Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2319211 C 2010/10/26

(11)(21) 2 319 211

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1999/01/27
(87) Date publication PCT/PCT Publication Date: 1999/08/05
(45) Date de délivrance/Issue Date: 2010/10/26
(85) Entrée phase nationale/National Entry: 2000/07/27
(86) N° demande PCT/PCT Application No.: EP 1999/000478
(87) N° publication PCT/PCT Publication No.: 1999/038967
(30) Priorité/Priority: 1998/01/30 (IT MI98A000179)

(51) Cl.Int./Int.Cl. *C12N 15/12* (2006.01), *A61K 38/18* (2006.01), *C07K 14/475* (2006.01), *C07K 19/00* (2006.01), *A61K 38/00* (2006.01)

(72) Inventeurs/Inventors:
MEDICO, ENZO, IT;
MICHIELI, PAOLO, IT;
COLLESI, CHIARA, IT;
CASELLI, GIANFRANCO, IT;
COMOGLIO, PAOLO, IT

(73) Propriétaire/Owner:
DOMPE PHA.R.MA S.P.A., IT

(74) Agent: KIRBY EADES GALE BAKER

(54) Titre : PROTEINES RECOMBINEES DERIVEES DE HGF ET MSP
(54) Title: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

(57) Abrégé/Abstract:
Recombinant proteins deriving from recombination of structural domains deriving from the α subunits of HGF and/or MSP growth factors. The recombinant proteins of the present invention have biological activity, and protect cells from death (apoptosis) induced by chemotherapeutic drugs. These molecules can conveniently be used to prevent or to treat the toxic side effects of chemotherapeutic agents used in cancer therapy.

Abstract

Recombinant proteins deriving from recombination of structural domains deriving from the α subunits of HGF and/or MSP growth factors. The recombinant proteins of the present invention have biological activity, and protect cells from death (apoptosis) induced by chemotherapeutic drugs. These molecules can conveniently be used to prevent or to treat the toxic side effects of chemotherapeutic agents used in cancer therapy.

RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

Field of the invention

The present invention relates to recombinant proteins obtained from the combination of structural domains derived from the α subunits of hepatocyte growth factor (HGF) and macrophage stimulating protein (MSP).

In particular, the engineered factors of the invention are obtained by combination of the hairpin loop and kringle domains of the α chains of HGF and/or MSP, so as to obtain a structure having two superdomains with an intervening linker sequence. Moreover, the invention relates to DNA sequences encoding the above mentioned recombinant proteins, to the expression vectors comprising said DNA sequences and to host cells containing said expression vectors. The recombinant proteins of the present invention are biologically active, and their activity can be measured by determination of their ability to induce activation of the Met tyrosine kinase receptor, their "scattering" effect on epithelial cells, and their protective effect against cell death induced by chemotherapic drugs (vide infra). Therefore, these molecules can conveniently be used to prevent or treat the toxic side effects of the chemotherapeutical treatment of tumours, and to reduce iatrogenic cell damage induced by other types of drugs.

Technological background

Hepatocyte Growth Factor (HGF) and Macrophage Stimulating Protein (MSP) are highly related proteins both structurally and

functionally (Fig. 1 and 2). Both these factors are secreted as an inactive precursor, which is processed by specific proteases which recognise a cleavage site inside the molecule, dividing the protein in two subunits. These subunits, named α chain and β chain, are linked by a disulphide bond. Thus, the mature factor is an α-β dimeric protein. Only the mature (dimeric) form of the factor is able to activate its receptor at the surface of the target cells (the Met tyrosine kinase in the case of HGF and the Ron tyrosine kinase in the case of MSP) and therefore to mediate biological responses (Naldini, L. et al., 1992, EMBO J. 11: 4825-4833; Wang, M. et al., 1994, J. Biol. Chem. 269; 3436-3440; Bottaro, D. et al., 1991, Science 25: 802-804; Naldini, L. et al., 1991, EMBO J. 10: 2867-2878; Wang, M. et al., 1994, Science 266: 117-119; Gaudino, G. et al., 1994, EMBO J. 13: 3524-3532).

The α chain of both factors contains a hairpin loop (HL) structure and four domains with a tangle-like structure named kringles (K1-K4; Nakamura T et al., 1989, Nature 342:440-443; Han, S. et al.., 1991, Biochemistry 30: 9768-9780). The precursor also contains a signal sequence (LS) of 31 amino acids (in the case of HGF) or of 18 amino acids (in the case of MSP), removed in rough endoplasmic reticulum, which directs the neoformed peptide to the secretive pathway. The β chain contains a box with a sequence homologous to that typical of serine proteases, but it has no catalytic activity (Nakamura T et al., 1989, Nature 342:440-443; Han, S. et al.., 1991, Biochemistry 30: 9768-9780). Both α and β chains contribute to the binding of the growth factor to the respective receptor (Met for HGF and

Ron for MSP).

HGF and MSP polypeptides are able to induce a variety of biological effects besides cell proliferation. The main biological activities of these molecules are: stimulation of cell division (mitogenesis); stimulation of motility (scattering); induction of polarisation and cell differentiation; induction of tubule formation (branched morphogenesis); increase of cell survival (protection from apoptosis). The tissues that respond to HGF and MSP stimulation are those where cells express the respective Met (HGF) and Ron (MSP) receptors. The most important target tissues of these factors are epithelial cells of different organs, such as liver, kidney, lung, breast, pancreas and stomach, and some cells of the hematopoietic and nervous systems. A detailed review of the biological effects of HGF and MSP in the various tissues can be found in Tamagnone, L. & Comoglio, P., 1997, Cytokine & Growth Factor Re-views, 8: 129-142, Elsevier Science Ltd.; Zarnegar, R. & Michalopoulos, G., 1995, J. Cell Biol. 129: 1177-1180; Medico, E. et al., 1996, Mol. Biol. Cell, 7: 495-504; Banu, N. et al., 1996, J. Immunol. 156: S2933-2940.

In the case of HGF, the hairpin loop and the first two kringles are known to contain the sites of direct interaction with the Met receptor (Lokker NA et al., 1992, EMBO J., 11:2503-2510; Lokker, N. et al., 1994, Protein Engineering 7: 895-903). Two naturally-occurring truncated forms of HGF produced by some cells by alternative splicing have been described. The first one comprises the first kringle (NK1-HGF Cioce, V. et al., 1996, J. Biol. Chem.

271: 13110-13115) whereas the second one spans to the second kringle (NK2-HGF Miyazawa, K. et al., 1991, Eur. J. Biochem. 197: 15-22). NK2-HGF induces cell scattering, but it is not mitogenic as the complete growth factor is (Hartmann, G. et al., 1992, Proc. Natl. Acad. Sci. USA 89: 11574-11578). However, NK2-HGF re-gains mitogenic activity in the presence of heparin, a glucosaminoglycan that binds HGF through a domain contained in the first kringle and which is likely to induce dimerization of NK2-HGF (Schwall, R. et al., 1996, J. Cell Biol. 133: 709-718). Moreover NK2-HGF, being a partial agonist of Met, behaves as a competitive inhibitor of HGF as far as the mitogenic activity is concerned (Chan, A. et al., 1991, Science 254: 1382-1385). NK1-HGF has also been described to exert partial stimulation of Met and competitive inhibition of HGF mitogenic activity (Cioce, V. et al., 1996, J. Biol. Chem. 271: 13110-13115). Anyway, a truncated factor is endowed with an activity markedly lower than the recombinant factors described in the invention, as shown in example 3.

In the case of MSP, the interaction sites with the Ron receptor are less understood: some preliminary studies suggest a situation opposite of that of HGF, i.e. the β chain directly binds the receptor whereas the α chain would act stabilizing the complex (Wang MH et al., 1997, J. Biol. Chem. 272:16999-17004).

The therapeutical use of molecules such as HGF and MSP is potentially valuable in a wide range of pathologies (Abdulla, S., 1997, Mol. Med. Today 3: 233). Nevertheless, a number of technical

as well as biological complications make the application of these molecules in clinics difficult. First of all, the pleiotropic character of these factors can cause poorly selective biological responses, which involve undesired side effects. For example, the use of HGF to prevent some side effects of the chemotherapeutic drug cisplatin has been proposed (Kawaida K et al., 1994, Proc. Natl. Acad. Sci. 91:4357-4361). Cancer patients treated with this drug can suffer kidney acute damage due to the cytotoxic action of cisplatin on proximal tubule epithelial cells. HGF is able to protect these cells against programmed death (apoptosis) induced by cisplatin, but at the same time it can induce an undesired proliferation of neoplastic cells. Other problems related to the pharmaceutical use of HGF and MSP are the necessity of their proteolytic activation and their stability, which causes technical problems. The NK1 and NK2 truncated forms of HGF do not require proteolytic activation, but they have a reduced biological activity.

Summary of the invention

Certain exemplary embodiments of the invention can provide for a recombinant protein comprising two superdomains, separated by a spacer sequence (linker), each superdomain consisting of a combination of HL and K1-K4 domains from the alpha chain of either HGF or MSP, according to the formula (I):

$$[A] - B - [C] - (D)y \qquad (I)$$

in which [A] corresponds to the sequence $(LS)_m$-HL-K1-$(K2)_n$-$(K3)_o$-$(K4)_p$ wherein the numbering of the following amino acids refers to the HGF and MSP sequences as provided in Figures 1a and 1b, respectively: LS is an

amino acid sequence corresponding to residues 1-31 of HGF or 1-18 of MSP; HL is an amino acid sequence starting between residues 32-70 of HGF α chain and ending between residues 96-127 of the identical chain; or it is an amino acid sequence starting between residues 19-56 of MSP α chain and ending between residues 78-109 of the identical chain; K1 is an amino acid sequence starting between residues 97-128 of HGF α chain and ending between residues 201-205 of the identical chain; or it is an amino acid sequence starting between residues 79-110 of MSP α chain and ending between residues 186-190 of the identical chain; K2 is an amino acid sequence starting between residues 202-206 of HGF α chain and ending between residues 283-299 of the identical chain; or it is an amino acid sequence starting between residues 187-191 of MSP α chain and ending between residues 268-282 of the identical chain; K3 is an amino acid sequence starting between residues 284-300 of HGF α chain and ending between residues 378-385 of the identical chain; or it is an amino acid sequence starting between residues 269-283 of MSP α chain and ending between residues 361-369 of the identical chain; K4 is an amino acid sequence starting between residues 379-386 of HGF α chain and ending between residues 464-487 of the identical chain; or it is an amino acid sequence starting between residues 362-370 of MSP α chain and ending between residues 448-481 of the identical chain; m, n, o, p are 0 or 1; the sum n + o + p is an integer from 1 to 3 or 0, with the proviso that $n \geq o \geq p$; B is the sequence $[(X)_q\ Y]_r$, wherein X = Gly and Y = Ser, or Cys, or Met, or Ala; q is an integer from 2 to 8; r is an integer from 1 to 9; [C] corresponds to the sequence HL-K1-$(K2)_s$-$(K3)_t$-$(K4)_u$ wherein HL, K1-K4 are as defined

above, s, t, u are 0 or 1; the sum s + t + u is an integer from 1 to 3 or 0, with the proviso that $s \geq t \geq u$; D is the sequence W-Z, wherein W is a proteolytic site, Z is a tag sequence for the purification and detection of the protein; y is 0 or 1.

The present invention provides recombinant molecules composed of a combination of structural domains derived from the α chains of HGF and/or MSP, which overcome the problems of the prior art molecules described above. The molecules of this invention are composed of two superdomains connected by a linker. Each superdomain is composed of a combination of the HL and K1-K4 domains of the α chain of HGF and/or MSP. These engineered factors induce selective biological responses, do not require

proteolytic activation, are stable and are more active than the truncated forms of HGF described previously.

Detailed disclosure of the invention

The present invention relates to recombinant proteins (which will be hereinafter referred to indifferently as proteins, molecules, engineered or recombinant factors) characterised by a structure that comprises two superdomains, each consisting of a combination of HL and K1-K4 domains derived from the α chain of HGF and/or MSP, linked by a spacer sequence or a linker. In particular, the invention relates to proteins of general formula (I)

$$[A] - B - [C] - (D)_y \qquad (I)$$

in which

[A] corresponds to the sequence $(LS)_m$-HL-K1-$(K2)_n$-$(K3)_o$-$(K4)_p$

wherein (the numbering of the following amino acids refers to the HGF and MSP sequences as reported in Fig. 1 and 2, respectively):

LS is an amino acid sequence corresponding to residues 1-31 of HGF or 1-18 of MSP;

HL is an amino acid sequence derived from the α chain of HGF starting between residues 32-70 and ending between residues 96-127; or it is an amino acid sequence derived from the α chain of MSP starting between residues 19-56 and ending between residues 78-109;

K1 is an amino acid sequence derived from the α chain of HGF starting between residues 97-128 and ending between residues 201-205; or it is an amino acid sequence derived from the α chain of MSP starting between residues 79-110 and ending between residues 186-190;

K2 is an amino acid sequence derived from the α chain of HGF starting between residues 202-206 and ending between residues 283-299; or it is an amino acid sequence derived from the α chain of MSP starting between residues 187-191 and ending between residues 268-282;

K3 is an amino acid sequence derived from the α chain of HGF starting between residues 284-300 and ending between residues 378-385; or it is an amino acid sequence derived from the α chain of MSP starting between residues 269-283 and ending between residues 361-369;

K4 is an amino acid sequence derived from the α chain of HGF starting between residues 379-386 and ending between residues 464-487; or it is an amino acid sequence derived from the α chain of MSP starting between residues 362-370 and ending between residues 448-481;

m, n, o, p can be 0 or 1;

the sum n + o + p is an integer from 1 to 3 or 0, with the proviso that $n \geq o \geq p$;

B is the sequence $[(X)_q\ Y]_r$, wherein X = Gly and Y = Ser, or Cys, or Met, or Ala;

q is an integer from 2 to 8;

r is an integer from 1 to 9;

[C] corresponds to the sequence HL-K1-$(K2)_s$-$(K3)_t$-$(K4)_u$

wherein HL, K1-K4 are as defined above,

s, t, u are 0 or 1; the sum s + t + u is an integer from 1 to 3 or 0, with the proviso that $s \geq t \geq u$;

D is the sequence W-Z, wherein W is a conventional proteolytic site, Z is any tag sequence useful for the purification and detection of the protein; y

is 0 or 1.

Non-limiting examples of W are consensus sequences for enterokinase protease, thrombin, factor Xa and IgA protease.

Preferred proteins of general formula (I), are those in which: the HL domain is a sequence of HGF α chain ranging from amino acids 32 to 127, or a sequence of MPS α chain ranging from amino acids 19 to 98; the K1 domain is a sequence of HGF α chain ranging from amino acids 128 to 203, or a sequence of MPS α chain ranging from amino acids 99 to 188; the K2 domain is a sequence of HGF α chain ranging from amino acids 204 to 294, or a sequence of MPS α chain ranging from amino acids 189 to 274; the K3 domain is a sequence of HGF α chain ranging from amino acids 286 to 383, or a sequence of MPS α chain ranging from amino acids 275 and 367; the K4 domain is a sequence of HGF α chain ranging from amino acids 384 to 487, or a sequence of MPS α chain ranging from amino acids 368 and 477.

Among the possible combinations of the domains of general formula (I), the following (II) and (III) are preferred, concerning two recombinant factors named Metron Factor-1 and Magic Factor-1, respectively:

LS_{MSP}-HL_{MSP}-$K1_{MSP}$-$K2_{MSP}$-L-HL_{HGF}-$K1_{HGF}$-$K2_{HGF}$-D (Metron Factor-1) (II)

and

LS_{HGF}-HL_{HGF}-$K1_{HGF}$-$K2_{HGF}$-L-HL_{HGF}-$K1_{HGF}$-$K2_{HGF}$-D (Magic Factor-1) (III)

For both molecules, L is a linker sequence $(Gly_4Ser)_3$, D is a tag

sequence Asp_4-Lys-His_6.

For Metron Factor-1, LS_{MSP} is the sequence 1-18 of MSP, HL_{MSP} is the sequence 19-56 of MSP, $K1_{MSP}$ is the sequence 99-188 of MSP, $K2_{MSP}$ is the sequence 189-274 of MSP, HL_{HGF} is the sequence 32-127 of HGF, $K1_{HGF}$ is the sequence 128-203 of HGF, $K2_{HGF}$ is the sequence 204-294 of HGF.

For Magic Factor-1, HL_{HGF}, $K1_{HGF}$, $K2_{HGF}$ are as defined above, LS_{HGF} is the sequence 1-31 of HGF.

The hybrid molecules of the invention are prepared by genetic engineering techniques according to a strategy involving the following steps:

a) construction of DNA encoding the desired protein;

b) insertion of DNA in an expression vector;

c) transformation of a host cell with recombinant DNA (rDNA);

d) culture of the transformed host cell so as to express the recombinant protein;

e) extraction and purification of the produced recombinant protein.

The DNA sequences corresponding to HGF or MSP structural domains can be obtained by synthesis or starting from DNA encoding for the two natural factors. For example, screening of cDNA libraries can be carried out using suitable probes, so as to isolate HGF or MSP cDNA. Alternatively, HGF or MSP cDNA can be obtained by reverse transcription from purified mRNA from suitable cells.

cDNAs coding for the fragments of HGF and MSP β chains can be amplificated by PCR (Mullis, K.B. and Faloona, F.A., 1987, Methods in

Enzymol. 155, 335-350), and the amplification products can be recombined making use of suitable restriction sites, naturally occurring in the factor sequences or artificially introduced in the oligonucleotide sequence used for the amplification.

In greater detail, one of the above mentioned strategies can be the following:

the portions of DNA encoding the LS, HL, K1, K2, K3 and K4 domains are amplificated by PCR from HGF or MSP cDNA and then recombined to obtain the hybrid sequences corresponding to [A] and [C]. Oligonucleotides recognising sequences located at the two ends of the domains to be amplificated are used as primers. Primers are designed so as to contain a sequence allowing recombination between the DNA of a domain and the adjacent one. Said recombination can be carried out by endonuclease cleavage and subsequent ligase reaction, or making use of the recombinant PCR method (Innis, NA et al.., 1990, in PCR Protocols, Academic Press, 177-183).

The sequence encoding the domain B (linker) can be obtained by synthesis of a double chain oligonucleotide, which can be inserted between [A] and [C] using suitable restriction sites.

The resulting three fragments encoding for [A], [B] and [C] are then inserted in the correct sequence in a suitable vector. In this step it can be decided whether to add or not the domain D (tag), obtained by synthesis analogously to domain B, downstream fragment [C].

The recombinant expression vector can contain, in addition to the recombinant construct, a promoter, a ribosome binding site, an initiation

codon, a stop codon, optionally a consensus site for expression enhancers.

The vector can also comprise a selection marker for isolating the host cells containing the DNA construct. Yeast or bacteria plasmids, such as plasmids suitable for Escherichia Coli, can be used as vectors, as well as bacteriophages, viruses, retroviruses, or DNA.

The vectors are cloned preferably in bacterial cells, for example in Escherichia Coli, as described in Sambrook J., 1989, Molecular Cloning, Cold Spring Harbor Laboratory Press, New York, and the colonies can be selected, for example, by hybridisation with radiolabelled oligonucleotide probes; subsequently, the rDNA sequence extracted from the positive colonies is determined by known methods.

The vector with the recombinant construct can be introduced in the host cell according to the competent cell method, the protoplast method, the calcium phosphate method, the DEAE-dextran method, the electric impulses method, the in vitro packaging method, the viral vector method, the micro-injection method, or other suitable techniques.

Host cells can be prokaryotic or eukaryotic, such as bacteria, yeasts or mammal cells, and they will be such as to effectively produce the recombinant protein.

After transformation, cells are grown in a suitable medium, which can be for example MEM, DMEM or RPMI 1640 in the case of mammal host cells.

The recombinant protein is secreted in the culture medium from which it can be recovered and purified with different methods, such as mass exclusion, absorption, affinity chromatography, salting-out, precipitation, dialysis, ultrafiltration.

A simple, rapid system for the production of the molecules of the invention is, for example, transient expression in mammal cells.

Accordingly, the plasmid containing the recombinant DNA fragment, for example PMT2 (Sambrook, J. et al., 1989, Molecular Cloning, Cold Spring Harbor Laboratory Press), is transfected in suitable recipient cells, such as Cos7 (Sambrook, J. et al., supra) by the calcium phosphate technique or other equivalent techniques. Some days after transfection, the conditioned medium of the transfected cells is collected, cleared by centrifugation and analysed for its content in factor. For this analysis, antibodies directed against HGF or MSP, or against any tag sequence, can be used: the supernatant is immunoprecipitated and then analysed by western blot with the same antibody. The supernatant containing the recombinant factor can also be used directly for biochemical and biological tests. The protein can be purified, for example, using a poly-histidine tag sequence, by absorption on a nickel resin column and subsequent elution with imidazole.

The biochemical properties of the recombinant factors of the invention were tested in connection with their ability to activate Met and Ron receptors.

Sub-micromolar concentrations of the factors have proved to induce phosphorylation in Met tyrosine in human epithelial cells A549, whereas

they do not induce phosphorylation above basal values in cells expressing Ron. On the whole, the tests proved that the first two kringles of HGF maintain their ability to interact and to activate Met tyrosine kinase receptor, whereas the corresponding first two kringles of MSP are not sufficient for modulating the catalytic activity of the Ron receptor. However, the interaction with Ron, although at low affinity, can contribute to the recruitment of the factor at the cell surface, playing a similar role to low affinity receptors (of mature glycoprotein) which recruit the HGF intact molecule through the heparin-binding domain.

The molecules of the invention have a marked biological activity, measured by the scattering tests, and a protecting activity against cell apoptosis induced by cisplatin or etoposide.

In particular, the supernatant containing the recombinant factor has been found to promote scattering of epithelial cells of various nature even at nanomolar concentrations. In these tests, kidney epithelial cells (MDCK) or hepatocyte precursors (MLP29) were used.

In an in vitro experimental system, in which DNA fragmentation typical of apoptotic cells is evaluated by the TUNEL method (Gavrieli, Y. et al., 1992, J. Cell. Biol. 117, 493-501), the recombinant factors protect against apoptosis induced by chemotherapeutic drugs at levels comparable with HGF and remarkably higher than MSP. The engineered molecules proved to be active on human primary epithelial cells from proximal tubule (PTECs), on an immortalised PTECs line (Loc) and on the already cited murine hepatocytes MLP29.

Among the applications of the recombinant molecules of the

invention, the following can be cited:

- prevention of myelotoxicity; in particular they can be used for the expansion of marrow precursors, to increase proliferation of the hematopoietic precursors or to stimulate their entry in circle;
- prevention of liver and kidney toxicity, and of mucositis following antineoplastic treatments; in particular the recombinant factors can be used to prevent toxicity (apoptosis) on differentiated cell elements of liver, kidney and mucosa of the gastroenteral tract, and to stimulate staminal elements of cutis and mucosas to allow the regeneration of germinative layers;
- prevention of chemotherapeutic neurotoxicity.

In general, the proteins of the invention provide the following advantages, compared with the parent molecules HGF and MSP:

- they are smaller molecules with a more compact structure;
- they are more stable and are produced in higher amounts;
- they require no endoproteolytic cleavage for activation, which transforms the HGF and MSP precursors into the respective active forms;
- they can be engineered in combinations of different functional domains, thereby modulating the biological effects, increasing the favourable ones and reducing those undesired (for example, protection from apoptosis versus cell proliferation).

The invention has to be considered also directed at amino acid and nucleotide sequences referred to formula (I), having modifications which can, for example, derive from degeneration of genetic code, without

therefore modifying the amino acid sequence, or from the deletion, substitution, insertion, inversion or addition of nucleotides and/or bases according to all the possible methods known in the art.

Furthermore, the invention relates to the expression vectors comprising a sequence encoding for a protein of general formula (I), which can be plasmids, bacteriophages, viruses, retroviruses, or others, and to host cells containing said expression vectors.

Finally, the invention relates to the use of the recombinant proteins as therapeutical agents, and to pharmaceutical compositions containing an effective amount of the recombinant proteins together with pharmacologically acceptable excipients.

Description of the Figures

(In the following legends, -His located after the name of the parent factors, truncated or recombinant, or of the plasmids, means that the respective sequences contain a poly-histidine tag).

Figure 1:

a) Nucleotide and amino acid sequence of human HGF (Gene Bank # M73239; Weidner, K.M., et al., 1991, Proc. Acad. Sci. USA, 88:7001-7005). In contrast to the cited reference, in the numbering used herein, nucleotide No. 1 is the first base of the initiation codon (the A of the first ATG). The first amino acid is the corresponding methionine. The cDNA untranslated regions at 5' and 3' are neither represented nor considered in the numbering.

b) Nucleotide and amino acid sequence of human MSP (Gene Bank # L11924; Yoshimura, T., et al., 1993, J. Biol. Chem., 268:15461-

15468). In contrast to the cited reference, in the numbering used herein nucleotide No. 1 is the first base of the initiation codon (the A of the first ATG). The first amino acid is the corresponding methionine. The cDNA untranslated regions at 5' and 3' are neither represented nor considered in the numbering.

Figure 2:

a) Molecular structure of Metron Factor-1. The leader sequence is removed from the cells used for the production before secretion and is therefore absent in the mature molecule. The poly-histidine tag can be removed by digestion with the protease enterokinase.

b) Nucleotide and amino acid sequence of Metron-Factor-1. The nucleotide sequence starts with the EcoRI site and terminates with the SalI site (first six bases and last six bases, respectively). The initiation codon (ATG) and the stop codon (TAG) are underlined.

Figure 3:

a) Molecular structure of Magic Factor-1. The leader sequence is removed from the cells used for the production before secretion and is therefore absent in the mature molecule. Poly-histidine tag can be removed by digestion with the protease enterokinase.

b) Nucleotide and amino acid sequence of Magic Factor-1. The nucleotide sequence starts with the SalI site (first six bases and last six bases, respectively). The initiation codon (ATG) and the stop codon (TAG) are underlined.

Figure 4:

Production of Metron-F-1 by transient transfection of mammal

cells. The conditioned supernatants from BOSC cells transfected with the control plasmid (CTRL) or with pRK7-Metron F-1-His were immunoprecipitated with an anti-MSP polyclonal antibody and detected by western blot with the same antibody.

Figure 5:

Quantitation of the recombinant proteins by western blot. (A) The proteins were absorbed on Sepharose-A-heparin beads and detected with an anti-poly-histidine monoclonal antibody. (B) The proteins were immunoprecipitated with an anti-MSP polyclonal antibody and detected with an anti-poly-histidine monoclonal antibody.

Figure 6:

Scattering test carried out on kidney epithelial cells (MDCK) using the recombinant proteins prepared by transient transfection. The protein content was quantified by western blot (see Fig. 5). (A) non-stimulated cells; (B) cells stimulated with control supernatant; (C) cells stimulated with HGF-His; (D) cells stimulated with NK2-HGF-His; (E) cells stimulated with Metron Factor-1; (F) cells stimulated with Magic Factor-1.

Figure 7:

Activation (phosphorylation) of Met receptor by the hybrid factor Metron Factor-1. Human epithelial cells (A549) were stimulated with supernatants conditioned from BOSC cells transfected with the control plasmid (CTRL) or with pRK7-Metron-F-1-His (METRON F-1) at the indicated dilutions. Cell lysates from the stimulated cells were immunoprecipitated with an anti-Met monoclonal antibody and detected by western blot with an anti-phosphotyrosine monoclonal antibody.

Figure 8:

Protective effect of Metron-F-1 against acute renal failure induced by $HgCl_2$ in vivo. Balb-c mice were injected i.v. with Metron-F-1 or vehicle at 0.5 h before and 6, 12, 24, 36 and 48 h after $HgCl_2$ i.v. administration. BUN and histological evaluation of renal necrosis were measured at 72 h.

Data expressed as mean + e.s. of 7 animals/group (BUN) or 3 animals/group (histology).

The following examples illustrate in greater detail the invention.

Example 1a: Preparation of the recombinant construct encoding Metron Factor-1

HGF cDNA was obtained by the RT-PCR technique (Reverse Transcriptase PCR; in: Innis, M. A., et al.., 1990, PCR Protocols, Academic Press, 21-27) from a human lung fibroblast cell line (MRC5; Naldini, L. et al., 1991, EMBO J. 10: 2867-2878). MSP cDNA was obtained with the same technique from human liver (Gaudino, G., et al., 1994, EMBO J. 13: 3524-3532).

The fragment corresponding to MSP LS-HL-K1-K2 was amplified by PCR using MSP cDNA as template and the following oligonucleotides as primers:

P1 (sense)

5' CGCGCGGAATTCCACCATGGGGTGGCTCCCACTCCT 3'

P2 (antisense)

5' CGCGCGCTCGAGGCGGGGCTGTGCCTCGGACCCGCA 3'

in which the underlined palindromic sequences are the restriction sites for

the enzymes EcoRI (oligonucleotide P1) and XhoI (oligonucleotide P2). The PCR product was digested with the restriction enzymes EcoRI and XhoI and then purified by electrophoresis on agarose gel.

The fragment corresponding to HL-K1-K2 of HGF was amplified by PCR using HGF cDNA as template and the following oligonucleotides as primers:

P3 (sense)

5' CGCGCGTCTAGAGGGACAAAGGAAAAGAAGAAATAC 3'

P4 (antisense)

5' CGCGCGAAGCTTTGTCAGCGCATGTTTTAATTGCAC 3'

in which the underlined palindromic sequences are the restriction sites for the enzymes XbaI (oligonucleotide P3) and HindIII (oligonucleotide P4). The PCR product was digested with the restriction enzymes XbaI and HindIII and then purified by electrophoresis on agarose gel.

For the linker sequence, the following partially complementary oligonucleotides were synthesised, and were subsequently annealed to obtain a double strand DNA fragment with sticky ends:

P5 (sense)

5'TCGAGGGCGGTGGCGGTTCTGGTGGCGGTGGCTCCGGCGGTGGCGGTTCT 3'

P6 (antisense)

5'CTAGAGAACCGCCACCGCCGGAGCCACCGCCACCAGAACCGCCACCGCCC 3'

in which the underlined bases are the sequences compatible with the restriction sites for the enzymes XhoI (oligonucleotide P5) and XbaI (oligonucleotide P6).

The resulting three DNA fragments were subcloned in the EcoRI-

HindIII sites of the expression vector pRK7 (Gaudino, G., et al., 1994, EMBO J. 13: 3524-3532), to obtain the recombinant plasmid pRK7-Metron-F-1, containing all the components of Metron Factor-1 except the tag sequence.

For the insertion of the tag sequence, the following partially complementary oligonucleotides were synthesised, and were subsequently annealed to obtain a double strand DNA fragment with sticky ends:

P7 (sense)

5'AGCTGACGACGACGACAAACACCACCACCACCACCACCACTAGGGTCGAC3'

P8 (antisense)

5'AGCTGTCGACCCTAGTGGTGGTGGTGGTGGTGGTGTTTGTCGTCGTCGTC3'

in which the underlined bases are compatible with the HindIII restriction site and the boxed palindromic sequences are the consensus sequences for the enzyme SalI. The resulting double strand DNA fragment was inserted in the restriction site HindIII of the recombinant plasmid obtained at the previous step (destroying the HindIII site and creating the SalI site), to obtain the plasmid pRK7-Metron-F-1-His.

Example 1b: Production of Metron Factor-1

The expression vector pRK7 contains a promoter of human cytomegalovirus immediate-early gene (CMV) and an episomal replication origin site of the DNA virus SV40. Therefore, this plasmid is particularly suitable for the expression of proteins in cells expressing the large T antigen of the virus SV40, such as kidney epithelial BOSC cells (Sambrook, J. et al.., 1989, Molecular Cloning, Cold Spring Harbor Laboratory Press). Metron Factor-1 can then be produced by transient

transfection of plasmid pRK7-Metron F-1-His in BOSC cells.

For transfection, 10^6 cells are seeded at day 0 in a 100 mm plate in 90% Dulbecco's Modified Eagle Medium (DMEM)-10% bovine calf serum (10 ml/plate). At day 1, cells are transfected with 10 µg/plate of pRK7-Metron-F-1-His by lipofection, using the protocol provided by the lipofectin producer (Gibco-BRL). At day 2, the DNA-containing medium is substituted by fresh medium with low content in serum (99.5% DMEM-0.5% bovine calf serum). At day 4 (48 hours after the end of the transfection), the medium is collected, cleared by centrifugation, and analysed for its content in Metron Factor-1.

This analysis can be carried out in different ways. For example, the recombinant protein present in the cleared supernatant can be immunoprecipitated with an anti-MSP antibody and then detected by western blot with the same antibody (Fig. 4). In the example shown in figure 4, 500 µl of supernatant (cleared by centrifugation, buffered in 25 mM HEPES and added with a protease inhibitors cocktail) were immunoprecipitated (2 hours at 4° C) with 20 µl of Sepharose™-A beads (Pharmacia) covalently conjugated with 2 µl of anti-MSP polyclonal antibody. The beads pellet was washed 3 times with 500 µl of washing buffer (20 mM HEPES pH 7.4; 150 mM NaCl; 0.1% Triton X-100™; 10% glycerol) and heated at 90° C for 2 minutes in 100 µl of Laemmli buffer. Eluted proteins were separated by SDS-PAGE on 8% BIS-acrylamide gel, transferred onto membrane (Hybond™-C; Amersham) and analysed by western blot. For this analysis, the same rabbit serum used for immunoprecipitation was employed as primary antibody with a 1:1000

dilution and protein A conjugated with peroxidase (Amersham) was used as secondary antibody. Protein A was detected by ECL (Amersham) following the protocol provided by the producer.

Alternatively, the recombinant protein can be partially purified by adsorption on Sepharose-A beads conjugated with heparin and subsequent analysis by western blot using antibodies directed to poly-histidine tag (Fig. 5).

In the example shown in figure 5, the Sepharose-A-heparin beads (20 µl; Pierce) were incubated (4 hours at 4° C) with 500 µl of supernatant (cleared by centrifugation, buffered in 25 mM HEPES and added with a protease inhibitors cocktail) in the presence of 500 mM NaCl, washed with suitable buffer (500 mM NaCl; 20 mM HEPES pH 7.4; 0.1% Triton X-100; 10% glycerol) and heated at 90° C for 2 minutes in 100 µl Laemmli buffer. Eluted proteins were separated by SDS-PAGE on 8% bis-acrylamide gel, transferred onto membrane (Hybond-C; Amersham) and analysed by western blot. For this analysis, a mouse monoclonal antibody to poly-histidine (Invitrogen) diluted 1:5000 was used as primary antibody and an anti-mouse IgG ovine antibody conjugated with peroxidase (Amersham) was used as secondary antibody. The secondary antibody was detected by ECL™ (Amersham) following the protocol provided by the producer.

The procedure of adsorption on heparin beads can also be used as protocol for the semi-purification of the recombinant protein. Furthermore, the molecule can additionally be purified making use of the poly-histidine affinity to heavy metals such as nickel. The protein containing poly-

histidine tag can be adsorbed on a nickel resin column (Invitrogen) and subsequently eluted with imidazole (the detailed protocol is provided by the manufacturer).

Example 1c: METRON-F-1 production in insect cells

The cDNA encoding for Metron-F1 was subcloned in a suitable expression vector (p-FASTBAC) to generate a recombinant plasmid containing the Metron-F1 gene (p-FASTBAC-Metron). A competent E. Coli strain (DH10 Bac) was transformed with p-FASTBAC-Metron to generate BACMID DNA. The DNA of positive colonies was isolated and checked by PCR to show the correct integration of the expression vector. Subsequently, the DNA from three clones was transfected into Sf9 insect cells with CellFECTIN reagent to produce virus particles. Virus titer was tested by a plaque assay. Single plaques were isolated and used for further propagation of the baculovirus. Viral stock was subsequently expanded in insect cells to scale up METRON-F-1 production. To verify protein expression, insect cells were infected with a multiplicity of infection (MOI) of 1 in a small-scale reactor. Samples of supernatants were analysed by SDS-PAGE followed by western blotting.

To produce amounts adequate for in vivo testing, insect cells were propagated in a 2.5-Liter stirred tank bioreactor. Cells were grown to a cell density of 1.106 ml^{-1} before they were infected with a MOI of 1. Cell suspension was harvested 3 days post infection. The supernatant containing the recombinant protein was separated by centrifugation. The presence of Metron F-1 in the supernatant was proved by SDS-PAGE followed by western blotting. Metron F-1 was pre-purified by a dual step

affinity chromatography on heparin sepharose (heparin-Hi Trap, Pharmacia) at 6° C. For in vivo testing or for further purification steps, the eluted fractions containing Metron F-1 were desalted by Sephadex™ G-25 chromatography (PD-10 or HiPrep 26/10, Pharmacia). Metron F-1 was further purified by chromatography on HisTrap columns (Pharmacia) and eluted by an imidazole gradient (0-0.5 M) using either a low-pressure system (Econo System, BIO-RAD) or an FPLC system (Pharmacia). Metron F-1 was eluted at an imidazole concentration of about 0.15 M. For in vivo testing, the eluted fractions containing Metron F-1 were freed of imidazole by Sephadex G-25 chromatography as already described, using the buffer to be used for animal treatment.

Example 2a: Preparation of the recombinant construct encoding for Magic Factor-1

HGF cDNA and the plasmid pRK7-Metron-F-1-His described above were used as starting DNA. The fragment corresponding to LS-HL-K1-K2 of HGF was amplificated by PCR using HGF cDNA as template and the following oligonucleotides as primers:

P9 (sense)

5' CGCGCGGGATCCGCCAGCCCGTCCAGCAGCACCATG 3'

P10 (antisense)

5' CGCGCGAAGCTTTGTCAGCGCATGTTTTAATTGCAC 3'

in which the underlined palindromic sequences are the restriction sites for the enzymes BamHI (oligonucleotide P9) and HindIII (oligonucleotide P10). The PCR product was digested with the restriction enzymes BamHI and HindIII and then purified by electrophoresis on agarose gel.

For the linker, the following partially complementary oligonucleotides were synthesized, and subsequently annealed to obtain a double strand DNA fragment with sticky ends:

P11 (sense)

5'AGCTTCGGGCGGTGGCGGTTCTGGTGGCGGTGGCTCCGGCGGTGGCGGTTCT3'

P12 (antisense)

5'CTAGAGAACCGCCACCGCCGGAGCCACCGCCACCAGAACCGCCACCGCCCGA3'

in which the underlined bases are the sequences compatible with the restriction sites for the enzymes HindIII (oligonucleotide P11) and XbaI (oligonucleotide P12). The fragment resulting by PCR and the double strand linker sequence were inserted in the plasmid pRK7-Metron-F-1-His in place of the fragment EcoRI-XbaI by means of an EcoRI-BamHI adapter, to obtain the plasmid pRK7-Magic-F-1-His.

Example 2b: Production of Magic Factor-1

Magic Factor-1 is produced on a small scale by transient transfection of BOSC cells analogously to what described for Metron Factor-1. Semi-purification is performed by adsorption on Sepharose-A beads conjugated with heparin followed by Western blot analysis using anti-poly-histidine antibodies (Fig. 5).

Example 3: Biological activity (scattering) on epithelial cells.

The biological activity of recombinant HGF, NK2-HGF, Metron Factor-1 and Magic Factor-1 was tested by a "scatter" assay on MDCK epithelial cells. For this functional test, cells are plated at day 0 in 96-well plates (10^3 cells/well) in 90% DMEM - 10% bovine calf serum. At day 1 the medium is substituted with fresh medium buffered with 50 mM

HEPES pH 7.4 and the supernatant containing the recombinant protein is added at different dilutions. At day 2 cells are washed with DPBS (Dulbecco's Phosphate Buffered Saline), fixed in 11% glutaraldehyde, stained with a Crystal-Violet solution and analysed by microscopy. The scattering activity is evaluated observing the morphology of the colonies, which are clustered in the negative control (non-stimulated cells or stimulated with supernatant containing no factors) whereas they are dispersed in the positive control (HGF-His). The morphology of the cells themselves also varies upon stimulation: in fact, as it can be observed in Fig. 6, cells stimulated with HGF-His and Metron Factor-1 have a more oblong, spindle-shaped form, characterised by protrusions of the cell membrane called pseudopodes. These morphological variations are the consequence of factor-induced activation of a genetic program involving the modification of a series of cellular parameters, such as digestion of cell matrix by specific proteases and increase in motility.

The Table summarises the results of different tests, obtained with factors HGF, NK2-HGF, Metron Factor-1 and Magic Factor-1 on MDCK cells. The scattering units reported indicate the maximum dilution of the conditioned supernatant containing the factor, at which motogenic activity could be observed. Values are normalised for the protein content determined by western blotting as described above (see Fig. 5). These data indicate that the hybrid factors Metron Factor-1 and Magic Factor-1 have a scattering activity approximately three magnitudes higher than that of the NK2-HGF-His truncated form and one magnitude higher than that of HGF-His parental factor.

	HGF-his	NK2-his	Metron F-1	Magic Factor-1
Scatter units	900 ± 29	6 ± 5	5500 ± 1532	7600 ± 150

Table. Scattering activity of factors HGF-His, NK2-HGF-His Metron Factor-1 measured on kidney epithelial cells (MDCK). The scattering units reported indicate the maximum dilution of the conditioned supernatant containing the factor, at which a motogenic activity can be observed. Values are normalised for the protein content determined by western blotting.

Example 4a: Test for the evaluation of protection against programmed cell death (apoptosis).

One of the most characterised side effect of the chemotherapeutic drug cisplatin is the induction of programmed cell death (apoptosis) of epithelial cells of the proximal tubule, which leads to acute renal failure (ARF). Thus, a factor that protects against cisplatin-induced cytotoxicity is highly desirable. An in vitro functional test has been used, which allows to evaluate the percentage of cisplatin-treated apoptotic cells in the presence or in the absence of a survival factor. This system utilises a cell line (LOC) derived from epithelial cells of human kidney proximal tubule, immortalised by ectopic expression of SV40 large T antigen. For the functional test, cells are plated at day 0 in 96-well plates (10^3 cells/well) in 90% DMEM - 10% bovine calf serum. At day 1, the medium is substituted with medium containing 0.5% bovine calf serum buffered with 50 mM HEPES pH 7.4, which is added with different dilutions of the supernatant containing the recombinant factor. Cells are pre-incubated with these

factors for 6 hours, and then further incubated in the presence of 10 μg/ml cisplatin. At day 2, cells are washed with DPBS and the percentage of apoptotic cells is evaluated by the TUNEL technique (Boehringer Mannheim). The same kind of tests can be performed using primary cultures of human epithelial cells of kidney proximal tubule (PTEC). These tests proved that Metron Factor-1 and Magic Factor-1 have protecting activity against cisplatin-induced programmed cell death.

Example 4b: Protection against cisplatin-induced cytotoxicity by transient gene delivery of Metron Factor-1 and Magic Factor-1

The protective effect of Metron F-1 and Magic F-1 against cisplatin-induced cytotoxicity was further demonstrated by a transient gene delivery approach. Simian kidney epithelial cells (COS) were transfected with a control empty vector, an expression vector for Metron F-1, or an expression vector for Magic F-1. Following transfection, cells were treated for 16 hours with cisplatin (20 μg/ml) and the percentage of surviving cells in each transfection was determined. Cisplatin treatment was calibrated to cause the death of approximately 20% of the cells in the negative control. Ectopic expression of Metron F-1 or Magic F-1 increased the survival rate to about 92.3% and 94.0%, respectively.

Example 5: Activation of the Met receptor by Metron Factor-1 and Magic Factor-1

The ability of Metron Factor-1 and Magic Factor-1 to activate the Met receptor was tested by analysing the ability of the recombinant factors to induce tyrosine phosphorylation of Met in human epithelial cells (A549). For this analysis, A549 cells at 90% confluence in a 100 mm petri

dish were stimulated for 10 minutes with 1 ml of conditioned supernatant containing Metron Factor-1, Magic Factor-1 or no factor (as negative control) diluted 1:2.5 or 1:10 in DMEM. After stimulation, cells were washed in ice with PBS, lysated in 200 µl of lysis solution (1% Triton X-100, 5 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 7.4), added with a cocktail of protease inhibitors, immunoprecipitated for 2 hours at 4° C with 10 µl of Sepharose-A beads covalently conjugated with an anti-Met monoclonal antibody (Naldini, L. et al., 1991, EMBO J. 10: 2867-2878), washed 3 times in the same lysis solution, and heated at 90°C for 2 minutes to elute the absorbed proteins. These were separated by SDS-PAGE on a 8% BIS-acrylamide gel, transferred onto a membrane (Hybond-C; Amersham) and analysed by western blot. A mouse monoclonal antibody against phosphotyrosine (UBI) diluted 1:10000 was used as primary antibody and an anti-mouse IgG ovine antibody conjugated with peroxidase (Amersham) was used as secondary antibody. The secondary antibody was detected by ECL (Amersham) following the protocol provided by the manufacturer. This analysis revealed that Metron F-1 and Magic F-1 potently activate the Met receptor (Fig. 7).

Example 6: Protection against chemotherapy-induced renal failure by Metron Factor-1 in vivo

Metron-F-1 was tested in a model of nephrotoxicity in Balb-c mice. The method used was substantially as described (Kawaida K et al., 1994, Hepatocyte growth factor prevents acute renal failure and accelerates renal regeneration in mice, Proc. Natl. Acad. Sci. 91:4357-4361). Briefly, renal failure was induced in male Balb-c mice weighing 20-25 g by an i.v.

injection of 7.5 mg/kg of $HgCl_2$ (7 animals/group). Renal damage was assessed by analysis of Blood Urea Nitrogen (BUN) and by histological evaluation, 72 h after $HgCl_2$ injection. Metron-F-1 was dissolved in 0.2 M NaCl, containing 0.01% Tween™ 80 and 0.25% human serum albumin and administered i.v. (100 µg/kg in a posological volume of 6.6 ml/kg) 0.5 h before and 6, 12, 24, 36 and 48 h after $HgCl_2$ injection. Controls animals were treated with the same amount of vehicle according to the same scheme.

Metron-F-1 significantly prevented the onset of acute renal failure induced by $HgCl_2$, evaluated in terms of BUN (figure 8). BUN values were closely paralleled by the histological findings, scored by an independent investigator.

In the following sequence listing:

SEQ. ID. NO. 1: Magic F-1 DNA coding sequence;

SEQ. ID. NO. 2: Magic F-1 amino acid sequence;

SEQ. ID. NO. 3: Metron F-1 DNA coding sequence;

SEQ. ID. NO. 4: Metron F-1 amino acid sequence.

SEQUENCE LISTING

<110> DOMPE' S.P.A

<120> RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

<130> 45166-NP

<140> 2,319,211
<141> 1999-01-27

<150> MI98A000179
<151> 1998-01-30

<160> 16

<170> PatentIn Ver. 2.1

<210> 1
<211> 1725
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: RECOMBINANT
C-DNA

<400> 1
atgtgggtga ccaaactcct gccagccctg ctgctgcagc atgtcctcct gcatctcctc 60
ctgctcccca tcgccatccc ctatgcagag ggacaaagga aaagaagaaa tacaattcat 120
gaattcaaaa aatcagcaaa gactacccta atcaaaatag atccagcact gaagataaaa 180
accaaaaaag tgaatactgc agaccaatgt gctaatagat gtactaggaa taaaggactt 240
ccattcactt gcaaggcttt tgtttttgat aaagcaagaa aacaatgcct ctggttcccc 300
ttcaatagca tgtcaagtgg agtgaaaaaa gaatttggcc atgaatttga cctctatgaa 360
aacaaagact acattagaaa ctgcatcatt ggtaaaggac gcagctacaa gggaacagta 420
tctatcacta agagtggcat caaatgtcag ccctggagtt ccatgatacc acacgaacac 480
agctatcggg gtaaagacct acaggaaaac tactgtcgaa atcctcgagg ggaagaaggg 540
ggaccctggt gtttcacaag caatccagag gtacgctacg aagtctgtga cattcctcag 600

tgttcagaag ttgaatgcat gacctgcaat ggggagagtt atcgaggtct catggatcat 660
acagaatcag gcaagatttg tcagcgctgg gatcatcaga caccacaccg gcacaaattc 720
ttgcctgaaa gatatcccga caagggcttt gatgataatt attgccgcaa tcccgatggc 780
cagccgaggc catggtgcta tactcttgac cctcacaccc gctgggagta ctgtgcaatt 840
aaaacatgcg ctgacaaagc ttcgggcggt ggcggttctg gtggcggtgg ctccggcggt 900
ggcggttctc tagagggaca aaggaaaaga agaaatacaa ttcatgaatt caaaaaatca 960
gcaaagacta ccctaatcaa aatagatcca gcactgaaga taaaaaccaa aaaagtgaat 1020
actgcagacc aatgtgctaa tagatgtact aggaataaag gacttccatt cacttgcaag 1080
gcttttgttt ttgataaagc aagaaaacaa tgcctctggt tccccttcaa tagcatgtca 1140
agtggagtga aaaaagaatt tggccatgaa tttgacctct atgaaaacaa agactacatt 1200
agaaactgca tcattggtaa aggacgcagc tacaagggaa cagtatctat cactaagagt 1260
ggcatcaaat gtcagccctg gagttccatg ataccacacg aacacagcta tcggggtaaa 1320
gacctacagg aaaactactg tcgaaatcct cgaggggaag aaggggggacc ctggtgtttc 1380
acaagcaatc cagaggtacg ctacgaagtc tgtgacattc ctcagtgttc agaagttgaa 1440
tgcatgacct gcaatgggga gagttatcga ggtctcatgg atcatacaga atcaggcaag 1500
atttgtcagc gctgggatca tcagacacca caccggcaca aattcttgcc tgaaagatat 1560
cccgacaagg gctttgatga taattattgc cgcaatcccg atggccagcc gaggccatgg 1620
tgctatactc ttgaccctca cacccgctgg gagtactgtg caattaaaac atgcgctgac 1680
aaagctgacg acgacgacaa acaccaccac caccaccacc actag 1725

<210> 2
<211> 574
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: RECOMBINANT
PROTEINS OBTAINED COMBINING HAIRPIN LOOP AND
KRINGLE DOMAINS OF HUMAN HGF AND MSP

<400> 2

Met	Trp	Val	Thr	Lys	Leu	Leu	Pro	Ala	Leu	Leu	Leu	Gln	His	Val	Leu
1				5				10						15	

Leu	His	Leu	Leu	Leu	Leu	Pro	Ile	Ala	Ile	Pro	Tyr	Ala	Glu	Gly	Gln
		20					25					30			

Arg Lys Arg Arg Asn Thr Ile His Glu Phe Lys Lys Ser Ala Lys Thr
35 40 45

Thr Leu Ile Lys Ile Asp Pro Ala Leu Lys Ile Lys Thr Lys Lys Val
50 55 60

Asn Thr Ala Asp Gln Cys Ala Asn Arg Cys Thr Arg Asn Lys Gly Leu
65 70 75 80

Pro Phe Thr Cys Lys Ala Phe Val Phe Asp Lys Ala Arg Lys Gln Cys
85 90 95

Leu Trp Phe Pro Phe Asn Ser Met Ser Ser Gly Val Lys Lys Glu Phe
100 105 110

Gly His Glu Phe Asp Leu Tyr Glu Asn Lys Asp Tyr Ile Arg Asn Cys
115 120 125

Ile Ile Gly Lys Gly Arg Ser Tyr Lys Gly Thr Val Ser Ile Thr Lys
130 135 140

Ser Gly Ile Lys Cys Gln Pro Trp Ser Ser Met Ile Pro His Glu His
145 150 155 160

Ser Tyr Arg Gly Lys Asp Leu Gln Glu Asn Tyr Cys Arg Asn Pro Arg
165 170 175

Gly Glu Glu Gly Gly Pro Trp Cys Phe Thr Ser Asn Pro Glu Val Arg
180 185 190

Tyr Glu Val Cys Asp Ile Pro Gln Cys Ser Glu Val Glu Cys Met Thr
195 200 205

Cys Asn Gly Glu Ser Tyr Arg Gly Leu Met Asp His Thr Glu Ser Gly
210 215 220

Lys Ile Cys Gln Arg Trp Asp His Gln Thr Pro His Arg His Lys Phe
225 230 235 240

Leu Pro Glu Arg Tyr Pro Asp Lys Gly Phe Asp Asp Asn Tyr Cys Arg
245 250 255

Asn Pro Asp Gly Gln Pro Arg Pro Trp Cys Tyr Thr Leu Asp Pro His
260 265 270

Thr Arg Trp Glu Tyr Cys Ala Ile Lys Thr Cys Ala Asp Lys Ala Ser
275 280 285

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Leu
290 295 300

Glu Gly Gln Arg Lys Arg Arg Asn Thr Ile His Glu Phe Lys Lys Ser
305 310 315 320

Ala Lys Thr Thr Leu Ile Lys Ile Asp Pro Ala Leu Lys Ile Lys Thr
325 330 335

Lys Lys Val Asn Thr Ala Asp Gln Cys Ala Asn Arg Cys Thr Arg Asn
340 345 350

Lys Gly Leu Pro Phe Thr Cys Lys Ala Phe Val Phe Asp Lys Ala Arg
355 360 365

Lys Gln Cys Leu Trp Phe Pro Phe Asn Ser Met Ser Ser Gly Val Lys
370 375 380

Lys Glu Phe Gly His Glu Phe Asp Leu Tyr Glu Asn Lys Asp Tyr Ile
385 390 395 400

Arg Asn Cys Ile Ile Gly Lys Gly Arg Ser Tyr Lys Gly Thr Val Ser
405 410 415

Ile Thr Lys Ser Gly Ile Lys Cys Gln Pro Trp Ser Ser Met Ile Pro
420 425 430

His Glu His Ser Tyr Arg Gly Lys Asp Leu Gln Glu Asn Tyr Cys Arg
435 440 445

Asn Pro Arg Gly Glu Glu Gly Gly Pro Trp Cys Phe Thr Ser Asn Pro
450 455 460

Glu Val Arg Tyr Glu Val Cys Asp Ile Pro Gln Cys Ser Glu Val Glu
465 470 475 480

Cys Met Thr Cys Asn Gly Glu Ser Tyr Arg Gly Leu Met Asp His Thr
485 490 495

Glu Ser Gly Lys Ile Cys Gln Arg Trp Asp His Gln Thr Pro His Arg
500 505 510

His Lys Phe Leu Pro Glu Arg Tyr Pro Asp Lys Gly Phe Asp Asp Asn
515 520 525

Tyr Cys Arg Asn Pro Asp Gly Gln Pro Arg Pro Trp Cys Tyr Thr Leu
530 535 540

Asp Pro His Thr Arg Trp Glu Tyr Cys Ala Ile Lys Thr Cys Ala Asp
545 550 555 560

Lys Ala Asp Asp Asp Asp Lys His His His His His His His
565 570

<210> 3
<211> 1692
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: RECOMBINANT
C-DNA

<400> 3
atggggtggc tcccactcct gctgcttctg actcaatgct taggggtccc tgggcagcgc 60
tcgccattga atgacttcca agtgctccgg ggcacagagc tacagcacct gctacatgcg 120
gtggtgcccg ggccttggca ggaggatgtg gcagatgctg aagagtgtgc tggtcgctgt 180
gggcccttaa tggactgccg ggccttccac tacaacgtga gcagccatgg ttgccaactg 240

ctgccatgga ctcaacactc gccccacacg aggctgcggc gttctgggcg ctgtgacctc 300
ttccagaaga aagactacgt acggacctgc atcatgaaca atggggttgg gtaccggggc 360
accatggcca cgaccgtggg tggcctgccc tgccaggctt ggagccacaa gttcccgaat 420
gatcacaagt acacgcccac tctccggaat ggcctggaag agaacttctg ccgtaaccct 480
gatggcgacc ccggaggtcc ttggtgctac acaacagacc ctgctgtgcg cttccagagc 540
tgcggcatca aatcctgccg ggaggccgcg tgtgtctggt gcaatggcga ggaataccgc 600
ggcgcggtag accgcacgga gtcagggcgc gagtgccagc gctgggatct tcagcacccg 660
caccagcacc ccttcgagcc gggcaagttc ctcgaccaag gtctggacga caactattgc 720
cggaatcctg acggctccga gcggccatgg tgctacacta cggatccgca gatcgagcga 780
gagttctgtg acctcccccg ctgcgggtcc gaggcacagc cccgcctcga gggcggtggc 840
ggttctggtg gcggtggctc cggcggtggc ggttctctag agggacaaag gaaaagaaga 900
aatacaattc atgaattcaa aaaatcagca aagactaccc taatcaaaat agatccagca 960
ctgaagataa aaaccaaaaa agtgaatact gcagaccaat gtgctaatag atgtactagg 1020
aataaaggac ttccattcac ttgcaaggct tttgtttttg ataaagcaag aaaacaatgc 1080
ctctggttcc ccttcaatag catgtcaagt ggagtgaaaa aagaatttgg ccatgaattt 1140
gacctctatg aaaacaaaga ctacattaga aactgcatca ttggtaaagg acgcagctac 1200
aagggaacag tatctatcac taagagtggc atcaaatgtc agccctggag ttccatgata 1260
ccacacgaac acagctatcg gggtaaagac ctacaggaaa actactgtcg aaatcctcga 1320
ggggaagaag ggggaccctg gtgtttcaca agcaatccag aggtacgcta cgaagtctgt 1380
gacattcctc agtgttcaga agttgaatgc atgacctgca atggggagag ttatcgaggt 1440
ctcatggatc atacagaatc aggcaagatt tgtcagcgct gggatcatca gacaccacac 1500
cggcacaaat tcttgcctga aagatatccc gacaagggct ttgatgataa ttattgccgc 1560
aatcccgatg gccagccgag gccatggtgc tatactcttg accctcacac ccgctgggag 1620
tactgtgcaa ttaaaacatg cgctgacaaa gctgacgacg acgacaaaca ccaccaccac 1680
caccaccact ag 1692

<210> 4
<211> 563
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: RECOMBINANT PROTEINS OBTAINED COMBINING HAIRPIN LOOP AND KRINGLE DOMAINS OF HUMAN HGF AND MSP

<400> 4

Met Gly Trp Leu Pro Leu Leu Leu Leu Leu Thr Gln Cys Leu Gly Val
1 5 10 15

Pro Gly Gln Arg Ser Pro Leu Asn Asp Phe Gln Val Leu Arg Gly Thr
20 25 30

Glu Leu Gln His Leu Leu His Ala Val Val Pro Gly Pro Trp Gln Glu
35 40 45

Asp Val Ala Asp Ala Glu Glu Cys Ala Gly Arg Cys Gly Pro Leu Met
50 55 60

Asp Cys Arg Ala Phe His Tyr Asn Val Ser Ser His Gly Cys Gln Leu
65 70 75 80

Leu Pro Trp Thr Gln His Ser Pro His Thr Arg Leu Arg Arg Ser Gly
85 90 95

Arg Cys Asp Leu Phe Gln Lys Lys Asp Tyr Val Arg Thr Cys Ile Met
100 105 110

Asn Asn Gly Val Gly Tyr Arg Gly Thr Met Ala Thr Thr Val Gly Gly
115 120 125

Leu Pro Cys Gln Ala Trp Ser His Lys Phe Pro Asn Asp His Lys Tyr
130 135 140

Thr Pro Thr Leu Arg Asn Gly Leu Glu Glu Asn Phe Cys Arg Asn Pro
145 150 155 160

Asp Gly Asp Pro Gly Gly Pro Trp Cys Tyr Thr Thr Asp Pro Ala Val
165 170 175

Arg Phe Gln Ser Cys Gly Ile Lys Ser Cys Arg Glu Ala Ala Cys Val
180 185 190

Trp Cys Asn Gly Glu Glu Tyr Arg Gly Ala Val Asp Arg Thr Glu Ser
195 200 205

Gly Arg Glu Cys Gln Arg Trp Asp Leu Gln His Pro His Gln His Pro
210 215 220

Phe Glu Pro Gly Lys Phe Leu Asp Gln Gly Leu Asp Asp Asn Tyr Cys
225 230 235 240

Arg Asn Pro Asp Gly Ser Glu Arg Pro Trp Cys Tyr Thr Thr Asp Pro
245 250 255

Gln Ile Glu Arg Glu Phe Cys Asp Leu Pro Arg Cys Gly Ser Glu Ala
260 265 270

Gln Pro Arg Leu Glu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
275 280 285

Gly Gly Gly Ser Leu Glu Gly Gln Arg Lys Arg Arg Asn Thr Ile His
290 295 300

Glu Phe Lys Lys Ser Ala Lys Thr Thr Leu Ile Lys Ile Asp Pro Ala
305 310 315 320

Leu Lys Ile Lys Thr Lys Lys Val Asn Thr Ala Asp Gln Cys Ala Asn
325 330 335

Arg Cys Thr Arg Asn Lys Gly Leu Pro Phe Thr Cys Lys Ala Phe Val
340 345 350

Phe Asp Lys Ala Arg Lys Gln Cys Leu Trp Phe Pro Phe Asn Ser Met
355 360 365

Ser Ser Gly Val Lys Lys Glu Phe Gly His Glu Phe Asp Leu Tyr Glu
370 375 380

Asn Lys Asp Tyr Ile Arg Asn Cys Ile Ile Gly Lys Gly Arg Ser Tyr
385 390 395 400

Lys Gly Thr Val Ser Ile Thr Lys Ser Gly Ile Lys Cys Gln Pro Trp
405 410 415

Ser Ser Met Ile Pro His Glu His Ser Tyr Arg Gly Lys Asp Leu Gln
420 425 430

Glu Asn Tyr Cys Arg Asn Pro Arg Gly Glu Glu Gly Gly Pro Trp Cys
435 440 445

Phe Thr Ser Asn Pro Glu Val Arg Tyr Glu Val Cys Asp Ile Pro Gln
450 455 460

Cys Ser Glu Val Glu Cys Met Thr Cys Asn Gly Glu Ser Tyr Arg Gly
465 470 475 480

Leu Met Asp His Thr Glu Ser Gly Lys Ile Cys Gln Arg Trp Asp His
485 490 495

Gln Thr Pro His Arg His Lys Phe Leu Pro Glu Arg Tyr Pro Asp Lys
500 505 510

Gly Phe Asp Asp Asn Tyr Cys Arg Asn Pro Asp Gly Gln Pro Arg Pro
515 520 525

Trp Cys Tyr Thr Leu Asp Pro His Thr Arg Trp Glu Tyr Cys Ala Ile
530 535 540

Lys Thr Cys Ala Asp Lys Ala Asp Asp Asp Asp Lys His His His His
545 550 555 560

His His His

<210> 5
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 5
cgcgcggaat tccaccatgg ggtggctccc actcct 36

<210> 6
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 6
cgcgcgctcg aggcggggct gtgcctcgga cccgca 36

<210> 7
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 7
cgcgcgtcta gagggacaaa ggaaaagaag aaatac 36

<210> 8
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 8

cgcgcgaagc tttgtcagcg catgttttaa ttgcac 36

<210> 9

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 9

tcgagggcgg tggcggttct ggtggcggtg gctccggcgg tggcggttct 50

<210> 10

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 10

ctagagaacc gccaccgccg gagccaccgc caccagaacc gccaccgccc 50

<210> 11

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 11

agctgacgac gacgacaaac accaccacca ccaccaccac tagggtcgac 50

<210> 12

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

OLIGONUCLEOTIDE PRIMER

<400> 12

agctgtcgac cctagtggtg gtggtggtgg tggtgtttgt cgtcgtcgtc 50

<210> 13

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

OLIGONUCLEOTIDE PRIMER

<400> 13

cgcgcgggat ccgccagccg ctccagcagc accatg 36

<210> 14

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

OLIGONUCLEOTIDE PRIMER

<400> 14

cgcgcgaagc tttgtcagcg catgttttaa ttgcac 36

<210> 15

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 15

agcttcgggc ggtggcggtt ctggtggcgg tggctccggc ggtggcggtt ct 52

<210> 16

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
OLIGONUCLEOTIDE PRIMER

<400> 16

ctagagaacc gccaccgccg gagccaccgc caccagaacc gccaccgccc ga 52

Claims

1. A recombinant protein comprising two superdomains, separated by a spacer sequence (linker), each superdomain consisting of a combination of HL and K1-K4 domains from the alpha chain of either HGF or MSP, according to the formula (I):

$$[A] - B - [C] - (D)y \qquad (I)$$

in which

[A] corresponds to the sequence $(LS)_m$-HL-K1-$(K2)_n$-$(K3)_o$-$(K4)_p$

wherein the numbering of the following amino acids refers to the HGF and MSP sequences as provided in Figures 1a and 1b, respectively:

LS is an amino acid sequence corresponding to residues 1-31 of HGF or 1-18 of MSP;

HL is an amino acid sequence starting between residues 32-70 of HGF α chain and ending between residues 96-127 of the identical chain; or it is an amino acid sequence starting between residues 19-56 of MSP α chain and ending between residues 78-109 of the identical chain;

K1 is an amino acid sequence starting between residues 97-128 of HGF α chain and ending between residues 201-205 of the identical chain; or it is an amino acid sequence starting between residues 79-110 of MSP α chain and ending between residues 186-190 of the identical chain;

K2 is an amino acid sequence starting between residues 202-206 of HGF α chain and ending between residues 283-299 of the identical chain; or it is an

amino acid sequence starting between residues 187-191 of MSP α chain and ending between residues 268-282 of the identical chain;

K3 is an amino acid sequence starting between residues 284-300 of HGF α chain and ending between residues 378-385 of the identical chain; or it is an amino acid sequence starting between residues 269-283 of MSP α chain and ending between residues 361-369 of the identical chain;

K4 is an amino acid sequence starting between residues 379-386 of HGF α chain and ending between residues 464-487 of the identical chain; or it is an amino acid sequence starting between residues 362-370 of MSP α chain and ending between residues 448-481 of the identical chain;

m, n, o, p are 0 or 1;

the sum n + o + p is an integer from 1 to 3 or 0, with the proviso that $n \geq o \geq p$;

B is the sequence $[(X)_q\ Y]_r$, wherein X = Gly and Y = Ser, or Cys, or Met, or Ala;

q is an integer from 2 to 8;

r is an integer from 1 to 9;

[C] corresponds to the sequence HL-K1-$(K2)_s$-$(K3)_t$-$(K4)_u$

wherein HL, K1-K4 are as defined above,

s, t, u are 0 or 1; the sum s + t + u is an integer from 1 to 3 or 0, with the proviso that $s \geq t \geq u$;

D is the sequence W-Z, wherein W is a proteolytic site, Z is a tag sequence for the purification and detection of the protein; y is 0 or 1.

2. The recombinant protein according to claim 1, in which the HL domain is a sequence of HGF α chain ranging from amino acids 32 to 127, or a sequence of MPS α chain ranging from amino acids 19 to 98; the K1 domain is a sequence of HGF α chain ranging from amino acids 128 to 203, or a sequence of MPS α chain ranging from amino acids 99 to 188; the K2 domain is a sequence of HGF α chain ranging from amino acids 204 to 294, or a sequence of MPS α chain ranging from amino acids 189 to 274; the K3 domain is a sequence of HGF α chain ranging from amino acids 286 to 383, or a sequence of MPS α chain ranging from amino acids 275 to 367; the K4 domain is a sequence of HGF α chain ranging from amino acids 384 to 487, or a sequence of MPS α chain ranging from amino acids 368 to 477.

3. The recombinant protein according to claim 1 or 2 of formula (II):

$$LS_{MSP}\text{-}HL_{MSP}\text{-}K1_{MSP}\text{-}K2_{MSP}\text{-}L\text{-}HL_{HGF}\text{-}K1_{HGF}\text{-}K2_{HGF}\text{-}D \qquad (II)$$

in which LS_{MSP} is the sequence 1-18 of MSP, HL_{MSP} is the sequence 19-56 of MSP, $K1_{MSP}$ is the sequence 99-188 of MSP, $K2_{MSP}$ is the sequence 189-274 of MSP, HL_{HGF} is the sequence 32-127 of HGF, $K1_{HGF}$ is the sequence 128-203 of HGF, $K2_{HGF}$ is the sequence 204-294 of HGF, L is the sequence $(Gly_4Ser)_3$, D is the sequence Asp_4-Lys-His_6.

4. The recombinant protein according to claim 1 or 2 of formula (III):

$$LS_{HGF}\text{-}HL_{HGF}\text{-}K1_{HGF}\text{-}K2_{HGF}\text{-}L\text{-}HL_{HGF}\text{-}K1_{HGF}\text{-}K2_{HGF}\text{-}D \qquad (III)$$

in which

LS_{HGF} is an amino acid sequence corresponding to residues 1-31 of HGF;

HL_{HGF} is an amino acid sequence starting between residues 32-70 of HGF α chain and ending between residues 96-127 of the identical chain;

$K1_{HGF}$ is an amino acid sequence starting between residues 97-128 of HGF α chain and ending between residues 201-205 of the identical chain;

$K2_{HGF}$ is an amino acid sequence starting between residues 202-206 of HGF α chain and ending between residues 283-299 of the identical chain;

L is the sequence $(Gly_4Ser)_3$;

D is the sequence Asp_4-Lys-His_6; and

LS_{HGF} is the sequence 1-31 of HGF.

5. A nucleotide sequence encoding for the recombinant protein of any one of claims 1-4.

6. An expression vector comprising the nucleotide sequence of claim 5.

7. A prokaryotic or eukaryotic host cell transformed with the expression vector of claim 6.

8. A process for preparing the recombinant protein of any one of claims 1-4, which comprises the following steps:

a) construction of DNA encoding the desired protein;

b) insertion of DNA in an expression vector;

c) transformation of a host cell with recombinant DNA (rDNA);

d) culture of the transformed host cell so as to express the recombinant protein; and

e) extraction and purification of the produced recombinant protein.

9. The process according to claim 8, wherein the host cell is a kidney epithelial BOSC cell or an SF9 insect cell.

10. The recombinant protein of any one of claims 1-4 for use as a prophylactic or therapeutic agent to treat toxic side effects of a chemotherapeutic agent.

11. Use of the recombinant protein of any one of claims 1-4 in the manufacture of a medicament for the prevention or treatment of chemotherapeutic-induced toxicity.

12. Use according to claim 11, wherein the chemotherapeutic-induced toxicity is selected from: myelotoxicity, kidney toxicity, neurotoxicity, mucotoxicity and hepatotoxicity.

13. A pharmaceutical composition containing an effective amount of the recombinant protein of any one of claims 1-4, in combination with at least one pharmacologically acceptable excipient.

FIG 1a

```
1   ATGTGGGTGACCAAACTCCTGCCAGCCCTGCTGCTGCAGCATGTCCTCCTGCATCTCCTC   60
    TACACCCACTGGTTTGAGGACGGTCGGGACGACGACGTCGTACAGGAGGACGTAGAGGAG

    1  M  W  V  T  K  L  L  P  A  L  L  L  Q  H  V  L  L  H  L  L   20

61  CTGCTCCCCATCGCCATCCCCTATGCAGAGGGACAAAGGAAAAGAAGAAATACAATTCAT   120
    GACGAGGGGTAGCGGTAGGGGATACGTCTCCCTGTTTCCTTTTCTTCTTTATGTTAAGTA

    21 L  L  P  I  A  I  P  Y  A  E  G  Q  R  K  R  R  N  T  I  H   40

121 GAATTCAAAAAATCAGCAAAGACTACCCTAATCAAAATAGATCCAGCACTGAAGATAAAA   180
    CTTAAGTTTTTTAGTCGTTTCTGATGGGATTAGTTTTATCTAGGTCGTGACTTCTATTTT

    41 E  F  K  K  S  A  K  T  T  L  I  K  I  D  P  A  L  K  I  K   60

181 ACCAAAAAAGTGAATACTGCAGACCAATGTGCTAATAGATGTACTAGGAATAAAGGACTT   240
    TGGTTTTTTCACTTATGACGTCTGGTTACACGATTATCTACATGATCCTTATTTCCTGAA

    61 T  K  K  V  N  T  A  D  Q  C  A  N  R  C  T  R  N  K  G  L   80

241 CCATTCACTTGCAAGGCTTTTGTTTTTGATAAAGCAAGAAAACAATGCCTCTGGTTCCCC   300
    GGTAAGTGAACGTTCCGAAAACAAAAACTATTTCGTTCTTTTGTTACGGAGACCAAGGGG

    81 P  F  T  C  K  A  F  V  F  D  K  A  R  K  Q  C  L  W  F  P   100

301 TTCAATAGCATGTCAAGTGGAGTGAAAAAAGAATTTGGCCATGAATTTGACCTCTATGAA   360
    AAGTTATCGTACAGTTCACCTCACTTTTTTCTTAAACCGGTACTTAAACTGGAGATACTT

    101 F  N  S  M  S  S  G  V  K  K  E  F  G  H  E  F  D  L  Y  E   120

361 AACAAAGACTACATTAGAAACTGCATCATTGGTAAAGGACGCAGCTACAAGGGAACAGTA   420
    TTGTTTCTGATGTAATCTTTGACGTAGTAACCATTTCCTGCGTCGATGTTCCCTTGTCAT

    121 N  K  D  Y  I  R  N  C  I  I  G  K  G  R  S  Y  K  G  T  V   140

421 TCTATCACTAAGAGTGGCATCAAATGTCAGCCCTGGAGTTCCATGATACCACACGAACAC   480
    AGATAGTGATTCTCACCGTAGTTTACAGTCGGGACCTCAAGGTACTATGGTGTGCTTGTG

    141 S  I  T  K  S  G  I  K  C  Q  P  W  S  S  M  I  P  H  E  H   160
```

(continued)

(continued)

```
481  AGCTATCGGGGTAAAGACCTACAGGAAAACTACTGTCGAAATCCTCGAGGGGAAGAAGGG  540
     TCGATAGCCCCATTTCTGGATGTCCTTTTGATGACAGCTTTAGGAGCTCCCCTTCTTCCC

161  S  Y  R  G  K  D  L  Q  E  N  Y  C  R  N  P  R  G  E  E  G   180

541  GGACCCTGGTGTTTCACAAGCAATCCAGAGGTACGCTACGAAGTCTGTGACATTCCTCAG  600
     CCTGGGACCACAAAGTGTTCGTTAGGTCTCCATGCGATGCTTCAGACACTGTAAGGAGTC

181  G  P  W  C  F  T  S  N  P  E  V  R  Y  E  V  C  D  I  P  Q   200

601  TGTTCAGAAGTTGAATGCATGACCTGCAATGGGGAGAGTTATCGAGGTCTCATGGATCAT  660
     ACAAGTCTTCAACTTACGTACTGGACGTTACCCCTCTCAATAGCTCCAGAGTACCTAGTA

201  C  S  E  V  E  C  M  T  C  N  G  E  S  Y  R  G  L  M  D  H   220

661  ACAGAATCAGGCAAGATTTGTCAGCGCTGGGATCATCAGACACCACACCGGCACAAATTC  720
     TGTCTTAGTCCGTTCTAAACAGTCGCGACCCTAGTAGTCTGTGGTGTGGCCGTGTTTAAG

221  T  E  S  G  K  I  C  Q  R  W  D  H  Q  T  P  H  R  H  K  F   240

721  TTGCCTGAAAGATATCCCGACAAGGGCTTTGATGATAATTATTGCCGCAATCCCGATGGC  780
     AACGGACTTTCTATAGGGCTGTTCCCGAAACTACTATTAATAACGGCGTTAGGGCTACCG

241  L  P  E  R  Y  P  D  K  G  F  D  D  N  Y  C  R  N  P  D  G   260

781  CAGCCGAGGCCATGGTGCTATACTCTTGACCCTCACACCCGCTGGGAGTACTGTGCAATT  840
     GTCGGCTCCGGTACCACGATATGAGAACTGGGAGTGTGGGCGACCCTCATGACACGTTAA

261  Q  P  R  P  W  C  Y  T  L  D  P  H  T  R  W  E  Y  C  A  I   280

841  AAAACATGCGCTGACAATACTATGAATGACACTGATGTTCCTTTGGAAACAACTGAATGC  900
     TTTTGTACGCGACTGTTATGATACTTACTGTGACTACAAGGAAACCTTTGTTGACTTACG

281  K  T  C  A  D  N  T  M  N  D  T  D  V  P  L  E  T  T  E  C   300

901  ATCCAAGGTCAAGGAGAAGGCTACAGGGGCACTGTCAATACCATTTGGAATGGAATTCCA  960
     TAGGTTCCAGTTCCTCTTCCGATGTCCCCGTGACAGTTATGGTAAACCTTACCTTAAGGT

301  I  Q  G  Q  G  E  G  Y  R  G  T  V  N  T  I  W  N  G  I  P   320

961  TGTCAGCGTTGGGATTCTCAGTATCCTCACGAGCATGACATGACTCCTGAAAATTTCAAG  1020
     ACAGTCGCAACCCTAAGAGTCATAGGAGTGCTCGTACTGTACTGAGGACTTTTAAAGTTC

321  C  Q  R  W  D  S  Q  Y  P  H  E  H  D  M  T  P  E  N  F  K   340

1021 TGCAAGGACCTACGAGAAAATTACTGCCGAAATCCAGATGGGTCTGAATCACCCTGGTGT  1080
     ACGTTCCTGGATGCTCTTTTAATGACGGCTTTAGGTCTACCCAGACTTAGTGGGACCACA

341  C  K  D  L  R  E  N  Y  C  R  N  P  D  G  S  E  S  P  W  C   360

1081 TTTACCACTGATCCAAACATCCGAGTTGGCTACTGCTCCCAAATTCCAAACTGTGATATG  1140
     AAATGGTGACTAGGTTTGTAGGCTCAACCGATGACGAGGGTTTAAGGTTTGACACTATAC

361  F  T  T  D  P  N  I  R  V  G  Y  C  S  Q  I  P  N  C  D  M   380
```

(continued)

(continued)

```
                TCACATGGACAAGATTGTTATCGTGGGAATGGCAAAAATTATATGGGCAACTTATCCCAA
1141            ----------+----------+----------+----------+----------+----------+  1200
                AGTGTACCTGTTCTAACAATAGCACCCTTACCGTTTTTAATATACCCGTTGAATAGGGTT

 381  S   H   G   Q   D   C   Y   R   G   N   G   K   N   Y   M   G   N   L   S   Q    400

                ACAAGATCTGGACTAACATGTTCAATGTGGGACAAGAACATGGAAGACTTACATCGTCAT
1201            ----------+----------+----------+----------+----------+----------+  1260
                TGTTCTAGACCTGATTGTACAAGTTACACCCTGTTCTTGTACCTTCTGAATGTAGCAGTA

 401  T   R   S   G   L   T   C   S   M   W   D   K   N   M   E   D   L   H   R   H    420

                ATCTTCTGGGAACCAGATGCAAGTAAGCTGAATGAGAATTACTGCCGAAATCCAGATGAT
1261            ----------+----------+----------+----------+----------+----------+  1320
                TAGAAGACCCTTGGTCTACGTTCATTCGACTTACTCTTAATGACGGCTTTAGGTCTACTA

 421  I   F   W   E   P   D   A   S   K   L   N   E   N   Y   C   R   N   P   D   D    440

                GACGCTCATGGACCCTGGTGCTACACGGGAAATCCACTCATTCCTTGGGATTATTGCCCT
1321            ----------+----------+----------+----------+----------+----------+  1380
                CTGCGAGTACCTGGGACCACGATGTGCCCTTTAGGTGAGTAAGGAACCCTAATAACGGGA

 441  D   A   H   G   P   W   C   Y   T   G   N   P   L   I   P   W   D   Y   C   P    460

                ATTTCTCGTTGTGAAGGTGATACCACACCTACAATAGTCAATTTAGACCATCCCGTAATA
1381            ----------+----------+----------+----------+----------+----------+  1440
                TAAAGAGCAACACTTCCACTATGGTGTGGATGTTATCAGTTAAATCTGGTAGGGCATTAT

 461  I   S   R   C   E   G   D   T   T   P   T   I   V   N   L   D   H   P   V   I    480

                TCTTGTGCCAAAACGAAACAATTGCGAGTTGTAAATGGGATTCCAACACGAACAAACATA
1441            ----------+----------+----------+----------+----------+----------+  1500
                AGAACACGGTTTTGCTTTGTTAACGCTCAACATTTACCCTAAGGTTGTGCTTGTTTGTAT

 481  S   C   A   K   T   K   Q   L   R   V   V   N   G   I   P   T   R   T   N   I    500

                GGATGGATGGTTAGTTTGAGATACAGAAATAAACATATCTGCGGAGGATCATTGATAAAG
1501            ----------+----------+----------+----------+----------+----------+  1560
                CCTACCTACCAATCAAACTCTATGTCTTTATTTGTATAGACGCCTCCTAGTAACTATTTC

 501  G   W   M   V   S   L   R   Y   R   N   K   H   I   C   G   G   S   L   I   K    520

                GAGAGTTGGGTTCTTACTGCACGACAGTGTTTCCCTTCTCGAGACTTGAAAGATTATGAA
1561            ----------+----------+----------+----------+----------+----------+  1620
                CTCTCAACCCAAGAATGACGTGCTGTCACAAAGGGAAGAGCTCTGAACTTTCTAATACTT

 521  E   S   W   V   L   T   A   R   Q   C   F   P   S   R   D   L   K   D   Y   E    540

                GCTTGGCTTGGAATTCATGATGTCCACGGAAGAGGAGATGAGAAATGCAAACAGGTTCTC
1621            ----------+----------+----------+----------+----------+----------+  1680
                CGAACCGAACCTTAAGTACTACAGGTGCCTTCTCCTCTACTCTTTACGTTTGTCCAAGAG

 541  A   W   L   G   I   H   D   V   H   G   R   G   D   E   K   C   K   Q   V   L    560

                AATGTTTCCCAGCTGGTATATGGCCCTGAAGGATCAGATCTGGTTTTAATGAAGCTTGCC
1681            ----------+----------+----------+----------+----------+----------+  1740
                TTACAAAGGGTCGACCATATACCGGGACTTCCTAGTCTAGACCAAAATTACTTCGAACGG

 561  N   V   S   Q   L   V   Y   G   P   E   G   S   D   L   V   L   M   K   L   A    580

                AGGCCTGCTGTCCTGGATGATTTTGTTAGTACGATTGATTTACCTAATTATGGATGCACA
1741            ----------+----------+----------+----------+----------+----------+  1800
                TCCGGACGACAGGACCTACTAAAACAATCATGCTAACTAAATGGATTAATACCTACGTGT

 581  R   P   A   V   L   D   D   F   V   S   T   I   D   L   P   N   Y   G   C   T    600
```

(continued)

(continued)

```
                ATTCCTGAAAAGACCAGTTGCAGTGTTTATGGCTGGGGCTACACTGGATTGATCAACTAT
1801            ---------+---------+---------+---------+---------+---------+  1860
                TAAGGACTTTTCTGGTCAACGTCACAAATACCGACCCCGATGTGACCTAACTAGTTGATA

 601  I  P  E  K  T  S  C  S  V  Y  G  W  G  Y  T  G  L  I  N  Y    620

                GATGGCCTATTACGAGTGGCACATCTCTATATAATGGGAAATGAGAAATGCAGCCAGCAT
1861            ---------+---------+---------+---------+---------+---------+  1920
                CTACCGGATAATGCTCACCGTGTAGAGATATATTACCCTTTACTCTTTACGTCGGTCGTA

 621  D  G  L  L  R  V  A  H  L  Y  I  M  G  N  E  K  C  S  Q  H    640

                CATCGAGGGAAGGTGACTCTGAATGAGTCTGAAATATGTGCTGGGGCTGAAAAGATTGGA
1921            ---------+---------+---------+---------+---------+---------+  1980
                GTAGCTCCCTTCCACTGAGACTTACTCAGACTTTATACACGACCCCGACTTTTCTAACCT

 641  H  R  G  K  V  T  L  N  E  S  E  I  C  A  G  A  E  K  I  G    660

                TCAGGACCATGTGAGGGGGATTATGGTGGCCCACTTGTTTGTGAGCAACATAAAATGAGA
1981            ---------+---------+---------+---------+---------+---------+  2040
                AGTCCTGGTACACTCCCCCTAATACCACCGGGTGAACAAACACTCGTTGTATTTTACTCT

 661  S  G  P  C  E  G  D  Y  G  G  P  L  V  C  E  Q  H  K  M  R    680

                ATGGTTCTTGGTGTCATTGTTCCTGGTCGTGGATGTGCCATTCCAAATCGTCCTGGTATT
2041            ---------+---------+---------+---------+---------+---------+  2100
                TACCAAGAACCACAGTAACAAGGACCAGCACCTACACGGTAAGGTTTAGCAGGACCATAA

 681  M  V  L  G  V  I  V  P  G  R  G  C  A  I  P  N  R  P  G  I    700

                TTTGTCCGAGTAGCATATTATGCAAAATGGATACACAAAATTATTTTAACATATAAGGTA
2101            ---------+---------+---------+---------+---------+---------+  2160
                AAACAGGCTCATCGTATAATACGTTTTACCTATGTGTTTTAATAAAATTGTATATTCCAT

 701  F  V  R  V  A  Y  Y  A  K  W  I  H  K  I  I  L  T  Y  K  V    720

                CCACAGTCATAG
2161            ---------+==   2172
                GGTGTCAGTATC

 721  P  Q  S  *    723
```

FIG 1b

```
1   ATGGGGTGGCTCCCACTCCTGCTGCTTCTGACTCAATGCTTAGGGGTCCCTGGGCAGCGC
    ----------+---------+---------+---------+---------+---------+ 60
    TACCCCACCGAGGGTGAGGACGACGAAGACTGAGTTACGAATCCCCAGGGACCCGTCGCG

1   M  G  W  L  P  L  L  L  L  L  T  Q  C  L  G  V  P  G  Q  R    20

    TCGCCATTGAATGACTTCCAAGTGCTCCGGGGCACAGAGCTACAGCACCTGCTACATGCG
    ---------+---------+---------+---------+---------+---------+ 120
61  AGCGGTAACTTACTGAAGGTTCACGAGGCCCCGTGTCTCGATGTCGTGGACGATGTACGC

21  S  P  L  N  D  F  Q  V  L  R  G  T  E  L  Q  H  L  L  H  A    40

    GTGGTGCCCGGGCCTTGGCAGGAGGATGTGGCAGATGCTGAAGAGTGTGCTGGTCGCTGT
    ---------+---------+---------+---------+---------+---------+ 180
121 CACCACGGGCCCGGAACCGTCCTCCTACACCGTCTACGACTTCTCACACGACCAGCGACA

41  V  V  P  G  P  W  Q  E  D  V  A  D  A  E  E  C  A  G  R  C    60

    GGGCCCTTAATGGACTGCCGGGCCTTCCACTACAACGTGAGCAGCCATGGTTGCCAACTG
    ---------+---------+---------+---------+---------+---------+ 240
181 CCCGGGAATTACCTGACGGCCCGGAAGGTGATGTTGCACTCGTCGGTACCAACGGTTGAC

61  G  P  L  M  D  C  R  A  F  H  Y  N  V  S  S  H  G  C  Q  L    80

    CTGCCATGGACTCAACACTCGCCCCACACGAGGCTGCGGCGTTCTGGGCGCTGTGACCTC
    ---------+---------+---------+---------+---------+---------+ 300
241 GACGGTACCTGAGTTGTGAGCGGGGTGTGCTCCGACGCCGCAAGACCCGCGACACTGGAG

81  L  P  W  T  Q  H  S  P  H  T  R  L  R  R  S  G  R  C  D  L   100

    TTCCAGAAGAAAGACTACGTACGGACCTGCATCATGAACAATGGGGTTGGGTACCGGGGC
    ---------+---------+---------+---------+---------+---------+ 360
301 AAGGTCTTCTTTCTGATGCATGCCTGGACGTAGTACTTGTTACCCCAACCCATGGCCCCG

101 F  Q  K  K  D  Y  V  R  T  C  I  M  N  N  G  V  G  Y  R  G   120

    ACCATGGCCACGACCGTGGGTGGCCTGCCCTGCCAGGCTTGGAGCCACAAGTTCCCGAAT
    ---------+---------+---------+---------+---------+---------+ 420
361 TGGTACCGGTGCTGGCACCCACCGGACGGGACGGTCCGAACCTCGGTGTTCAAGGGCTTA

121 T  M  A  T  T  V  G  G  L  P  C  Q  A  W  S  H  K  F  P  N   140

    GATCACAAGTACACGCCCACTCTCCGGAATGGCCTGGAAGAGAACTTCTGCCGTAACCCT
    ---------+---------+---------+---------+---------+---------+ 480
421 CTAGTGTTCATGTGCGGGTGAGAGGCCTTACCGGACCTTCTCTTGAAGACGGCATTGGGA

141 D  H  K  Y  T  P  T  L  R  N  G  L  E  E  N  F  C  R  N  P   160
```

(continued)

(continued)

```
481  GATGGCGACCCCGGAGGTCCTTGGTGCTACACAACAGACCCTGCTGTGCGCTTCCAGAGC  540
     ---------+---------+---------+---------+---------+---------+
     CTACCGCTGGGGCCTCCAGGAACCACGATGTGTTGTCTGGGACGACACGCGAAGGTCTCG

161  D  G  D  P  G  G  P  W  C  Y  T  T  D  P  A  V  R  F  Q  S   180

531  TGCGGCATCAAATCCTGCCGGGAGGCCGCGTGTGTCTGGTGCAATGGCGAGGAATACCGC  600
     ---------+---------+---------+---------+---------+---------+
     ACGCCGTAGTTTAGGACGGCCCTCCGGCGCACACAGACCACGTTACCGCTCCTTATGGCG

181  C  G  I  K  S  C  R  E  A  A  C  V  W  C  N  G  E  E  Y  R   200

601  GGCGCGGTAGACCGCACGGAGTCAGGGCGCGAGTGCCAGCGCTGGGATCTTCAGCACCCG  660
     ---------+---------+---------+---------+---------+---------+
     CCGCGCCATCTGGCGTGCCTCAGTCCCGCGCTCACGGTCGCGACCCTAGAAGTCGTGGGC

201  G  A  V  D  R  T  E  S  G  R  E  C  Q  R  W  D  L  Q  H  P   220

661  CACCAGCACCCCTTCGAGCCGGGCAAGTTCCTCGACCAAGGTCTGGACGACAACTATTGC  720
     ---------+---------+---------+---------+---------+---------+
     GTGGTCGTGGGGAAGCTCGGCCCGTTCAAGGAGCTGGTTCCAGACCTGCTGTTGATAACG

221  H  Q  H  P  F  E  P  G  K  F  L  D  Q  G  L  D  D  N  Y  C   240

721  CGGAATCCTGACGGCTCCGAGCGGCCATGGTGCTACACTACGGATCCGCAGATCGAGCGA  780
     ---------+---------+---------+---------+---------+---------+
     GCCTTAGGACTGCCGAGGCTCGCCGGTACCACGATGTGATGCCTAGGCGTCTAGCTCGCT

241  R  N  P  D  G  S  E  R  P  W  C  Y  T  T  D  P  Q  I  E  R   260

781  GAGTTCTGTGACCTCCCCCGCTGCGGGTCCGAGGCACAGCCCCGCCAAGAGGCCACAACT  840
     ---------+---------+---------+---------+---------+---------+
     CTCAAGACACTGGAGGGGGCGACGCCCAGGCTCCGTGTCGGGGCGGTTCTCCGGTGTTGA

261  E  F  C  D  L  P  R  C  G  S  E  A  Q  P  R  Q  E  A  T  T   280

841  GTCAGCTGCTTCCGCGGGAAGGGTGAGGGCTACCGGGGCACAGCCAATACCACCACTGCG  900
     ---------+---------+---------+---------+---------+---------+
     CAGTCGACGAAGGCGCCCTTCCCACTCCCGATGGCCCCGTGTCGGTTATGGTGGTGACGC

281  V  S  C  F  R  G  K  G  E  G  Y  R  G  T  A  N  T  T  T  A   300

901  GGCGTACCTTGCCAGCGTTGGGACGCGCAAATCCCGCATCAGCACCGATTTACGCCAGAA  960
     ---------+---------+---------+---------+---------+---------+
     CCGCATGGAACGGTCGCAACCCTGCGCGTTTAGGGCGTAGTCGTGGCTAAATGCGGTCTT

301  G  V  P  C  Q  R  W  D  A  Q  I  P  H  Q  H  R  F  T  P  E   320

961  AAATACGCGTGCAAAGACCTTCGGGAGAACTTCTGCCGGAACCCCGACGGCTCAGAGGCG  1020
     ---------+---------+---------+---------+---------+---------+
     TTTATGCGCACGTTTCTGGAAGCCCTCTTGAAGACGGCCTTGGGGCTGCCGAGTCTCCGC

321  K  Y  A  C  K  D  L  R  E  N  F  C  R  N  P  D  G  S  E  A   340

1021 CCCTGGTGCTTCACACTGCGGCCCGGCATGCGCGCGGCCTTTTGCTACCAGATCCGGCGT  1080
     ---------+---------+---------+---------+---------+---------+
     GGGACCACGAAGTGTGACGCCGGGCCGTACGCGCGCCGGAAAACGATGGTCTAGGCCGCA

341  P  W  C  F  T  L  R  P  G  M  R  A  A  F  C  Y  Q  I  R  R   360

1081 TGTACAGACGACGTGCGGCCCCAGGACTGCTACCACGGCGCAGGGGAGCAGTACCGCGGC  1140
     ---------+---------+---------+---------+---------+---------+
     ACATGTCTGCTGCACGCCGGGGTCCTGACGATGGTGCCGCGTCCCCTCGTCATGGCGCCG

361  C  T  D  D  V  R  P  Q  D  C  Y  H  G  A  G  E  Q  Y  R  G   380
```

(continued)

(continued)

```
                ACGGTCAGCAAGACCCGCAAGGGTGTCCAGTGCCAGCGCTGGTCCGCTGAGACGCCGCAC
1141            ---------+---------+---------+---------+---------+---------+  1200
                TGCCAGTCGTTCTGGGCGTTCCCACAGGTCACGGTCGCGACCAGGCGACTCTGCGGCGTG

 381  T  V  S  K  T  R  K  G  V  Q  C  Q  R  W  S  A  E  T  P  H      400

                AAGCCGCAGTTCACGTTTACCTCCGAACCGCATGCACAACTGGAGGAGAACTTCTGCCGG
1201            ---------+---------+---------+---------+---------+---------+  1260
                TTCGGCGTCAAGTGCAAATGGAGGCTTGGCGTACGTGTTGACCTCCTCTTGAAGACGGCC

 401  K  P  Q  F  T  F  T  S  E  P  H  A  Q  L  E  E  N  F  C  R      420

                AACCCAGATGGGGATAGCCATGGGCCCTGGTGCTACACGATGGACCCAAGGACCCCATTC
1261            ---------+---------+---------+---------+---------+---------+  1320
                TTGGGTCTACCCCTATCGGTACCCGGGACCACGATGTGCTACCTGGGTTCCTGGGGTAAG

 421  N  P  D  G  D  S  H  G  P  W  C  Y  T  M  D  P  R  T  P  F      440

                GACTACTGTGCCCTGCGACGCTGCGCTGATGACCAGCCGCCATCAATCCTGGACCCCCCA
1321            ---------+---------+---------+---------+---------+---------+  1380
                CTGATGACACGGGACGCTGCGACGCGACTACTGGTCGGCGGTAGTTAGGACCTGGGGGGT

 441  D  Y  C  A  L  R  R  C  A  D  D  Q  P  P  S  I  L  D  P  P      460

                GACCAGGTGCAGTTTGAGAAGTGTGGCAAGAGGGTGGATCGGCTGGATCAGCGGCGTTCC
1381            ---------+---------+---------+---------+---------+---------+  1440
                CTGGTCCACGTCAAACTCTTCACACCGTTCTCCCACCTAGCCGACCTAGTCGCCGCAAGG

 461  D  Q  V  Q  F  E  K  C  G  K  R  V  D  R  L  D  Q  R  R  S      480

                AAGCTGCGCGTGGTTGGGGGCCATCCGGGCAACTCACCCTGGACAGTCAGCTTGCGGAAT
1441            ---------+---------+---------+---------+---------+---------+  1500
                TTCGACGCGCACCAACCCCCGGTAGGCCCGTTGAGTGGGACCTGTCAGTCGAACGCCTTA

 481  K  L  R  V  V  G  G  H  P  G  N  S  P  W  T  V  S  L  R  N      500

                CGGCAGGGCCAGCATTTCTGCGGGGGGTCTCTAGTGAAGGAGCAGTGGATACTGACTGCC
1501            ---------+---------+---------+---------+---------+---------+  1560
                GCCGTCCCGGTCGTAAAGACGCCCCCCAGAGATCACTTCCTCGTCACCTATGACTGACGG

 501  R  Q  G  Q  H  F  C  G  G  S  L  V  K  E  Q  W  I  L  T  A      520

                CGGCAGTGCTTCTCCTCCTGCCATATGCCTCTCACGGGCTATGAGGTATGGTTGGGCACC
1561            ---------+---------+---------+---------+---------+---------+  1620
                GCCGTCACGAAGAGGAGGACGGTATACGGAGAGTGCCCGATACTCCATACCAACCCGTGG

 521  R  Q  C  F  S  S  C  H  M  P  L  T  G  Y  E  V  W  L  G  T      540

                CTGTTCCAGAACCCACAGCATGGAGAGCCAAGCCTACAGCGGGTCCCAGTAGCCAAGATG
1621            ---------+---------+---------+---------+---------+---------+  1680
                GACAAGGTCTTGGGTGTCGTACCTCTCGGTTCGGATGTCGCCCAGGGTCATCGGTTCTAC

 541  L  F  Q  N  P  Q  H  G  E  P  S  L  Q  R  V  P  V  A  K  M      560

                GTGTGTGGGCCCTCAGGCTCCCAGCTTGTCCTGCTCAAGCTGGAGAGATCTGTGACCCTG
1681            ---------+---------+---------+---------+---------+---------+  1740
                CACACACCCGGGAGTCCGAGGGTCGAACAGGACGAGTTCGACCTCTCTAGACACTGGGAC

 561  V  C  G  P  S  G  S  Q  L  V  L  L  K  L  E  R  S  V  T  L      580

                AACCAGCGTGTGGCCCTGATCTGCCTGCCCCCTGAATGGTATGTGGTGCCTCCAGGGACC
1741            ---------+---------+---------+---------+---------+---------+  1800
                TTGGTCGCACACCGGGACTAGACGGACGGGGGACTTACCATACACCACGGAGGTCCCTGG

 581  N  Q  R  V  A  L  I  C  L  P  P  E  W  Y  V  V  P  P  G  T      600
```

(continued)

(continued)

```
                AAGTGTGAGATTGCAGGCTGGGGTGAGACCAAAGGTACGGGTAATGACACAGTCCTAAAT
1801            ---------+---------+---------+---------+---------+---------+  1860
                TTCACACTCTAACGTCCGACCCCACTCTGGTTTCCATGCCCATTACTGTGTCAGGATTTA

 601  K  C  E  I  A  G  W  G  E  T  K  G  T  G  N  D  T  V  L  N    620

                GTGGCCTTTCTGAATGTTATCTCCAACCAGGAGTGTAACATCAAGCACCGAGGACGTGTG
1861            ---------+---------+---------+---------+---------+---------+  1920
                CACCGGAAAGACTTACAATAGAGGTTGGTCCTCACATTGTAGTTCGTGGCTCCTGCACAC

 621  V  A  F  L  N  V  I  S  N  Q  E  C  N  I  K  H  R  G  R  V    640

                CGGGAGAGTGAGATGTGCACTGAGGGACTGTTGGCCCCTGTGGGGGCCTGTGAGGGTGAC
1921            ---------+---------+---------+---------+---------+---------+  1980
                GCCCTCTCACTCTACACGTGACTCCCTGACAACCGGGGACACCCCCGGACACTCCCACTG

 641  R  E  S  E  M  C  T  E  G  L  L  A  P  V  G  A  C  E  G  D    660

                TACGGGGGCCCACTTGCCTGCTTTACCCACAACTGCTGGGTCCTGGAAGGAATTATAATC
1981            ---------+---------+---------+---------+---------+---------+  2040
                ATGCCCCCGGGTGAACGGACGAAATGGGTGTTGACGACCCAGGACCTTCCTTAATATTAG

 661  Y  G  G  P  L  A  C  F  T  H  N  C  W  V  L  E  G  I  I  I    680

                CCCAACCGAGTATGCGCAAGGTCCCGCTGGCCAGCTGTCTTCACGCGTGTCTCTGTGTTT
2041            ---------+---------+---------+---------+---------+---------+  2100
                GGGTTGGCTCATACGCGTTCCAGGGCGACCGGTCGACAGAAGTGCGCACAGAGACACAAA

 681  P  N  R  V  C  A  R  S  R  W  P  A  V  F  T  R  V  S  V  F    700

                GTGGACTGGATTCACAAGGTCATGAGACTGGGTTAG
2101            ---------+---------+---------+---=== 2136
                CACCTGACCTAAGTGTTCCAGTACTCTGACCCAATC

 701  V  D  W  I  H  K  V  M  R  L  G  *    711
```

FIG 2a

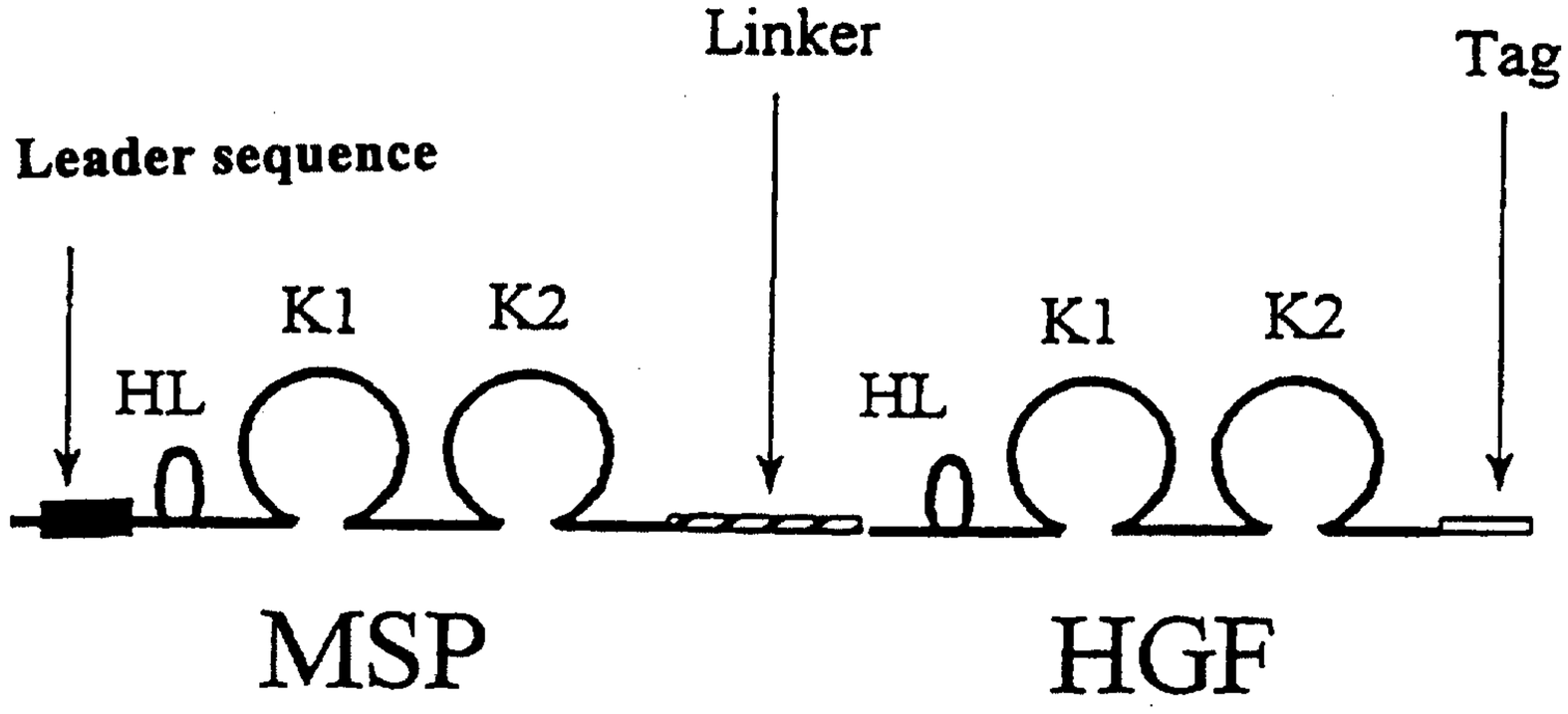


FIG 2b

```
1   GAATTCCACCATGGGGTGGCTCCCACTCCTGCTGCTTCTGACTCAATGCTTAGGGGTCCC   60
    ---------+========+---------+---------+---------+---------+---------+
    CTTAAGGTGGTACCCCACCGAGGGTGAGGACGACGAAGACTGAGTTACGAATCCCCAGGG

1                 M   G   W   L   P   L   L   L   L   L   T   Q   C   L   G   V   P   17

61  TGGGCAGCGCTCGCCATTGAATGACTTCCAAGTGCTCCGGGGCACAGAGCTACAGCACCT  120
    ---------+---------+---------+---------+---------+---------+---------+
    ACCCGTCGCGAGCGGTAACTTACTGAAGGTTCACGAGGCCCCGTGTCTCGATGTCGTGGA

18    G   Q   R   S   P   L   N   D   F   Q   V   L   R   G   T   E   L   Q   H   L   37

121 GCTACATGCGGTGGTGCCCGGGCCTTGGCAGGAGGATGTGGCAGATGCTGAAGAGTGTGC  180
    ---------+---------+---------+---------+---------+---------+---------+
    CGATGTACGCCACCACGGGCCCGGAACCGTCCTCCTACACCGTCTACGACTTCTCACACG

38    L   H   A   V   V   P   G   P   W   Q   E   D   V   A   D   A   E   E   C   A   57

181 TGGTCGCTGTGGGCCCTTAATGGACTGCCGGGCCTTCCACTACAACGTGAGCAGCCATGG  240
    ---------+---------+---------+---------+---------+---------+---------+
    ACCAGCGACACCCGGGAATTACCTGACGGCCCGGAAGGTGATGTTGCACTCGTCGGTACC

58    G   R   C   G   P   L   M   D   C   R   A   F   H   Y   N   V   S   S   H   G   77

241 TTGCCAACTGCTGCCATGGACTCAACACTCGCCCCACACGAGGCTGCGGCGTTCTGGGCG  300
    ---------+---------+---------+---------+---------+---------+---------+
    AACGGTTGACGACGGTACCTGAGTTGTGAGCGGGGTGTGCTCCGACGCCGCAAGACCCGC

78    C   Q   L   L   P   W   T   Q   H   S   P   H   T   R   L   R   R   S   G   R   97

301 CTGTGACCTCTTCCAGAAGAAAGACTACGTACGGACCTGCATCATGAACAATGGGGTTGG  360
    ---------+---------+---------+---------+---------+---------+---------+
    GACACTGGAGAAGGTCTTCTTTCTGATGCATGCCTGGACGTAGTACTTGTTACCCCAACC

98    C   D   L   F   Q   K   K   D   Y   V   R   T   C   I   M   N   N   G   V   G   117

361 GTACCGGGGCACCATGGCCACGACCGTGGGTGGCCTGCCCTGCCAGGCTTGGAGCCACAA  420
    ---------+---------+---------+---------+---------+---------+---------+
    CATGGCCCCGTGGTACCGGTGCTGGCACCCACCGGACGGGACGGTCCGAACCTCGGTGTT

118   Y   R   G   T   M   A   T   T   V   G   G   L   P   C   Q   A   W   S   H   K   137

421 GTTCCCGAATGATCACAAGTACACGCCCACTCTCCGGAATGGCCTGGAAGAGAACTTCTG  480
    ---------+---------+---------+---------+---------+---------+---------+
    CAAGGGCTTACTAGTGTTCATGTGCGGGTGAGAGGCCTTACCGGACCTTCTCTTGAAGAC

138   F   P   N   D   H   K   Y   T   P   T   L   R   N   G   L   E   E   N   F   C   157
```

(continued)

```
481 CCGTAACCCTGATGGCGACCCCGGAGGTCCTTGGTGCTACACAACAGACCCTGCTGTGCG  540
    ---------+---------+---------+---------+---------+---------+---------+
    GGCATTGGGACTACCGCTGGGGCCTCCAGGAACCACGATGTGTTGTCTGGGACGACACGC

158   R   N   P   D   G   D   P   G   G   P   W   C   Y   T   T   D   P   A   V   R   177
```

(continued)

```
                CTTCCAGAGCTGCGGCATCAAATCCTGCCGGGAGGCCGCGTGTGTCTGGTGCAATGGCGA
541             ----------+---------+---------+---------+---------+---------+  600
                GAAGGTCTCGACGCCGTAGTTTAGGACGGCCCTCCGGCGCACACAGACCACGTTACCGCT

178              F  Q  S  C  G  I  K  S  C  R  E  A  A  C  V  W  C  N  G  E   197

                GGAATACCGCGGCGCGGTAGACCGCACGGAGTCAGGGCGCGAGTGCCAGCGCTGGGATCT
601             ----------+---------+---------+---------+---------+---------+  660
                CCTTATGGCGCCGCGCCATCTGGCGTGCCTCAGTCCCGCGCTCACGGTCGCGACCCTAGA

198              E  Y  R  G  A  V  D  R  T  E  S  G  R  E  C  Q  R  W  D  L   217

                TCAGCACCCGCACCAGCACCCCTTCGAGCCGGGCAAGTTCCTCGACCAAGGTCTGGACGA
661             ----------+---------+---------+---------+---------+---------+  720
                AGTCGTGGGCGTGGTCGTGGGGAAGCTCGGCCCGTTCAAGGAGCTGGTTCCAGACCTGCT

218              Q  H  P  H  Q  H  P  F  E  P  G  K  F  L  D  Q  G  L  D  D   237

                CAACTATTGCCGGAATCCTGACGGCTCCGAGCGGCCATGGTGCTACACTACGGATCCGCA
721             ----------+---------+---------+---------+---------+---------+  780
                GTTGATAACGGCCTTAGGACTGCCGAGGCTCGCCGGTACCACGATGTGATGCCTAGGCGT

238              N  Y  C  R  N  P  D  G  S  E  R  P  W  C  Y  T  T  D  P  Q   257

                GATCGAGCGAGAGTTCTGTGACCTCCCCCGCTGCGGGTCCGAGGCACAGCCCCGCCTCGA
781             ----------+---------+---------+---------+---------+---------+  840
                CTAGCTCGCTCTCAAGACACTGGAGGGGGCGACGCCCAGGCTCCGTGTCGGGGCGGAGCT

258              I  E  R  E  F  C  D  L  P  R  C  G  S  E  A  Q  P  R  L  E   277

                GGGCGGTGGCGGTTCTGGTGGCGGTGGCTCCGGCGGTGGCGGTTCTCTAGAGGGACAAAG
841             ----------+---------+---------+---------+---------+---------+  900
                CCCGCCACCGCCAAGACCACCGCCACCGAGGCCGCCACCGCCAAGAGATCTCCCTGTTTC

278              G  G  G  G  S  G  G  G  G  S  G  G  G  G  S  L  E  G  Q  R   297

                GAAAAGAAGAAATACAATTCATGAATTCAAAAAATCAGCAAAGACTACCCTAATCAAAAT
901             ----------+---------+---------+---------+---------+---------+  960
                CTTTTCTTCTTTATGTTAAGTACTTAAGTTTTTTAGTCGTTTCTGATGGGATTAGTTTTA

298              K  R  R  N  T  I  H  E  F  K  K  S  A  K  T  T  L  I  K  I   317

                AGATCCAGCACTGAAGATAAAAACCAAAAAAGTGAATACTGCAGACCAATGTGCTAATAG
961             ----------+---------+---------+---------+---------+---------+ 1020
                TCTAGGTCGTGACTTCTATTTTTGGTTTTTTCACTTATGACGTCTGGTTACACGATTATC

318              D  P  A  L  K  I  K  T  K  K  V  N  T  A  D  Q  C  A  N  R   337

                ATGTACTAGGAATAAAGGACTTCCATTCACTTGCAAGGCTTTTGTTTTTGATAAAGCAAG
1021            ----------+---------+---------+---------+---------+---------+ 1080
                TACATGATCCTTATTTCCTGAAGGTAAGTGAACGTTCCGAAAACAAAAACTATTTCGTTC

338              C  T  R  N  K  G  L  P  F  T  C  K  A  F  V  F  D  K  A  R   357

                AAAACAATGCCTCTGGTTCCCCTTCAATAGCATGTCAAGTGGAGTGAAAAAAGAATTTGG
1081            ----------+---------+---------+---------+---------+---------+ 1140
                TTTTGTTACGGAGACCAAGGGGAAGTTATCGTACAGTTCACCTCACTTTTTTCTTAAACC

358              K  Q  C  L  W  F  P  F  N  S  M  S  S  G  V  K  K  E  F  G   377

                CCATGAATTTGACCTCTATGAAAACAAAGACTACATTAGAAACTGCATCATTGGTAAAGG
1141            ----------+---------+---------+---------+---------+---------+ 120C
                GGTACTTAAACTGGAGATACTTTTGTTTCTGATGTAATCTTTGACGTAGTAACCATTTCC

378              H  E  F  D  L  Y  E  N  K  D  Y  I  R  N  C  I  I  G  K  G   397
```

(continued)

(continued)

```
                ACGCAGCTACAAGGGAACAGTATCTATCACTAAGAGTGGCATCAAATGTCAGCCCTGGAG
1201            ---------+---------+---------+---------+---------+---------+  1260
                TGCGTCGATGTTCCCTTGTCATAGATAGTGATTCTCACCGTAGTTTACAGTCGGGACCTC

 398             R   S   Y   K   G   T   V   S   I   T   K   S   G   I   K   C   Q   P   W   S   417

                TTCCATGATACCACACGAACACAGCTATCGGGGTAAAGACCTACAGGAAAACTACTGTCG
1261            ---------+---------+---------+---------+---------+---------+  1320
                AAGGTACTATGGTGTGCTTGTGTCGATAGCCCCATTTCTGGATGTCCTTTTGATGACAGC

 418             S   M   I   P   H   E   H   S   Y   R   G   K   D   L   Q   E   N   Y   C   R   437

                AAATCCTCGAGGGGAAGAAGGGGGACCCTGGTGTTTCACAAGCAATCCAGAGGTACGCTA
1321            ---------+---------+---------+---------+---------+---------+  1380
                TTTAGGAGCTCCCCTTCTTCCCCCTGGGACCACAAAGTGTTCGTTAGGTCTCCATGCGAT

 438             N   P   R   G   E   E   G   G   P   W   C   F   T   S   N   P   E   V   R   Y   457

                CGAAGTCTGTGACATTCCTCAGTGTTCAGAAGTTGAATGCATGACCTGCAATGGGGAGAG
1381            ---------+---------+---------+---------+---------+---------+  1440
                GCTTCAGACACTGTAAGGAGTCACAAGTCTTCAACTTACGTACTGGACGTTACCCCTCTC

 458             E   V   C   D   I   P   Q   C   S   E   V   E   C   M   T   C   N   G   E   S   477

                TTATCGAGGTCTCATGGATCATACAGAATCAGGCAAGATTTGTCAGCGCTGGGATCATCA
1441            ---------+---------+---------+---------+---------+---------+  1500
                AATAGCTCCAGAGTACCTAGTATGTCTTAGTCCGTTCTAAACAGTCGCGACCCTAGTAGT

 478             Y   R   G   L   M   D   H   T   E   S   G   K   I   C   Q   R   W   D   H   Q   497

                GACACCACACCGGCACAAATTCTTGCCTGAAAGATATCCCGACAAGGGCTTTGATGATAA
1501            ---------+---------+---------+---------+---------+---------+  1560
                CTGTGGTGTGGCCGTGTTTAAGAACGGACTTTCTATAGGGCTGTTCCCGAAACTACTATT

 498             T   P   H   R   H   K   F   L   P   E   R   Y   P   D   K   G   F   D   D   N   517

                TTATTGCCGCAATCCCGATGGCCAGCCGAGGCCATGGTGCTATACTCTTGACCCTCACAC
1561            ---------+---------+---------+---------+---------+---------+  1620
                AATAACGGCGTTAGGGCTACCGGTCGGCTCCGGTACCACGATATGAGAACTGGGAGTGTG

 518             Y   C   R   N   P   D   G   Q   P   R   P   W   C   Y   T   L   D   P   H   T   537

                CCGCTGGGAGTACTGTGCAATTAAAACATGCGCTGACAAAGCTGACGACGACGACAAACA
1621            ---------+---------+---------+---------+---------+---------+  1680
                GGCGACCCTCATGACACGTTAATTTTGTACGCGACTGTTTCGACTGCTGCTGCTGTTTGT

 538             R   W   E   Y   C   A   I   K   T   C   A   D   K   A   D   D   D   D   K   H   557

                CCACCACCACCACCACCACTAGGGTCGAC
1681            ---------+---------+---------  1709
                GGTGGTGGTGGTGGTGGTGATCCCAGCTG

 558             H   H   H   H   H   H   *                 563
```

FIG 3a

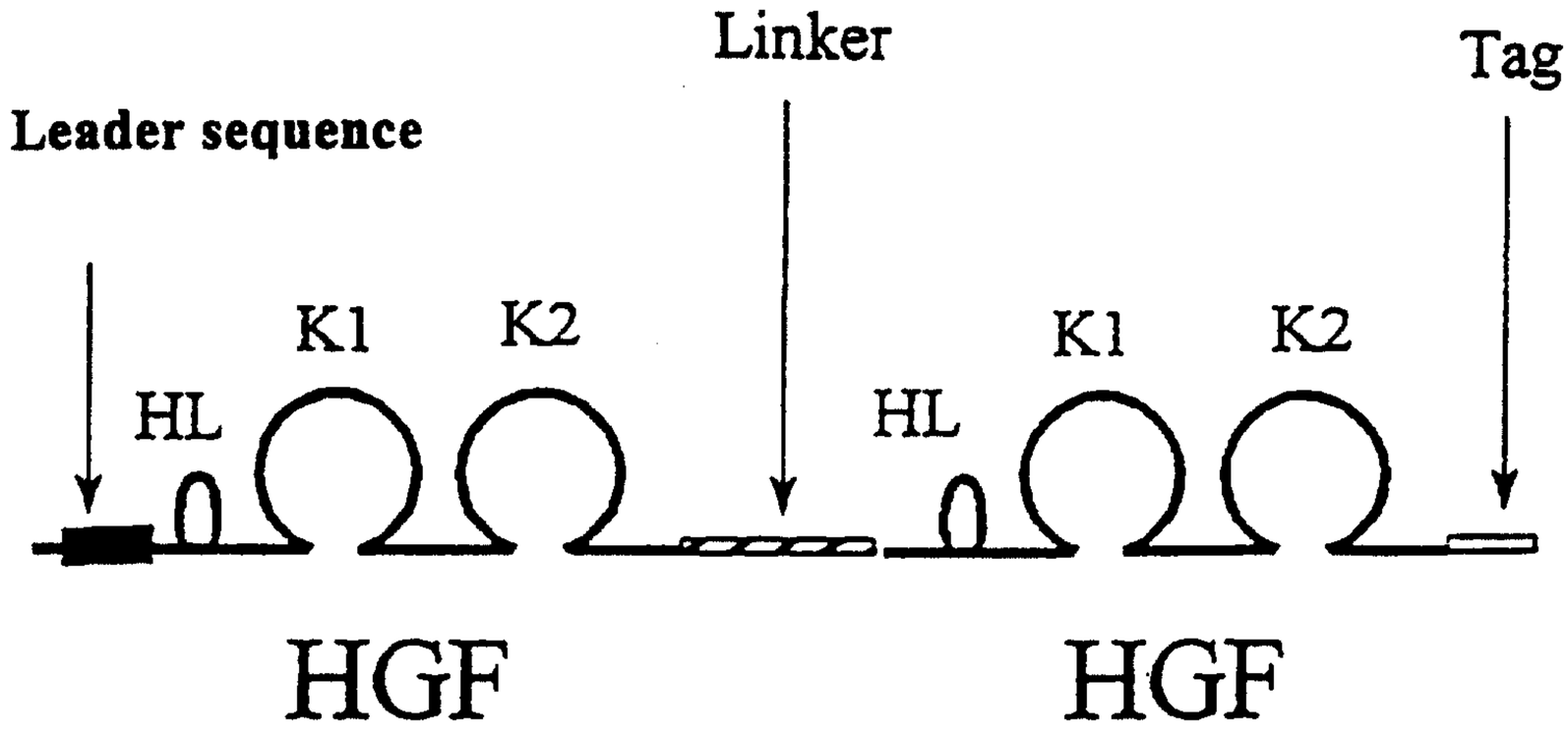


FIG 3b

```
                GGATCCGCCAGCCCGTCCAGCAGCACCATGTGGGTGACCAAACTCCTGCCAGCCCTGCTG
1               ---------+---------+---------+---------+---------+---------+  60
                CCTAGGCGGTCGGGCAGGTCGTCGTGGTACACCCACTGGTTTGAGGACGGTCGGGACGAC

1                                   M   W   V   T   K   L   L   P   A   L   L   11

                CTGCAGCATGTCCTCCTGCATCTCCTCCTGCTCCCCATCGCCATCCCCTATGCAGAGGGA
61              ---------+---------+---------+---------+---------+---------+  120
                GACGTCGTACAGGAGGACGTAGAGGAGGACGAGGGGTAGCGGTAGGGGATACGTCTCCCT

12  L   Q   H   V   L   L   H   L   L   L   L   P   I   A   I   P   Y   A   E   G   31

                CAAAGGAAAAGAAGAAATACAATTCATGAATTCAAAAAATCAGCAAAGACTACCCTAATC
121             ---------+---------+---------+---------+---------+---------+  180
                GTTTCCTTTTCTTCTTTATGTTAAGTACTTAAGTTTTTTAGTCGTTTCTGATGGGATTAG

32  Q   R   K   R   R   N   T   I   H   E   F   K   K   S   A   K   T   T   L   I   51

                AAAATAGATCCAGCACTGAAGATAAAAACCAAAAAAGTGAATACTGCAGACCAATGTGCT
181             ---------+---------+---------+---------+---------+---------+  240
                TTTTATCTAGGTCGTGACTTCTATTTTTGGTTTTTTCACTTATGACGTCTGGTTACACGA

52  K   I   D   P   A   L   K   I   K   T   K   K   V   N   T   A   D   Q   C   A   71

                AATAGATGTACTAGGAATAAAGGACTTCCATTCACTTGCAAGGCTTTTGTTTTTGATAAA
241             ---------+---------+---------+---------+---------+---------+  300
                TTATCTACATGATCCTTATTTCCTGAAGGTAAGTGAACGTTCCGAAAACAAAAACTATTT

72  N   R   C   T   R   N   K   G   L   P   F   T   C   K   A   F   V   F   D   K   91

                GCAAGAAAACAATGCCTCTGGTTCCCCTTCAATAGCATGTCAAGTGGAGTGAAAAAAGAA
301             ---------+---------+---------+---------+---------+---------+  360
                CGTTCTTTTGTTACGGAGACCAAGGGGAAGTTATCGTACAGTTCACCTCACTTTTTTCTT

92  A   R   K   Q   C   L   W   F   P   F   N   S   M   S   S   G   V   K   K   E   111

                TTTGGCCATGAATTTGACCTCTATGAAAACAAAGACTACATTAGAAACTGCATCATTGGT
361             ---------+---------+---------+---------+---------+---------+  420
                AAACCGGTACTTAAACTGGAGATACTTTTGTTTCTGATGTAATCTTTGACGTAGTAACCA

112 F   G   H   E   F   D   L   Y   E   N   K   D   Y   I   R   N   C   I   I   G   131

                AAAGGACGCAGCTACAAGGGAACAGTATCTATCACTAAGAGTGGCATCAAATGTCAGCCC
421             ---------+---------+---------+---------+---------+---------+  480
                TTTCCTGCGTCGATGTTCCCTTGTCATAGATAGTGATTCTCACCGTAGTTTACAGTCGGG

132 K   G   R   S   Y   K   G   T   V   S   I   T   K   S   G   I   K   C   Q   P   151

                TGGAGTTCCATGATACCACACGAACACAGCTATCGGGGTAAAGACCTACAGGAAAACTAC
481             ---------+---------+---------+---------+---------+---------+  540
                ACCTCAAGGTACTATGGTGTGCTTGTGTCGATAGCCCCATTTCTGGATGTCCTTTTGATG

152 W   S   S   M   I   P   H   E   H   S   Y   R   G   K   D   L   Q   E   N   Y   171
```

(continued)

(continued)

```
541  TGTCGAAATCCTCGAGGGGAAGAAGGGGGACCCTGGTGTTTCACAAGCAATCCAGAGGTA
     ACAGCTTTAGGAGCTCCCCTTCTTCCCCCTGGGACCACAAAGTGTTCGTTAGGTCTCCAT  600
172  C  R  N  P  R  G  E  E  G  G  P  W  C  F  T  S  N  P  E  V   191

601  CGCTACGAAGTCTGTGACATTCCTCAGTGTTCAGAAGTTGAATGCATGACCTGCAATGGG
     GCGATGCTTCAGACACTGTAAGGAGTCACAAGTCTTCAACTTACGTACTGGACGTTACCC  660
192  R  Y  E  V  C  D  I  P  Q  C  S  E  V  E  C  M  T  C  N  G   211

661  GAGAGTTATCGAGGTCTCATGGATCATACAGAATCAGGCAAGATTTGTCAGCGCTGGGAT
     CTCTCAATAGCTCCAGAGTACCTAGTATGTCTTAGTCCGTTCTAAACAGTCGCGACCCTA  720
212  E  S  Y  R  G  L  M  D  H  T  E  S  G  K  I  C  Q  R  W  D   231

721  CATCAGACACCACACCGGCACAAATTCTTGCCTGAAAGATATCCCGACAAGGGCTTTGAT
     GTAGTCTGTGGTGTGGCCGTGTTTAAGAACGGACTTTCTATAGGGCTGTTCCCGAAACTA  780
232  H  Q  T  P  H  R  H  K  F  L  P  E  R  Y  P  D  K  G  F  D   251

781  GATAATTATTGCCGCAATCCCGATGGCCAGCCGAGGCCATGGTGCTATACTCTTGACCCT
     CTATTAATAACGGCGTTAGGGCTACCGGTCGGCTCCGGTACCACGATATGAGAACTGGGA  840
252  D  N  Y  C  R  N  P  D  G  Q  P  R  P  W  C  Y  T  L  D  P   271

841  CACACCCGCTGGGAGTACTGTGCAATTAAAACATGCGCTGACAAAGCTTCGGGCGGTGGC
     GTGTGGGCGACCCTCATGACACGTTAATTTTGTACGCGACTGTTTCGAAGCCCGCCACCG  900
272  H  T  R  W  E  Y  C  A  I  K  T  C  A  D  K  A  S  G  G  G   291

901  GGTTCTGGTGGCGGTGGCTCCGGCGGTGGCGGTTCTCTAGAGGGACAAAGGAAAAGAAGA
     CCAAGACCACCGCCACCGAGGCCGCCACCGCCAAGAGATCTCCCTGTTTCCTTTTCTTCT  960
292  G  S  G  G  G  G  S  G  G  G  G  S  L  E  G  Q  R  K  R  R   311

961  AATACAATTCATGAATTCAAAAAATCAGCAAAGACTACCCTAATCAAAATAGATCCAGCA
     TTATGTTAAGTACTTAAGTTTTTTAGTCGTTTCTGATGGGATTAGTTTTATCTAGGTCGT  1020
312  N  T  I  H  E  F  K  K  S  A  K  T  T  L  I  K  I  D  P  A   331

1021 CTGAAGATAAAAACCAAAAAAGTGAATACTGCAGACCAATGTGCTAATAGATGTACTAGG
     GACTTCTATTTTTGGTTTTTTCACTTATGACGTCTGGTTACACGATTATCTACATGATCC  1080
332  L  K  I  K  T  K  K  V  N  T  A  D  Q  C  A  N  R  C  T  R   351

1081 AATAAAGGACTTCCATTCACTTGCAAGGCTTTTGTTTTTGATAAAGCAAGAAAACAATGC
     TTATTTCCTGAAGGTAAGTGAACGTTCCGAAAACAAAAACTATTTCGTTCTTTTGTTACG  1140
352  N  K  G  L  P  F  T  C  K  A  F  V  F  D  K  A  R  K  Q  C   371

1141 CTCTGGTTCCCCTTCAATAGCATGTCAAGTGGAGTGAAAAAAGAATTTGGCCATGAATTT
     GAGACCAAGGGGAAGTTATCGTACAGTTCACCTCACTTTTTTCTTAAACCGGTACTTAAA  1200
372  L  W  F  P  F  N  S  M  S  S  G  V  K  K  E  F  G  H  E  F   391
```

(continued)

(continued)

```
1201  GACCTCTATGAAAACAAAGACTACATTAGAAACTGCATCATTGGTAAAGGACGCAGCTAC
      ---------+---------+---------+---------+---------+---------+---------+
      CTGGAGATACTTTTGTTTCTGATGTAATCTTTGACGTAGTAACCATTTCCTGCGTCGATG  1260

 392  D  L  Y  E  N  K  D  Y  I  R  N  C  I  I  G  K  G  R  S  Y    411

1261  AAGGGAACAGTATCTATCACTAAGAGTGGCATCAAATGTCAGCCCTGGAGTTCCATGATA
      ---------+---------+---------+---------+---------+---------+---------+
      TTCCCTTGTCATAGATAGTGATTCTCACCGTAGTTTACAGTCGGGACCTCAAGGTACTAT  1320

 412  K  G  T  V  S  I  T  K  S  G  I  K  C  Q  P  W  S  S  M  I    431

1321  CCACACGAACACAGCTATCGGGGTAAAGACCTACAGGAAAACTACTGTCGAAATCCTCGA
      ---------+---------+---------+---------+---------+---------+---------+
      GGTGTGCTTGTGTCGATAGCCCCATTTCTGGATGTCCTTTTGATGACAGCTTTAGGAGCT  1380

 432  P  H  E  H  S  Y  R  G  K  D  L  Q  E  N  Y  C  R  N  P  R    451

1381  GGGGAAGAAGGGGGACCCTGGTGTTTCACAAGCAATCCAGAGGTACGCTACGAAGTCTGT
      ---------+---------+---------+---------+---------+---------+---------+
      CCCCTTCTTCCCCCTGGGACCACAAAGTGTTCGTTAGGTCTCCATGCGATGCTTCAGACA  1440

 452  G  E  E  G  G  P  W  C  F  T  S  N  P  E  V  R  Y  E  V  C    471

1441  GACATTCCTCAGTGTTCAGAAGTTGAATGCATGACCTGCAATGGGGAGAGTTATCGAGGT
      ---------+---------+---------+---------+---------+---------+---------+
      CTGTAAGGAGTCACAAGTCTTCAACTTACGTACTGGACGTTACCCCTCTCAATAGCTCCA  1500

 472  D  I  P  Q  C  S  E  V  E  C  M  T  C  N  G  E  S  Y  R  G -  491

1501  CTCATGGATCATACAGAATCAGGCAAGATTTGTCAGCGCTGGGATCATCAGACACCACAC
      ---------+---------+---------+---------+---------+---------+---------+
      GAGTACCTAGTATGTCTTAGTCCGTTCTAAACAGTCGCGACCCTAGTAGTCTGTGGTGTG  1560

 492  L  M  D  H  T  E  S  G  K  I  C  Q  R  W  D  H  Q  T  P  H    511

1561  CGGCACAAATTCTTGCCTGAAAGATATCCCGACAAGGGCTTTGATGATAATTATTGCCGC
      ---------+---------+---------+---------+---------+---------+---------+
      GCCGTGTTTAAGAACGGACTTTCTATAGGGCTGTTCCCGAAACTACTATTAATAACGGCG  1620

 512  R  H  K  F  L  P  E  R  Y  P  D  K  G  F  D  D  N  Y  C  R    531

1621  AATCCCGATGGCCAGCCGAGGCCATGGTGCTATACTCTTGACCCTCACACCCGCTGGGAG
      ---------+---------+---------+---------+---------+---------+---------+
      TTAGGGCTACCGGTCGGCTCCGGTACCACGATATGAGAACTGGGAGTGTGGGCGACCCTC  1680

 532  N  P  D  G  Q  P  R  P  W  C  Y  T  L  D  P  H  T  R  W  E    551

1681  TACTGTGCAATTAAAACATGCGCTGACAAAGCTGACGACGACGACAAACACCACCACCAC
      ---------+---------+---------+---------+---------+---------+---------+
      ATGACACGTTAATTTTGTACGCGACTGTTTCGACTGCTGCTGCTGTTTGTGGTGGTGGTG  1740

 552  Y  C  A  I  K  T  C  A  D  K  A  D  D  D  D  K  H  H  H  H    571

1741  CACCACCACTAGGGTCGAC
      ---------+---------
      GTGGTGGTGATCCCAGCTG  1759

 572  H  H  H  *                                                     574
```

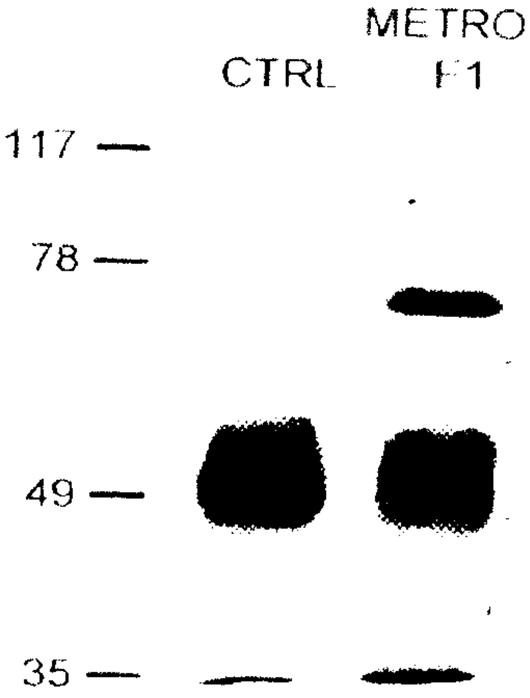


Fig 4

Fig 5A

Fig 5B

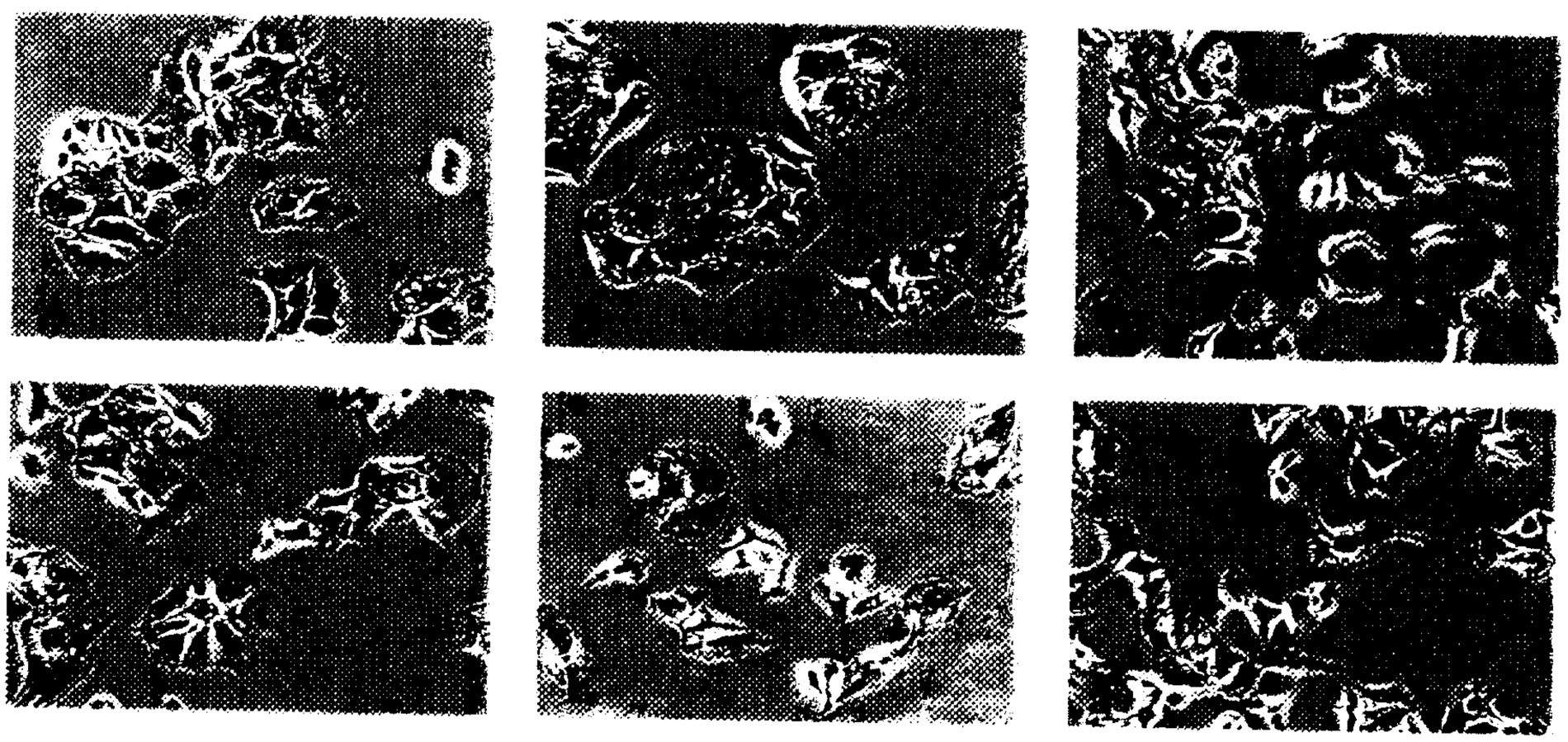


Fig 6

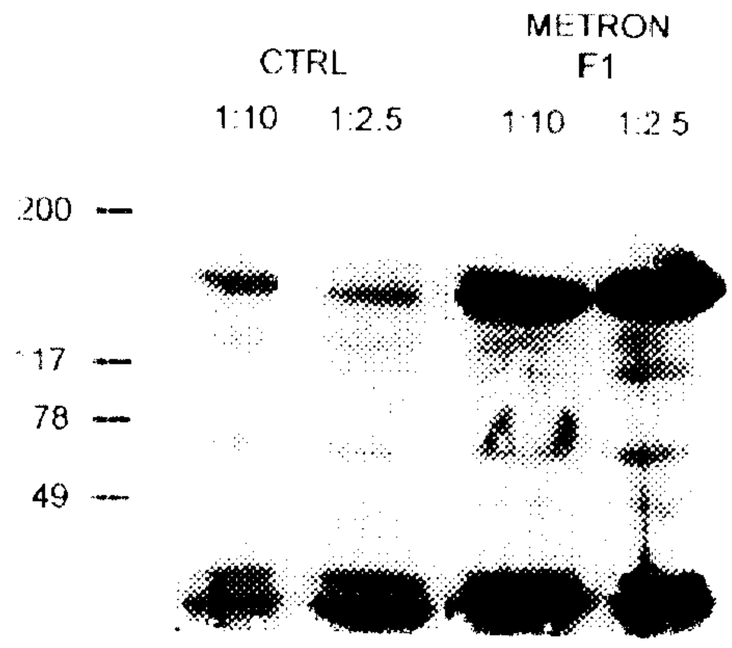
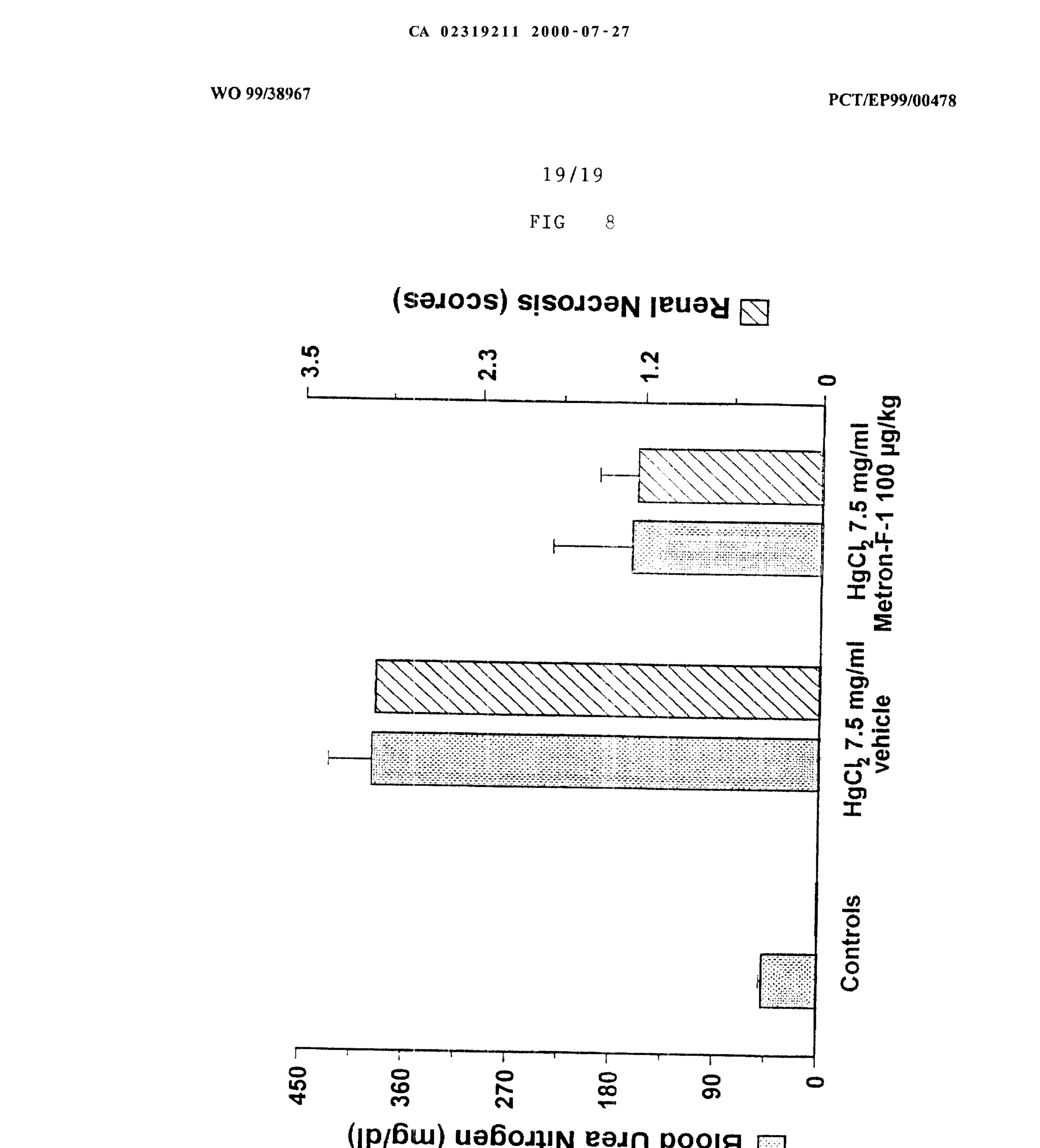


Fig 7

FIG 8


Renal Necrosis (scores)
3.5
2.3
1.2
0
Controls
HgCl2 7.5 mg/ml vehicle
HgCl2 7.5 mg/ml Metron-F-1 100 µg/kg
450
360
270
180
90
0
Blood Urea Nitrogen (mg/dl)