the art and employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland OH)), Taq polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the ABI 377 DNA sequencers (Perkin Elmer).

Extending the Polynucleotide Sequence

The polynucleotide sequence encoding NGCP may be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. Gobinda et al (1993; PCR Methods Applic 2:318-22) disclose "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia T et al (1988) Nucleic Acids Res 16:8186). The primers may be designed using OLIGO® 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom M et al (1991) PCR Methods Applic 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and

ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method which may be used to retrieve unknown sequence is walking PCR (Parker JD et al (1991) Nucleic Acids Res 19:3055-60), a method for targeted gene walking. Alternatively, PCR, nested primers, PromoterFinder™ (Clontech, Palo Alto CA) and PromoterFinder libraries can be used to walk in genomic DNA. This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

Capillary electrophoresis may be used to analyze either the size or confirm the nucleotide sequence in sequencing or PCR products. Systems for rapid sequencing are available from Perkin Elmer, Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity is converted to electrical signal using appropriate software (eg. Genotyper™ and Sequence Navigator™ from Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis is particularly suited to the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample. The reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min has been reported (Ruiz-Martinez MC et al (1993) Anal Chem 65:2851-8).

Expression of the Nucleotide Sequence

In accordance with the present invention, polynucleotide sequences which encode NGCP, fragments of the polypeptide, fusion proteins or functional equivalents thereof may be used in recombinant DNA molecules that direct the expression of NGCP in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence, may be used to clone and express NGCP. As will be understood by those of skill in the art, it may be advantageous to produce NGCP-encoding nucleotide sequences possessing non-naturally occurring codons. Codons

preferred by a particular prokaryotic or eukaryotic host (Murray E et al (1989) Nuc Acids Res 17:477-508) can be selected, for example, to increase the rate of NGCP expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered in order to alter NGCP-encoding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, eg, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

In another embodiment of the invention, a natural, modified or recombinant NGCP-encoding sequence may be ligated to a heterologous sequence to encode a fusion protein. For example, for screening of peptide libraries for inhibitors of NGCP activity, it may be useful to encode a chimeric NGCP protein that is recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between NGCP and the heterologous protein sequence, so that the NGCP may be cleaved and substantially purified away from the heterologous moiety.

In an alternate embodiment of the invention, the sequence encoding NGCP may be synthesized, whole or in part, using chemical methods well known in the art (see Caruthers MH et al (1980) Nuc Acids Res Symp Ser 215-23, Horn T et al(1980) Nuc Acids Res Symp Ser 225-32, etc). Alternatively, the protein itself could be produced using chemical methods to synthesize an amino acid sequence for NGCP, whole or in part. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge JY et al (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (eg, Creighton (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co, New York NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (eg, the Edman degradation procedure; Creighton, supra). Additionally the amino acid sequence of NGCP, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from

other proteins, or any part thereof, to produce a variant polypeptide.

Expression Systems

In order to express a biologically active NGCP, the nucleotide sequence encoding NGCP or its functional equivalent, is inserted into an appropriate expression vector, ie, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing a sequence encoding NGCP and appropriate transcriptional or translational controls. These methods include in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination or genetic recombination. Such techniques are described in Sambrook et al (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY and Ausubel FM et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY.

A variety of expression vector/host systems may be utilized to contain and express a sequence encoding NGCP. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (eg, baculovirus); plant cell systems transfected with virus expression vectors (eg, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (eg, Ti or pBR322 plasmid); or animal cell systems.

The "control elements" or "regulatory sequences" of these systems vary in their strength and specificities and are those nontranslated regions of the vector, enhancers, promoters, and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla CA) or pSport1 (Gibco BRL) and ptrp-lac hybrids and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (eg, heat shock, RUBISCO; and storage protein genes) or from plant viruses (eg, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from the mammalian genes or from mammalian viruses are most appropriate. If it is necessary to generate

a cell line that contains multiple copies of the sequence encoding NGCP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for NGCP. For example, when large quantities of NGCP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding NGCP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster (1989) J Biol Chem 264:5503-5509); and the like. pGEX vectors (Promega, Madison WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems are designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel et al (supra) and Grant et al (1987) Methods in Enzymology 153:516-544.

In cases where plant expression vectors are used, the expression of a sequence encoding NGCP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al (1984) Nature 310:511-514) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al (1987) EMBO J 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al (1984) EMBO J 3:1671-1680; Broglie et al (1984) Science 224:838-843); or heat shock promoters (Winter J and Sinibaldi RM (1991) Results Probl Cell Differ 17:85-105) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S or Murry LE in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill New York NY, pp 191-196 or Weissbach and Weissbach (1988) Methods for Plant Molecular Biology, Academic Press, New York NY, pp 421-463.

An alternative expression system which could be used to express NGCP is an insect

system. In one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequence encoding NGCP may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the sequence encoding NGCP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect S. frugiperda cells or Trichoplusia larvae in which NGCP is expressed (Smith et al (1983) J Virol 46:584; Engelhard EK et al (1994) Proc Nat Acad Sci 91:3224-7).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a sequence encoding NGCP may be ligated into an adenovirus transcription/ translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci 81:3655-59). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a sequence encoding NGCP. These signals include the ATG initiation codon and adjacent sequences. In cases where the sequence encoding NGCP, its initiation codon and upstream sequences are inserted into the most appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf D et al (1994) Results Probl Cell Differ 20:125-62; Bittner et al (1987) Methods in Enzymol 153:516-544).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, 293, WI38, etc have specific cellular

machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express NGCP may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I et al (1980) Cell 22:817-23) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler M et al (1980) Proc Natl Acad Sci 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin F et al (1981) J Mol Biol 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes CA et al (1995) Methods Mol Biol 55:121-131).

Identification of Transformants Containing the Polynucleotide Sequence

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the sequence encoding NGCP is inserted within a marker gene sequence, recombinant cells containing the sequence encoding NGCP can be identified by the absence of marker gene

function. Alternatively, a marker gene can be placed in tandem with the sequence encoding NGCP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem sequence as well.

Alternatively, host cells which contain the sequence encoding NGCP and expressing NGCP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of the nucleic acid or protein.

The presence of the polynucleotide sequence encoding NGCP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of the sequence encoding NGCP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the nucleic acid sequence to detect transformants containing DNA or RNA encoding NGCP. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides which can be used as a probe or amplimer.

A variety of protocols for detecting and measuring the expression of NGCP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NGCP is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton R et al (1990, Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox DE et al (1983, J Exp Med 158:1211).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the NGCP-encoding sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced as shown in US Patent No. 4,816,567 incorporated herein by reference.

Purification of NGCP

Host cells transformed with a nucleotide sequence encoding NGCP may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing sequence encoding NGCP can be designed with signal sequences which direct secretion of NGCP through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may join the sequence encoding NGCP to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins (Kroll DJ et al (1993) DNA Cell Biol 12:441-53; cf discussion of vectors infra containing fusion proteins).

NGCP may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and NGCP is useful to facilitate purification. One such expression vector provides for expression of a fusion protein comprising the sequence encoding NGCP and nucleic acid sequence encoding 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification while the enterokinase cleavage site provides a means for purifying NGCP from the fusion protein.

In addition to recombinant production, fragments of NGCP may be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al (1969) Solid-Phase Peptide

Synthesis, WH Freeman Co, San Francisco; Merrifield J (1963) J Am Chem Soc 85:2149-2154). In vitro protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance with the instructions provided by the manufacturer. Various fragments of NGCP may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Uses of NGCP

The discovery of NGCP provides the opportunity to provide a better understanding of and the means for intervening at the molecular level in various pathogenic processes. NGCP or its derivatives may be used therapeutically in the treatment of disease states such as leukemia and other cancers, immune deficiencies, or excessive immune responses.

Transfection of minute amounts of NGCP into tumor cells may reduce tumorigenicity. Purified NGCP could be administered to induce proteolytic activity of neutrophilic granulocytes resulting in tumor regression.

Since it appears that the expression of NGCP in mammals (Broxmeyer, supra) is controlled by a feedback mechanism, NGCP can be administered to a patient for its myelosuppressive effect on bone marrow progenitor cell populations. NGCP, with or without preincubation in acetonitril, would be expected to suppress granulocyte-macrophage, erythroid and multipotential progenitor cells; therefore, daily chemokine treatment may be useful for patients with acute or chronic leukemia.

A therapeutic composition comprising NGCP may have application in the prevention and treatment of individuals subject to disease conditions which compromise the immune system. Such conditions include, but are not limited to, HIV infection, Job-Buckley syndrome, lazy leukocyte syndrome, acquired agranulocytosis, and Chediak-Higashi syndrome. Although the pathologies of these diseases differ, common features include impairment of the immune system with subsequent predisposition to infections. Administration of minute amounts of NGCP might be expected to attract and activate leukocytes to the site for systemic defense.

In contrast, suppression of NGCP expression via antisense technology or introduction of antagonists, inhibitors or anti-NGCP antibodies may be useful in preventing pathogenesis. In those situations, it is useful to prevent the chemoattraction and activity of excessive leukocytes which cause cellular destruction. Such conditions include idiopathic dilated cardiomyopathy, emphysema, lupus, myasthenia gravis, pancreatitis, and rheumatoid arthritis.

NGCP Antibodies

NGCP-specific antibodies are useful for the diagnosis and treatment of conditions and diseases associated with expression of NGCP. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Neutralizing antibodies, ie, those which inhibit dimer formation, are especially preferred for diagnostics and therapeutics.

NGCP for antibody induction does not require biological activity; however, the protein fragment, or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably, they should mimic a portion of the amino acid sequence of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of NGCP amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Procedures well known in the art can be used for the production of antibodies to NGCP.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with NGCP or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal antibodies to NGCP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (1975 Nature 256:495-497), the human B-cell hybridoma technique (Kosbor et al (1983) Immunol Today 4:72; Cote et al (1983) Proc Natl Acad Sci 80:2026-2030) and the EBV-hybridoma technique (Cole et al (1985) Monoclonal Antibodies and Cancer Therapy, Alan R Liss Inc, New York NY, pp 77-96).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison et al (1984) Proc Natl Acad Sci

81:6851-6855; Neuberger et al (1984) Nature 312:604-608; Takeda et al (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies (US Patent No. 4,946,778) can be adapted to produce NGCP-specific single chain antibodies

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi et al (1989, Proc Natl Acad Sci 86: 3833-3837), and Winter G and Milstein C (1991; Nature 349:293-299).

Antibody fragments which contain specific binding sites for NGCP may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD et al (1989) Science 256:1275-1281).

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the formation of complexes between NGCP and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on a specific NGCP protein is preferred, but a competitive binding assay may also be employed. These assays are described in Maddox DE et al (1983, J Exp Med 158:1211).

Diagnostic Assays Using NGCP Specific Antibodies

Particular NGCP antibodies are useful for the diagnosis of conditions or diseases characterized by expression of NGCP or in assays to monitor patients being treated with NGCP, its fragments, agonists or inhibitors. Diagnostic assays for NGCP include methods utilizing the antibody and a label to detect NGCP in human body fluids or extracts of cells or tissues. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, the polypeptides and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known, several of which were described above.

A variety of protocols for measuring NGCP, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent

activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NGCP is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, DE et al (1983, J Exp Med 158:1211).

In order to provide a basis for diagnosis, normal or standard values for NGCP expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with antibody to NGCP under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified by comparing various artificial membranes containing known quantities of NGCP with both control and disease samples from biopsied tissues. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

Drug Screening

NGCP, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The fragment employed in such a test may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between NGCP and the agent being tested, may be measured.

Another technique for drug screening which may be used for high throughput screening of compounds having suitable binding affinity to the NGCP is described in detail in "Determination of Amino Acid Sequence Antigenicity" by Geysen HN, WO Application 84/03564, published on September 13, 1984, and incorporated herein by reference. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with fragments of NGCP and washed. Bound NGCP is then detected by methods well known in the art. Substantially purified NGCP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding NGCP specifically compete with a test compound for binding NGCP. In this manner, the antibodies can be used to detect the presence of any peptide

which shares one or more antigenic determinants with NGCP.

Uses of the Polynucleotide Encoding NGCP

A polynucleotide sequence encoding NGCP or any part thereof may be used for diagnostic and/or therapeutic purposes. For diagnostic purposes, the sequence encoding NGCP of this invention may be used to detect and quantitate gene expression in biopsied tissues in which NGCP may be expressed in response to oncogenes. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of NGCP and to monitor regulation of NGCP levels during therapeutic intervention. Included in the scope of the invention are oligonucleotide sequences, antisense RNA and DNA molecules, and peptide nucleic acids, (PNA).

Another aspect of the subject invention is to provide for hybridization or PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding NGCP or closely related molecules. The specificity of the probe, whether it is made from a highly specific region, eg, 10 unique nucleotides in the 5' regulatory region, or a less specific region, eg, especially in the 3' region, and the stringency of the hybridization or amplification (maximal, high, intermediate or low) will determine whether the probe identifies only naturally occurring NGCP, alleles or related sequences.

Probes may also be used for the detection of related sequences and should preferably contain at least 50% of the nucleotides from any of these sequences encoding NGCP. The hybridization probes of the subject invention may be derived from the nucleotide sequence of SEQ ID NO:2 or from genomic sequence including promoter, enhancer elements and introns of the naturally occurring sequence encoding NGCP. Hybridization probes may be labeled by a variety of reporter groups, including radionuclides such as 32P or 35S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Other means for producing specific hybridization probes for DNAs include the cloning of nucleic acid sequences encoding NGCP or NGCP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides.

<u>Diagnostic Use</u>

Polynucleotide sequences encoding NGCP may be used for the diagnosis of conditions or diseases with which the expression of NGCP is associated. For example, polynucleotide

sequences encoding NGCP may be used in hybridization or PCR assays of fluids or tissues from biopsies to detect NGCP expression. The form of such qualitative or quantitative methods may include Southern or northern analysis, dot blot or other membrane-based technologies; PCR technologies; dip stick, pin, chip and ELISA technologies. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits.

The NGCP-encoding nucleotide sequences disclosed herein provide the basis for assays that detect activation or induction associated with inflammation or disease. The nucleotide sequence may be labeled by methods known in the art and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After an incubation period, the sample is washed with a compatible fluid which optionally contains a dye (or other label requiring a developer) if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye in the biopsied or extracted sample is significantly elevated over that of a comparable control sample, the nucleotide sequence has hybridized with nucleotide sequences in the sample, and the presence of elevated levels of nucleotide sequences encoding NGCP in the sample indicates the presence of the associated inflammation and/or disease.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regime in animal studies, in clinical trials, or in monitoring the treatment of an individual patient. In order to provide a basis for the diagnosis of disease, a normal or standard profile for NGCP expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with NGCP, or a portion thereof, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained for normal subjects with a dilution series of NGCP run in the same experiment where a known amount of substantially purified NGCP is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients affected by NGCP-associated diseases. Deviation between standard and subject values establishes the presence of disease.

Once disease is established, a therapeutic agent is administered and a treatment profile is generated. Such assays may be repeated on a regular basis to evaluate whether the values in the profile progress toward or return to the normal or standard pattern. Successive treatment profiles may be used to show the efficacy of treatment over a period of several days or several months.

PCR, may be used as described in US Patent Nos. 4,683,195 and 4,965,188 provides

additional uses for oligonucleotides based upon the sequence encoding NGCP. Such oligomers are generally chemically synthesized, but they may be generated enzymatically or produced from a recombinant source. Oligomers generally comprise two nucleotide sequences, one with sense orientation (5'->3') and one with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Additionally, methods which may be used to quantitate the expression of a particular molecule include radiolabeling (Melby PC et al 1993 J Immunol Methods 159:235-44) or biotinylating (Duplaa C et al 1993 Anal Biochem 229-36) nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated. Quantitation of multiple samples may be speeded up by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation. A definitive diagnosis of this type may allow health professionals to begin aggressive treatment and prevent further worsening of the condition. Similarly, further assays can be used to monitor the progress of a patient during treatment. Furthermore, the nucleotide sequences disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, and the like.

Therapeutic Use

Based upon its homology to EMAP-II and its expression profile, the polynucleotide encoding NGCP disclosed herein may be useful in the treatment of immune deficiency diseases.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express antisense of the sequence encoding NGCP. See, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra).

The polynucleotides comprising full length cDNA sequence and/or its regulatory elements enable researchers to use the sequence encoding NGCP as an investigative tool in sense (Youssoufian H and HF Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991)

Annu Rev Biochem 60:631-652) regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Genes encoding NGCP can be turned off by transfecting a cell or tissue with expression vectors which express high levels of a desired NGCP fragment. Such constructs can flood cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector (Mettler I, personal communication) and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the control regions of the sequence encoding NGCP, ie, the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, eg, between -10 and +10 regions of the leader sequence, are preferred. The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA were reviewed by Gee JE et al (In: Huber BE and BI Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco NY).

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of the sequence encoding NGCP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding NGCP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Methods for introducing vectors into cells or tissues include those methods discussed infra and which are equally suitable for in vivo, in vitro and ex vivo therapy. For ex vivo therapy, vectors are introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient is presented in US Patent Nos. 5,399,493 and 5,437,994, disclosed herein by reference. Delivery by transfection and by liposome are quite well known in the art.

Furthermore, the nucleotide sequences encoding NGCP disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

Detection and Mapping of Related Polynucleotide Sequences

The nucleic acid sequence encoding NGCP can also be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include in situ hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes,

bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price CM (1993; Blood Rev 7:127-34) and Trask BJ (1991; Trends Genet 7:149-54).

The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY. Fluorescent in situ hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a the sequence encoding NGCP on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers are invaluable in extending genetic maps. A recent example of an STS based map of the human genome was recently published by the Whitehead-MIT Center for Genomic Research (Hudson TJ et al (1995) Science 270:1945-1954). Often the placement of a gene on the chromosome of another mammalian species such as mouse (Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Database Release 10, April 28, 1995) may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once a disease or syndrome, such as ataxia telangiectasia (AT), has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti et al (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier or affected individuals.

Pharmaceutical Compositions

The present invention relates to pharmaceutical compositions which may comprise nucleotides, proteins, antibodies, agonists, antagonists, or inhibitors, alone or in combination

with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone, or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with excipient(s) or pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

Administration of Pharmaceutical Compositions

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of "Remington's Pharmaceutical Sciences" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene

glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, ie, dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Manufacture and Storage

The pharmaceutical compositions of the present invention may be manufactured in a manner that known in the art, eg, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5 that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of NGCP, such labeling would include amount, frequency and method of administration.

Therapeutically Effective Dose

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, eg, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, eg, ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account include the severity of the disease state, eg, tumor size and location; age, weight and gender of the patient; diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long acting pharmaceutical compositions might be administered

every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature. See US Patent Nos. 4,657,760; 5,206,344; or 5,225,212. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

It is contemplated, for example, that NGCP can be used as a therapeutic molecule to attract granulocytes to cancers for the purpose of destroying the cancerous cells or tissues.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

INDUSTRIAL APPLICABILITY

I PANCNOT05 cDNA Library Construction

The PANCNOT05 cDNA library was constructed from human pancreas tissue. The donor was a 2 year old Hispanic male who died from anoxia. The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury NJ) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.0, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water and DNase treated at 37°C. The RNA extraction was repeated with acid phenol pH 4.0 and precipitated with sodium acetate and ethanol as before. The mRNA was then isolated using the Qiagen Oligotex kit (QIAGEN Inc; Chatsworth CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Catalog #18248-013; Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (Catalog #275105-01; Pharmacia), and those cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5a™ competent cells (Catalog #18258-012; Gibco/BRL).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL Prep 96 Plasmid

Kit for Rapid Extraction Alkaline Lysis Plasmid Minipreps (Catalog #26173; QIAGEN, Inc). This kit enables the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, LIFE TECHNOLOGIES™) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger F and AR Coulson (1975; J Mol Biol 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown MA) and Applied Biosystems 377 DNA Sequencing Systems; and the reading frame was determined.

III Homology Searching of cDNA Clones and Their Deduced Proteins

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT- 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (TRW Inc, Los Angeles CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT™ 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul SF (1993) J Mol Evol 36:290-300; Altschul, SF et al (1990) J Mol Biol 215:403-10), was used to search for local

sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches which do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al. supra).

Analogous computer techniques using BLAST (Altschul SF 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte, Palo Alto CA). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

and it takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

V Extension of the Sequence Encoding NGCP

The nucleic acid sequence of SEQ ID NO:2 is used to design oligo-nucleotide primers for

extending a partial nucleotide sequence to full length or for obtaining 5'sequence from genomic libraries. One primer is synthesized to initiate extension in the antisense direction (XLR) and the other is synthesized to extend sequence in the sense direction (XLF). Primers allow the extension of the know sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest (US Patent Application 08/487,112, filed June 7, 1995, specifically incorporated by reference). The initial primers are designed from the cDNA using OLIGO® 4.06 Primer Analysis Software (National Biosciences), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

The original, selected cDNA libraries, or a human genomic library are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired, additional sets of primers are designed to further extend the known region.

By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR is performed using the Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5-10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were selected and cut out of the gel. Further purification involves using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc). After recovery of the DNA, Klenow enzyme was used to trim

single-stranded, nucleotide overhangs creating blunt ends which facilitate religation and cloning.

After ethanol precipitation, the products are redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase are added, and the mixture is incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) are transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook J et al, supra). After incubation for one hour at 37° C, the whole transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook J et al, supra) containing 2xCarb. The following day, several colonies are randomly picked from each plate and cultured in 150 μl of liquid LB/2xCarb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample is transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction are added to each well. Amplification is performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products are compared to the original partial cDNAs, and appropriate clones are selected, ligated into plasmid and sequenced.

VI Labeling and Use of Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 mCi of [γ-^{32}P] adenosine triphosphate (Amersham, Chicago IL) and T4 polynucleotide kinase (DuPont NEN®, Boston MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 super

fine resin column (Pharmacia). A portion containing 10^7 counts per minute of each of the sense and antisense oligonucleotides is used in a typical membrane based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale CA) for several hours, hybridization patterns are compared visually.

VII Antisense Molecules

The sequence encoding NGCP, or any part thereof, is used to inhibit in vivo or in vitro expression of naturally occurring sequence. Although use of antisense oligonucleotides, comprising about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. An oligonucleotide complementary to the coding sequence of NGCP as shown in Figures 1A and 1B is used to inhibit expression of naturally occurring sequence. The complementary oligonucleotide is designed from the most unique 5' sequence as shown in Figures 1A and 1B and used either to inhibit transcription by preventing promoter binding to the upstream nontranslated sequence or translation of a transcript encoding NGCP by preventing the ribosome from binding. Using an appropriate portion of the leader and 5' sequence of SEQ ID NO:2, an effective antisense oligonucleotide includes any 15-20 nucleotides spanning the region which translates into the signal or early coding sequence of the polypeptide as shown in Figures 1A and 1B.

VIII Expression of NGCP

Expression of NGCP is accomplished by subcloning the cDNAs into appropriate vectors and transfecting the vectors into host cells. In this case, the cloning vector, pSport, previously used for the generation of the cDNA library is used to express NGCP in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met and the subsequent 7 residues of ß-galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transfected bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first seven residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length NGCP. The signal sequence directs the secretion of NGCP into the bacterial growth media which can be used directly in the following assay for activity.

IX Assay for NGCP Activity

The chemotactic activity of NGCP is measured under agarose according to the method of Nelson et al (1975; J Immunol 115:1650). Mononuclear cells or granulocytes are exposed to serial dilutions of control media or media in which cells expressing recombinant NGCP were grown. After 2 hours incubation at 37°C, cells are fixed and stained. Spontaneous migration toward the control sample is compared to migration toward the test sample for various cell types.

The specificity of the chemoattraction is determined by performing the agarose assay on specific populations of cells. Blood cells obtained from venipuncture are fractionated by density gradient centrifugation and the chemotactic activity of NGCP is tested on enriched populations of neutrophils, peripheral blood mononuclear cells, granulocytes, monocytes and lymphocytes. Optionally, such enriched cell populations may be further fractionated using $CD8^-$ and $CD4^+$ specific antibodies for negative selection of $CD4^+$ and $CD8^+$ enriched T-cell populations, respectively.

X Production of NGCP Specific Antibodies

NGCP is substantially purified using PAGE electrophoresis (Sambrook, supra) is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence translated from NGCP is analyzed using DNAStar software (DNAStar Inc) to determine regions of high immunogenicity and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Analysis to select appropriate epitopes, such as those near the C-terminus or in hydrophilic regions (shown in Figures 4 and 5) is described by Ausubel FM et al (supra).

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma) by reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel FM et al, supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit

antisera, washing, and reacting with radioiodinated, goat anti-rabbit IgG.

XI Purification of Naturally Occurring NGCP Using Specific Antibodies

Naturally occurring or recombinant NGCP is substantially purified by immunoaffinity chromatography using antibodies specific for NGCP. An immunoaffinity column is constructed by covalently coupling NGCP antibody to an activated chromatographic resin such as CnBr-activated Sepharose (Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Membrane fractions from cells expressing NGCP are prepared by methods well known in the art. Alternatively, a recombinant NGCP fragment containing an appropriate signal sequence may be secreted in useful quantity into the medium in which transfected cells are grown.

The NGCP-containing preparation is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NGCP (eg, high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NGCP binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope such as urea or thiocyanate ion), and NGCP is collected.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF THE INVENTION: NOVEL GRANULOCYTE CHEMOTACTIC PROTEIN 2 VARIANT

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
(B) STREET: 3174 Porter Drive
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: U.S.
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) PCT APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Filed Herewith

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/713,288
(B) FILING DATE: 12-SEP-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Billings, Lucy J.
(B) REGISTRATION NUMBER: 36,749
(C) REFERENCE/DOCKET NUMBER: PF-0121 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 650-855-0555
(B) TELEFAX: 650-845-4166

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 500 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY:
(B) CLONE: Consensus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Gly Ala Ala Cys Cys Ala Ala Ala Gly Thr Gly Cys Thr Cys Thr Gly
 1               5                  10                  15
Thr Ala Thr Cys Cys Thr Cys Cys Ala Gly Thr Cys Thr Cys Cys Gly
            20                  25                  30
Cys Gly Cys Cys Thr Cys Cys Ala Cys Cys Cys Ala Gly Cys Thr Cys
        35                  40                  45
Ala Gly Gly Ala Ala Cys Cys Cys Gly Cys Gly Ala Ala Cys Cys Cys
    50                  55                  60
Thr Cys Thr Cys Thr Thr Gly Ala Cys Cys Ala Cys Thr Ala Thr Gly
65                  70                  75                  80
Ala Gly Cys Cys Thr Cys Cys Cys Gly Thr Cys Cys Ala Gly Cys Cys
                85                  90                  95
Gly Cys Gly Cys Gly Gly Cys Cys Cys Gly Thr Gly Thr Cys Cys Cys
            100                 105                 110
Gly Gly Gly Thr Cys Cys Thr Thr Cys Gly Gly Gly Cys Thr Cys Cys
        115                 120                 125
Thr Thr Gly Thr Gly Cys Gly Cys Gly Cys Thr Gly Cys Thr Cys Gly
    130                 135                 140
Cys Gly Cys Thr Gly Cys Thr Gly Cys Thr Cys Cys Thr Gly Cys Thr
145                 150                 155                 160
Gly Ala Cys Gly Cys Cys Gly Cys Cys Gly Gly Gly Gly Cys Cys Cys
                165                 170                 175
Cys Thr Cys Gly Cys Cys Ala Gly Cys Gly Cys Thr Gly Gly Thr Cys
            180                 185                 190
Cys Thr Gly Thr Cys Thr Cys Thr Gly Cys Thr Gly Thr Gly Cys Thr
        195                 200                 205
Gly Ala Cys Ala Gly Ala Gly Cys Thr Gly Cys Gly Thr Thr Gly Cys
    210                 215                 220
Ala Cys Thr Thr Gly Thr Thr Thr Ala Cys Gly Cys Gly Thr Thr Ala
225                 230                 235                 240
Cys Gly Cys Thr Gly Ala Gly Ala Gly Thr Ala Ala Ala Cys Cys Cys
                245                 250                 255
Cys Ala Ala Ala Ala Cys Gly Ala Thr Thr Gly Gly Thr Ala Ala Ala
            260                 265                 270
Cys Thr Gly Cys Ala Gly Gly Thr Gly Thr Thr Cys Cys Cys Cys Gly
        275                 280                 285
Cys Ala Gly Gly Cys Cys Cys Gly Cys Ala Gly Thr Gly Cys Thr Cys
    290                 295                 300
Cys Ala Ala Gly Gly Thr Gly Gly Ala Ala Gly Thr Gly Gly Thr Ala
305                 310                 315                 320
Gly Cys Cys Thr Cys Cys Cys Thr Gly Ala Ala Gly Ala Ala Cys Gly
                325                 330                 335
Gly Gly Ala Ala Gly Cys Ala Ala Gly Thr Thr Thr Gly Thr Cys Thr
            340                 345                 350
Gly Gly Ala Cys Cys Cys Gly Gly Ala Ala Gly Cys Cys Cys Cys Thr
        355                 360                 365
Thr Thr Thr Cys Thr Ala Ala Ala Gly Ala Ala Ala Gly Thr Cys Ala
    370                 375                 380
Thr Cys Cys Ala Gly Ala Ala Ala Ala Thr Thr Thr Thr Gly Gly Ala
385                 390                 395                 400
Cys Ala Gly Thr Gly Gly Ala Ala Ala Cys Ala Ala Gly Ala Ala Ala
                405                 410                 415
Ala Ala Cys Thr Gly Ala Gly Thr Ala Ala Cys Ala Ala Ala Ala Ala
            420                 425                 430
Ala Gly Ala Cys Cys Ala Thr Gly Cys Ala Thr Cys Ala Thr Ala Ala
        435                 440                 445
Ala Ala Thr Thr Gly Cys Cys Cys Ala Gly Thr Cys Thr Thr Cys Ala
    450                 455                 460
```

```
Gly Cys Gly Gly Ala Gly Cys Ala Gly Thr Thr Thr Cys Thr Gly
465                 470                 475                 480
Gly Ala Gly Ala Thr Cys Cys Cys Thr Gly Gly Ala Cys Cys Cys Ala
                485                 490                 495
Gly Thr Ala Ala
            500
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY:
(B) CLONE: Consensus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
MSSSRAARVG SGSCAATGAS AGVSAVTRCT CRVTRVNKTG KVAGCSKVVV ASKNGKVCDA      60
KKVKDSGNKK N                                                           71
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 607031

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ser Leu Leu Ser Ser Arg Ala Ala Arg Val Pro Gly Pro Ser Ser
 1               5                  10                  15
Ser Leu Cys Ala Leu Leu Val Leu Leu Leu Leu Leu Thr Gln Pro Gly
            20                  25                  30
Pro Ile Ala Ser Ala Gly Pro Ala Ala Ala Val Leu Arg Glu Leu Arg
        35                  40                  45
Cys Val Cys Leu Gln Thr Thr Gln Gly Val His Pro Lys Met Ile Ser
    50                  55                  60
Asn Leu Gln Val Phe Ala Ile Gly Pro Gln Cys Ser Lys Val Glu Val
65                  70                  75                  80
Val Ala Ser Leu Lys Asn Gly Lys Glu Ile Cys Leu Asp Pro Glu Ala
                85                  90                  95
Pro Phe Leu Lys Lys Val Ile Gln Lys Ile Leu Asp Gly Gly Asn Lys
            100                 105                 110
Glu Asn
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 462170

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Gly Pro Val Ser Ala Val Leu Thr Glu Leu Arg Cys Thr Cys Leu Arg
 1               5                  10                  15
Val Thr Leu Arg Val Asn Pro Lys Thr Ile Gly Lys Leu Gln Val Phe
            20                  25                  30
Pro Ala Gly Pro Gln Cys Ser Lys Val Glu Val Val Ala Ser Leu Lys
        35                  40                  45
Asn Gly Lys Gln Val Cys Leu Asp Pro Glu Ala Pro Phe Leu Lys Lys
    50                  55                  60
Val Ile Gln Lys Ile Leu Asp Ser Gly Asn Lys
65                  70                  75
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 415589

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Gly Pro Val Ala Ala Val Val Arg Glu Leu Arg Cys Val Cys Leu Thr
 1               5                  10                  15
Thr Thr Pro Gly Ile His Pro Lys Thr Val Ser Asp Leu Gln Val Ile
            20                  25                  30
Ala Ala Gly Pro Gln Cys Ser Lys Val Glu Val Ile Ala Thr Leu Lys
        35                  40                  45
Asn Gly Arg Glu Val Cys Leu Asp Pro Glu Ala Pro Leu Ile Lys Lys
    50                  55                  60
Ile Val Gln Lys Ile Leu Asp Ser Gly Lys Asn
65                  70                  75
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 477 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 973875

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATCTTCGCT CCTCCAATCT CCGCTCCTCC ACCCAGTTCA GGAACCCGCG ACCGCTCGCA 60
GCCTCTCTTG ACCACTATGA GCCTCCTGTC CAGCCGCGCG CCCGTGTCCC CGGTCCTTCG 120
AGCTCCTTGT GCGCGCTGTT GGTGCTGCTG CTGCTGCTGA CGTACCAGGG CCCATCGTCA 180
GCGCTGGTCC TGCCGCTGCT GTGTTGAGAG AGCTGCGTTG CGTTTGTTTA CAGACCACGN 240
AGGGAGTTCA TCCCAAAATG ATCAGTAATC TGCAAGTGTT CGNCATAGGG CCACAGTTGT 300
TNCAAGGTTG AAGTGGTAGC CTTNCTGAAG AACGGGAAGG AAATTTTGTT CTTGATTNCA 360
GAAGACCNTT TTTTNTAAAG ANAGTCATTN CAGAAAATTT TTGGGACGGT NGGAAACAAG 420
GNATAATTGA TTAAGGAGAA ATGAGGCACG NATGGGAAAA GTTTTNCCAG TTTTTCA 477

CLAIMS

1. A substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1, or fragments thereof.

2. An isolated and purified polynucleotide sequence encoding the polypeptide of claim 1.

3. An isolated polynucleotide sequence comprising the nucleic acid sequence of SEQ ID NO:2 or variants thereof.

4. An isolated polynucleotide sequence consisting of the complement of SEQ ID NO:2.

5. An isolated polynucleotide sequence which hybridizes under stringent conditions to SEQ ID NO:2.

6. A hybridization probe comprising SEQ ID NO:2, or fragments thereof.

7. A recombinant expression vector containing the polynucleotide sequence of claim 3.

8. A recombinant host cell containing the expression vector of claim 7.

9. A method for producing the polypeptide of SEQ ID NO:1, or fragments thereof, the method comprising the steps of:

a) culturing the host cell of claim 8 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

10. A pharmaceutical composition comprising the polypeptide of SEQ ID NO:1 in conjunction with a pharmaceutically acceptable excipient.

11. A method for the treatment of cancer, the method comprising administering to a subject an amount of the pharmaceutical composition of claim 10 which is sufficient to treat the cancer.

12. A purified antibody which binds specifically to the polypeptide of claim 1.

```
                9          18          27          36          45          54
5' NGA ACC AAA GTG CTC TGT ATC CTC CAG TCT CCG CGC CTC CAC CCA GCT CAG GAA

               63          72          81          90          99         108
   CCC GCG AAC CCT CTC TTG ACC ACT ATG AGC CTC CCG TCC AGC CGC GCG GCC CGT
                                   M   S   L   P   S   S   R   A   A   R

              117         126         135         144         153         162
   GTC CCG GGT CCT TCG GGC TCC TTG TGC GCG CTG CTC GCG CTG CTG CTC CTG CTG
   V   P   G   P   S   G   S   L   C   A   L   L   A   L   L   L   L   L

              171         180         189         198         207         216
   ACG CCG CCG GGG CCC CTC GCC AGC GCT GGT CCT GTC TCT GCT GTG CTG ACA GAG
   T   P   P   G   P   L   A   S   A   G   P   V   S   A   V   L   T   E

              225         234         243         252         261         270
   CTG CGT TGC ACT TGT TTA CGC GTT ACG CTG AGA GTA AAC CCC AAA ACG ATT GGT
   L   R   C   T   C   L   R   V   T   L   R   V   N   P   K   T   I   G

              279         288         297         306         315         324
   AAA CTG CAG GTG TTC CCC GCA GGC CCG CAG TGC TCC AAG GTG GAA GTG GTA GCC
   K   L   Q   V   F   P   A   G   P   Q   C   S   K   V   E   V   V   A

              333         342         351         360         369         378
   TCC CTG AAG AAC GGG AAG CAA GTT TGT CTG GAC CCG GAA GCC CCT TTT CTA AAG
   S   L   K   N   G   K   Q   V   C   L   D   P   E   A   P   F   L   K
```

FIGURE 1A

```
         387          396          405          414          423          432
AAA GTC ATC CAG AAA ATT TTG GAC AGT GGA AAC AAG AAA AAC TGA GTA ACA AAA
K   V   I   Q   K   I   L   D   S   G   N   K   K   N   *

         441          450          459          468          477          486
AAG ACC ATG CAT CAT AAA ATT GCC CAG TCT TCA GCG GAG CAG TTT TCT GGA GAT


         495          504          513
CCC TGG ACC CAG TAA AAA AAA AAA AAA  3'
```

FIGURE 1B

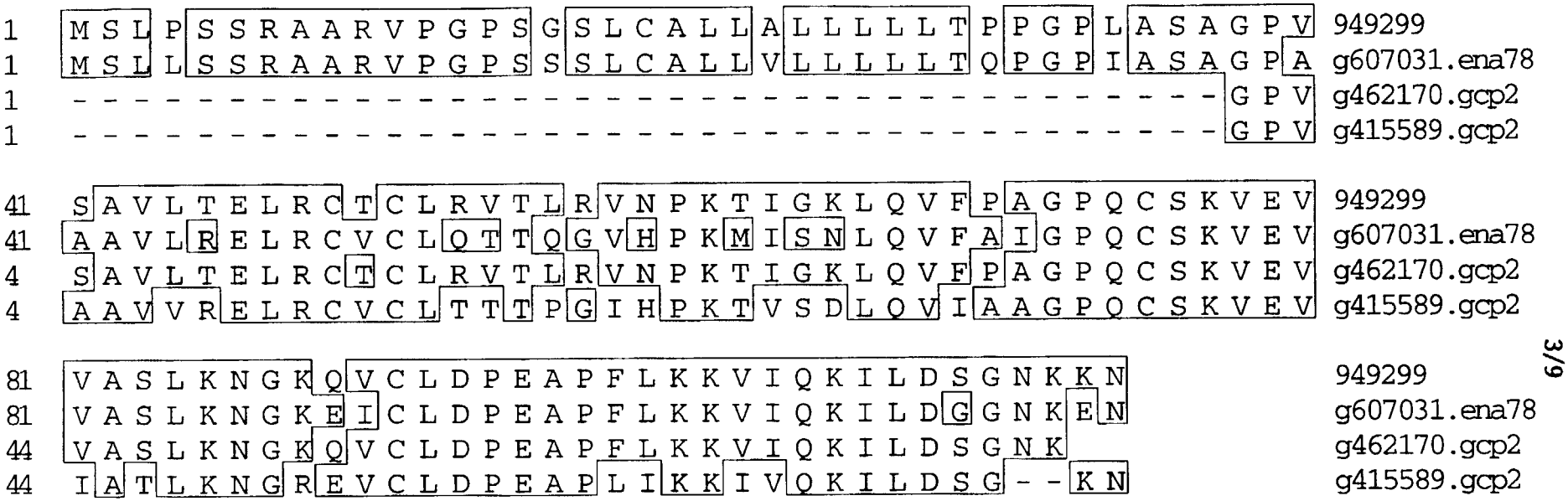

1 MSLPSSRAARVPGPSGSLCALLALLLLLTPPGPLASAGPV 949299
1 MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPA g607031.ena78
1 ---------- GPV g462170.gcp2
1 ---------- GPV g415589.gcp2
41 SAVLTELRCTCLRVTLRVNPKTIGKLQVFPAGPQCSKVEV 949299
41 AAVLRELRCVCLQTTQGVHPKMISNLQVFAIGPQCSKVEV g607031.ena78
4 SAVLTELRCTCLRVTLRVNPKTIGKLQVFPAGPQCSKVEV g462170.gcp2
4 AAVVRELRCVCLTTTPGIHPKTVSDLQVIAAGPQCSKVEV g415589.gcp2
81 VASLKNGKQVCLDPEAPFLKKVIQKILDSGNKKN 949299
81 VASLKNGKEICLDPEAPFLKKVIQKILDGGNKEN g607031.ena78
44 VASLKNGKQVCLDPEAPFLKKVIQKILDSGNK g462170.gcp2
44 IATLKNGREVCLDPEAPLIKKIVQKILDSG--KN g415589.gcp2

FIGURE 2

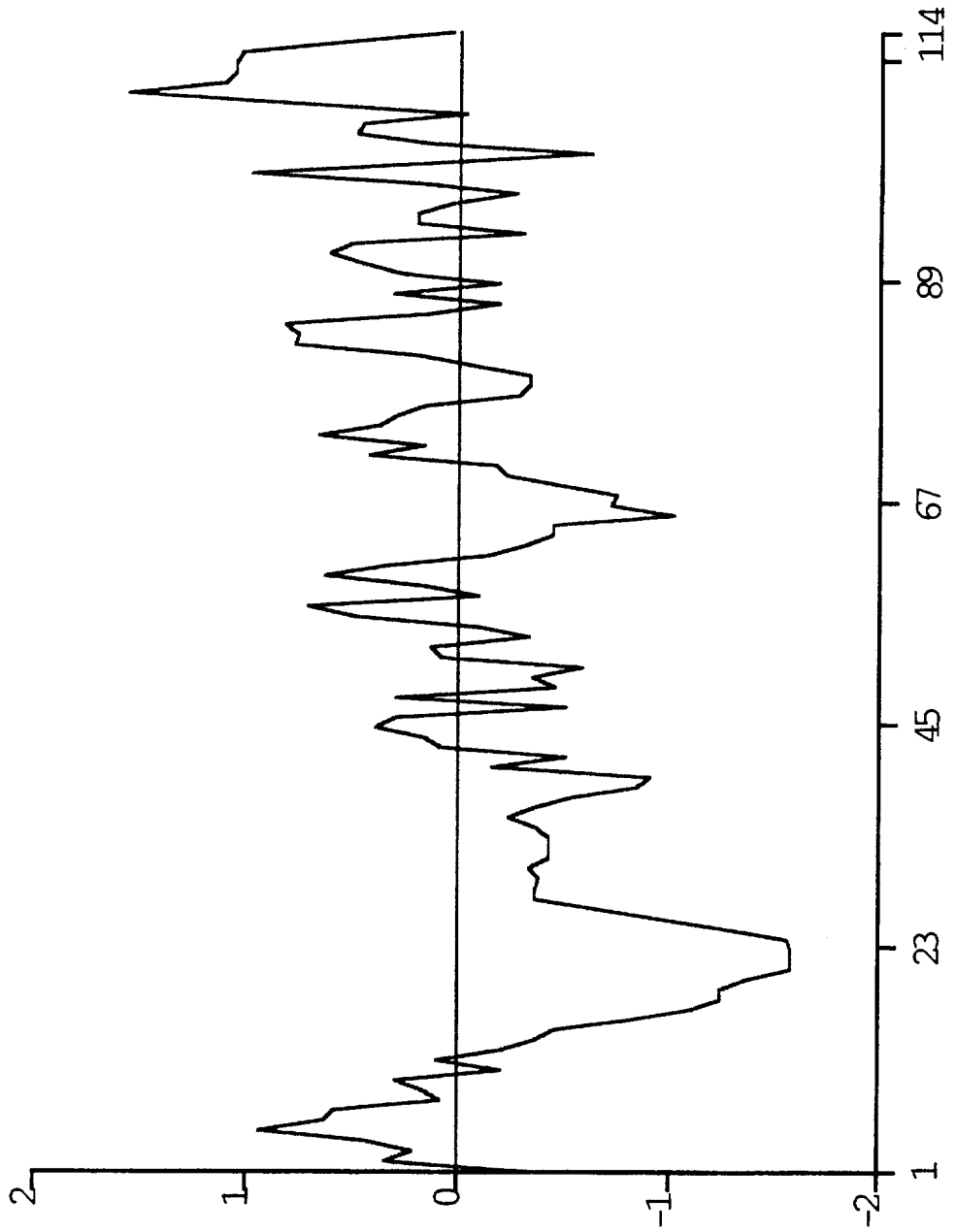


FIGURE 3

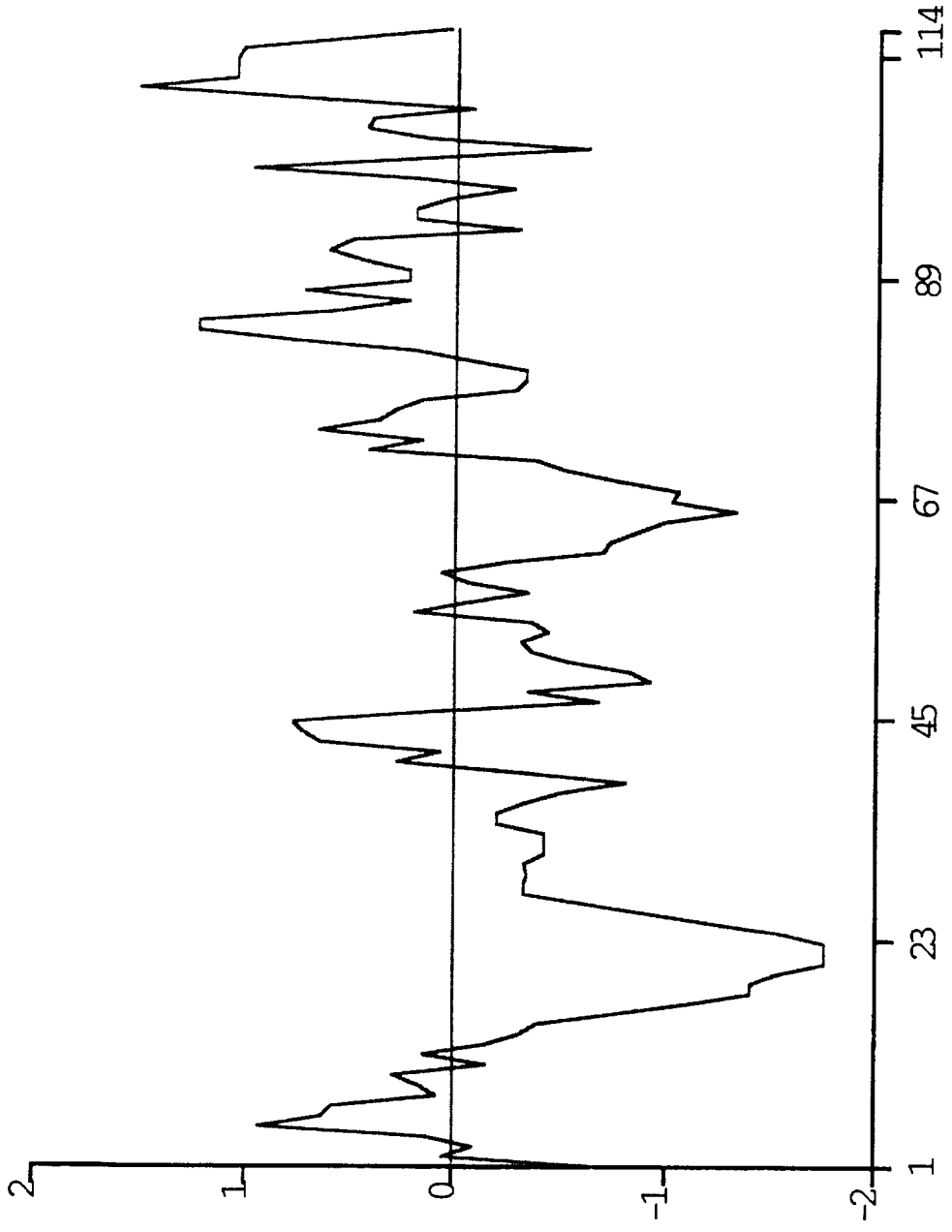


FIGURE 4

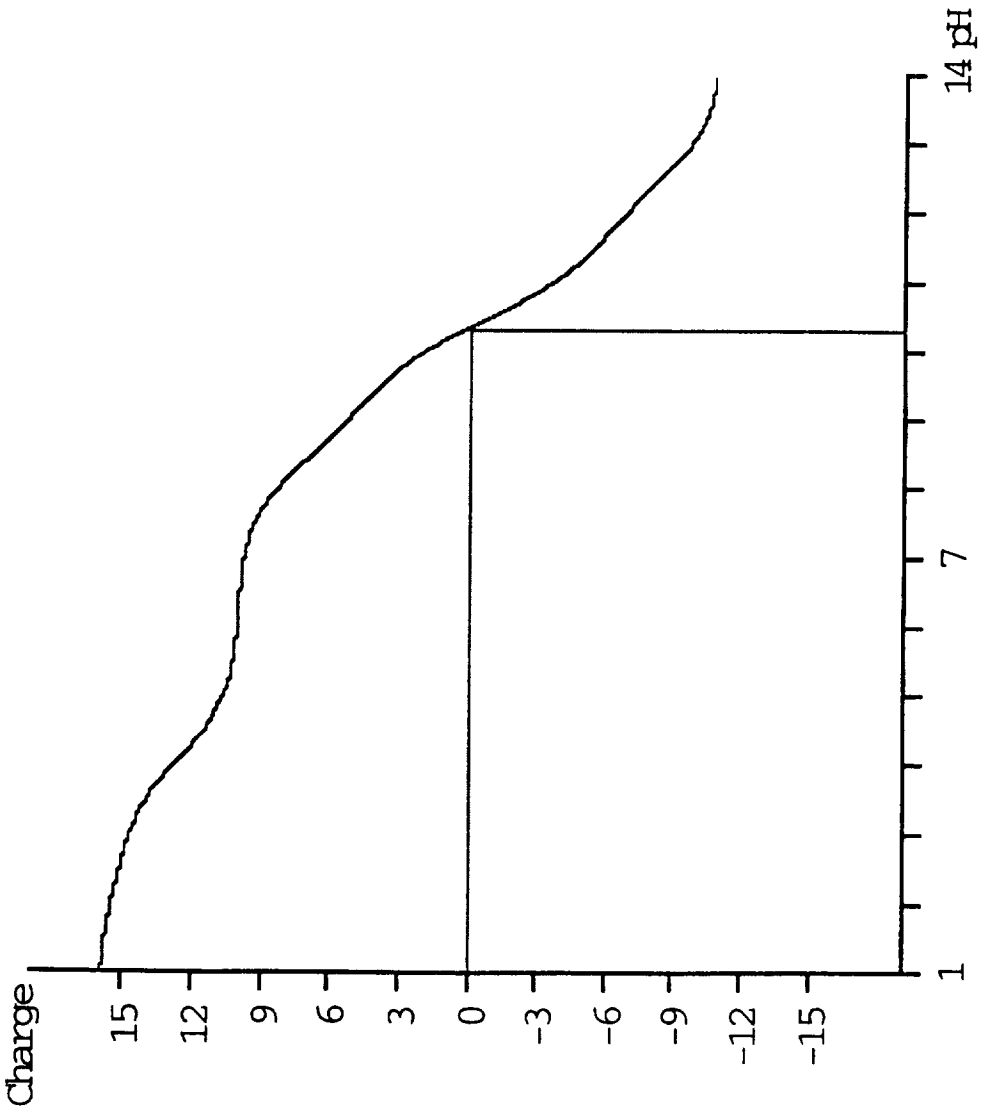


FIGURE 5

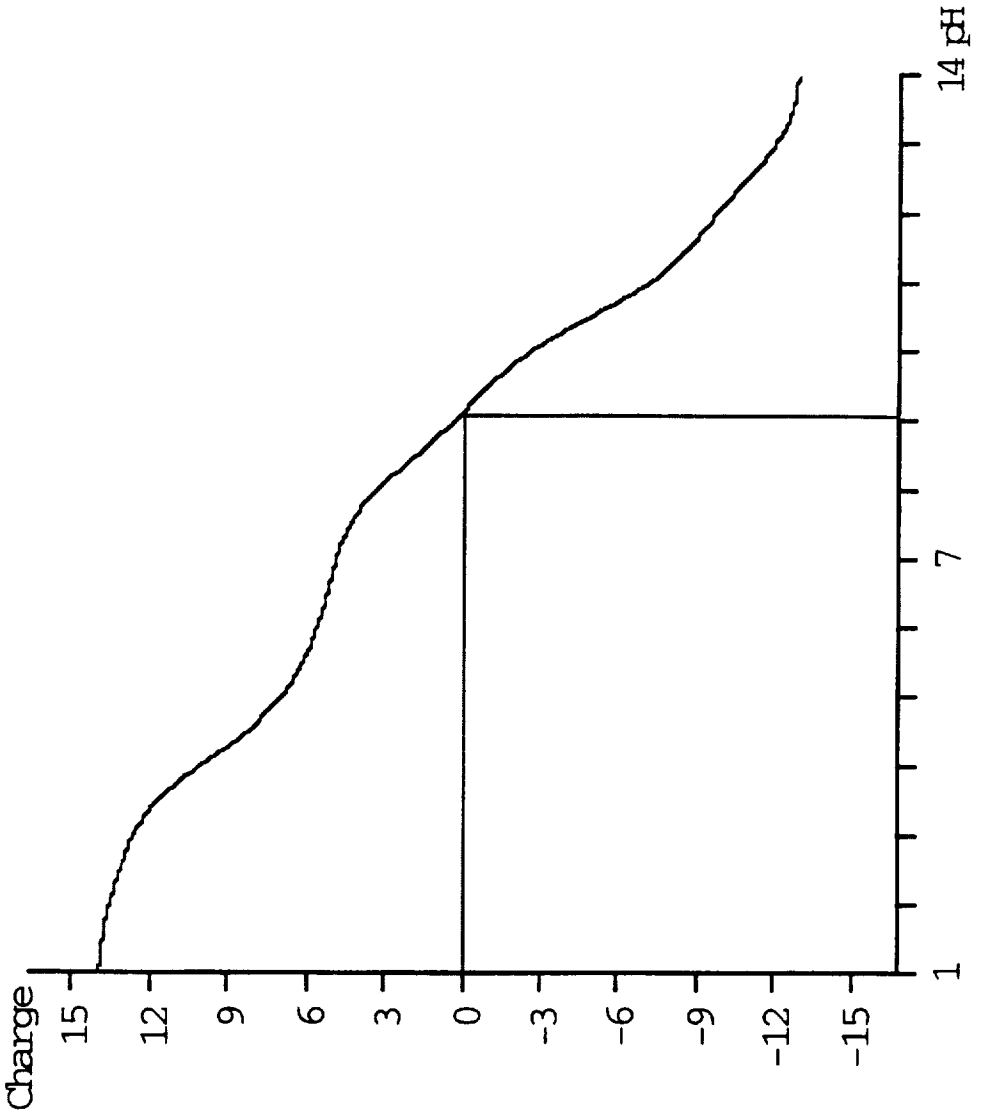


FIGURE 6

```
1   GAACCAAAGTGCTCTGTATCCTCCAGTCTCCGCGCCTCCA 949299.nt
1   --------AA T-C TTC GC-TCCTCCAATCTCCGCTCCTCCA g973875.nt

41  CCCAGCTCAGGAACCCGCGAAC----------CCTCTCTTGA 949299.nt
32  CCCAGTTCAGGAACCCGCGACCGCTCGCAGCCTCTCTTGA g973875.nt

73  CCACTATGAGCCTCCCGTCCAGCCGCGCGGCCCGTGTCCC 949299.nt
72  CCACTATGAGCCTCCTGTCCAGCCGCGCG-CCCGTGTCCC g973875.nt

113 GGGTCCTTCGGGCTCCTTGTGCGCGCTGCTCGCGCTGCTG 949299.nt
111 CGGTCCTTCGAGCTCCTTGTGCGCGCTGTTGGTGCTGCTG g973875.nt

153 CTCCTGCTGACGCCGCCGGGGCCCCTCGCCAGCGCTGGTC 949299.nt
151 CTGCTGCTGACGT-ACCAGGGCCCATCGTCAGCGCTGGTC g973875.nt

193 CTGTCTCTGCTGTGCTGACAGAGCTGCGTTGCACTTGTTT 949299.nt
190 CTGCCGCTGCTGTTGAGAGAGCTGCGTTGCGTTTGTTT g973875.nt

233 ACGCGTTACGCTGAGAGTAAACCCCAAAACGATTGGTAAA 949299.nt
230 ACAGACCACGNAGGGAGTTCATCCCAAAATGATCAGTAAT g973875.nt

273 CTGCAGGTGTTCCCCGCAGGCCCGCAGT-GCTCCAAGGTG 949299.nt
270 CTGCAAGTGTTCGNCATAGGGCCACAGTTGTTNCAAGGTT g973875.nt
```

FIGURE 7A

```
312  G A A G T G G T A G C C T C C C T G A A G A A C G G G A A G C A A G T T T - G T  949299.nt
310  G A A G T G G T A G C C T T N C T G A A G A A C G G G A A G G A A A T T T T G T  g973875.nt

351  C - - T G G A C C C G G A A G - - C C C C T T T T C T A A A G A A A G T C A T -  949299.nt
350  T C T T G A T T N C A G A A G A C C N T T T T T T N T A A A G A N A G T C A T T  g973875.nt

386  C C A G A A A A T T T T - - G G A C A G T G G A A A C A A G A A A A A C T G A G  949299.nt
390  N C A G A A A A T T T T T G G G A C G G T N G G A - - A A C A A G G N A T A A T  g973875.nt

424  T A A C A A A A A A G A C C A T G C A T C A T A A A A T T G C C C A G T C T T C  949299.nt
428  T G A T T A A G G A G A A - A T G A G G C A C G N A - - T G - - - - - - - - - -  g973875.nt

464  A G C G G A G C A G T T T T C T G G A G A T C C C T G G A C C C A G T A A A A A  949299.nt
455  - - - G G A A A A G T T T T N C C - A G T T T T T C - - - - - - - - - - - - - -  g973875.nt

504  A A A A A A A A A A                                                          949299.nt
477  - - - - - - - - - A                                                          g973875.nt
```


FIGURE 7B

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 97/16034

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/19 C07K14/52 A61K38/19 C12N15/11 C12Q1/68
C07K16/24

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K C12Q

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 12537 A (UNIV LEUVEN REGA INST ;DAMME JO VAN (BE); PROOST PAUL (BE)) 9 June 1994 see page 29 ---	1-12
X	PROOST P. ET AL: "Human and bovine granulocyte chemotactic protein-2: Complete amino acid sequence and fuctional characterization as chemokines" BIOCHEMISTRY., vol. 32, 1993, EASTON, PA US, pages 10170-10177, XP002049871 see figure 2 --- -/--	1-12

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
11 December 1997	20. 01. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Espen, J

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter ›nal Application No
PCT/US 97/16034

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHANG M.-S. ET AL: "Cloning and characterization of the human neutrophil-activating peptide (ENA-78) gene" JOURNAL OF BIOLOGICAL CHEMISTRY., vol. 269, 1994, MD US, pages 25277-25282, XP002049872 cited in the application see figures 1,2 ---	1-12
Y	POWER C.A. ET AL: "Cloning of a full-length cDNA encoding the neutrophil-activating peptide ENA-78 from human platelets" GENE., vol. 151, 1994, AMSTERDAM NL, pages 333-334, XP002049873 see figure 1 -----	1-12

1

INTERNATIONAL SEARCH REPORT

Ir. .national application No.
PCT/US 97/16034

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

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

see FURTHER INFORMATION sheet PCT/ISA/210

2. [X] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

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

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

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

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 97/16034

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

Claims Nos.: 1 (not completely)

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

The sequence given within SEQ ID No 1 of the present international application does not correspond to the sequences given in Fig. 1A/B. The search was carried out with those sequences given in said Fig. 1A/B.

Remark : Although claim 11 are directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the compound/composition.

INTERNATIONAL SEARCH REPORT

...formation on patent family members

Interr nal Application No
PCT/US 97/16034

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9412537 A	09-06-94	AU 5628394 A	22-06-94
		EP 0804486 A	05-11-97