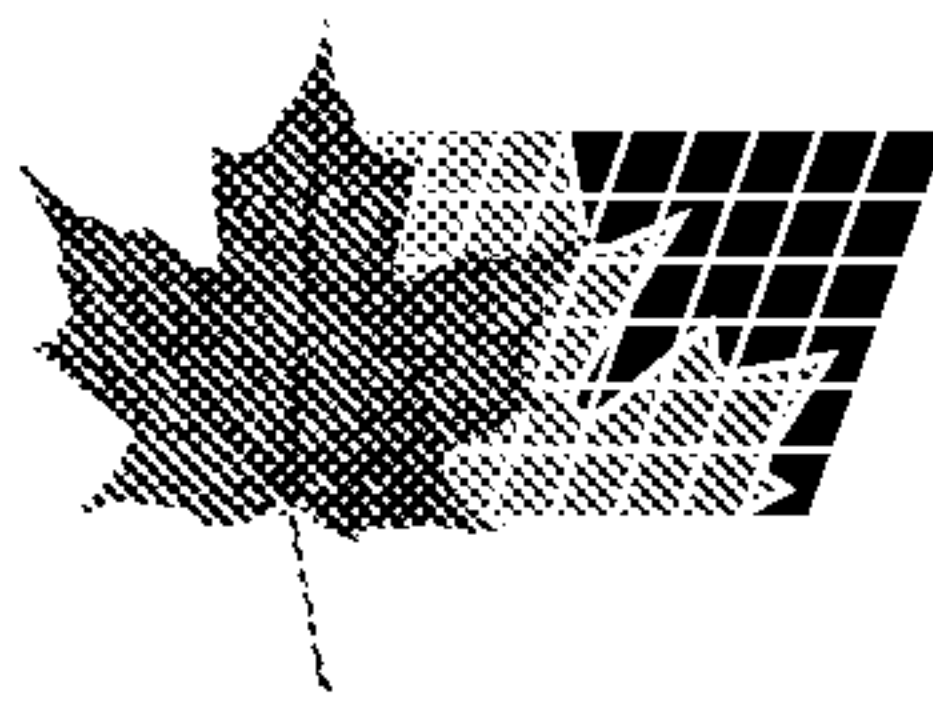




- (72) ZOU, HUA, US
(72) HENZEL, WILLIAM J., US
(72) WANG, XIAODONG, US
(71) GENENTECH, INC., US
(71) THE UNIVERSITY OF TEXAS SYSTEM BOARD OF REGENTS, US
(51) Int.Cl.⁷ C12N 15/12, G01N 33/68, C12N 15/62, C07K 14/47,
C07K 16/18, A61K 38/17, C12N 5/10, A01K 67/027, C07K 19/00
(30) 1997/06/05 (60/048,807) US
(30) 1997/08/07 (60/055,258) US
(54) **APAF-1, HOMOLOGUE HUMAIN DE CED-4, ACTIVATEUR DE
CASPASE-3**
(54) **APAF-1, THE CED-4 HUMAN HOMOLOG, AN ACTIVATOR OF
CASPASE-3**

(57) La présente invention concerne de nouveaux polypeptides appelés Apaf-1, capables de moduler une apoptose. La présente invention concerne également des compositions comprenant des chimères d'Apaf-1, un acide nucléique codant Apaf-1 et des anticorps contre d'Apaf-1.

(57) Novel polypeptides, designated Apaf-1, which are capable of modulating apoptosis are provided. Compositions including Apaf-1 chimeras, nucleic acid encoding Apaf-1, and antibodies to Apaf-1 are also provided.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, 15/62, C07K 14/47, 16/18, 19/00, C12N 5/10, A61K 38/17, G01N 33/68, A01K 67/027	A1	(11) International Publication Number: WO 98/55615 (43) International Publication Date: 10 December 1998 (10.12.98)
(21) International Application Number: PCT/US98/11773 (22) International Filing Date: 5 June 1998 (05.06.98) (30) Priority Data: 60/048,807 5 June 1997 (05.06.97) US 60/055,258 7 August 1997 (07.08.97) US (71) Applicants: GENENTECH, INC. [US/US]; 460 Point San Bruno Boulevard, South San Francisco, CA 94080-4990 (US). THE UNIVERSITY OF TEXAS BOARD OF REGENTS [US/US]; Southwestern Medical Center, Office of Legal Affairs and Technology Transfer, 5323 Harry Hines Boulevard, Dallas, TX 75235-9008 (US). (72) Inventors: ZOU, Hua; 2506A Rampast, Dallas, TX 75235 (US). HENZEL, William, J.; 3724 Southwood Avenue, San Mateo, CA 94403 (US). WANG, Xiaodong; 4063 Stone Hollow Way, Dallas, TX 75287 (US). (74) Agent: BRUESS, Steven, C.; Merchant, Gould, Smith, Edell, Welter & Schmidt, P.A., 3100 Norwest Center, 90 South Seventh Street, Minneapolis, MN 55402-4131 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: APAF-1, THE CED-4 HUMAN HOMOLOG, AN ACTIVATOR OF CASPASE-3 (57) Abstract Novel polypeptides, designated Apaf-1, which are capable of modulating apoptosis are provided. Compositions including Apaf-1 chimeras, nucleic acid encoding Apaf-1, and antibodies to Apaf-1 are also provided.		

APAF-1, THE CED-4 HUMAN HOMOLOG, AN ACTIVATOR OF CASPASE-3

Field of the Invention

The present invention relates generally to the identification, isolation,
5 and recombinant production of a novel protein, designated herein as "Apaf-1".

Background of the InventionApoptosis or "Programmed Cell Death"

Control of cell numbers in mammals is believed to be determined, in
10 part, by a balance between cell proliferation and cell death. One form of cell death,
sometimes referred to as necrotic cell death, is typically characterized as a
pathologic form of cell death resulting from some trauma or cellular injury. In
contrast, another "physiologic" form of cell death usually proceeds in an orderly or
controlled manner. This orderly or controlled form of cell death is often referred to
15 as "apoptosis" (see, *e.g.*, Barr, et al., *Bio/Technology*, 12:487-493 (1994); Steller, et
al., *Science*, 267:1445-1449 (1995)).

Apoptotic cell death naturally occurs in many physiological
processes, including embryonic development and clonal selection in the immune
system (Itoh, et al., *Cell*, 66:233-243 (1991)). Decreased levels of apoptotic cell
20 death have been associated with a variety of pathological conditions, including
cancer, lupus, and herpes virus infection (Thompson, *Science*, 267:1456-1462
(1995)). Increased levels of apoptotic cell death may be associated with a variety of
other pathological conditions, including AIDS, Alzheimer's disease, Parkinson's
disease, amyotrophic lateral sclerosis, multiple sclerosis, retinitis pigmentosa,
25 cerebella degeneration, aplastic anemia, myocardial infarction, stroke, reperfusion
injury, and toxin-induced liver disease (see, Thompson, *Supra*).

Apoptotic cell death is typically accompanied by one or more
characteristic morphological and biochemical changes in cells, such as condensation
of cytoplasm, loss of plasma membrane microvilli, segmentation of the nucleus,
30 degradation of chromosomal DNA or loss of mitochondrial function. A recognized
biochemical marker of apoptosis is the cleavage of chromatin into nucleosomal
fragments.

A variety of extrinsic and intrinsic signals are believed to trigger or induce such morphological and biochemical cellular changes (Raff, *Nature*, 356:397-400 (1992); Steller, *Supra*; Sachs, et al., *Blood*, 82:15 (1993)). For instance, they can be triggered by hormonal stimuli, such as glucocorticoid hormones for immature thymocytes, as well as withdrawal of certain growth factors (Watanabe-Fukunaga, et al., *Nature*, 356:314-317 (1992)). Also, some identified oncogenes such as *myc*, *rel*, and *ElA*, and tumor suppressers, like *p53*, have been reported to have a role in inducing apoptosis. Certain chemotherapy drugs and some forms of radiation have likewise been observed to have apoptosis-inducing activity (Thompson, *Supra*). Apoptosis is also triggered by the activation of a family of cysteine proteases having specificity for aspartic acid residues, including *Ced-3* of *C. elegans*, CCP32 (now caspase-3), Yama/Apopain of humans, and DCP-1 of *Drosophila*. These proteases are designated as caspases (Alnemri, et al., *Cell*, 87:171, (1996)).

15

The Apoptosis-Inducing Signaling Complex

As presently understood, the apoptosis program contains at least three important elements - activators, inhibitors, and effectors. In *C. elegans*, these elements are encoded respectively by three genes, *Ced-4*, *Ced-9* and *Ced-3* (Steller, *Science*, 267:1445 (1995); Chinnaiyan, et al., *Science*, 275:1122-1126 (1997)). Two genes, *Ced-3* and *Ced-4*, are required to initiate apoptosis (Yuan and Horvitz, *Development* 116:309-320, (1990)). *Ced-9*, which functions upstream of *Ced-3* and *Ced-4*, negatively regulates the apoptotic program by preventing activation of *Ced-3* and *Ced-4* (Hengartner, et al., *Cell* 76:665-676 (1994)).

25

The apoptotic program delineated in *C. elegans* is conserved in mammalian cells which contain homologues of *Ced-9* and *Ced-3*. One of these homologues, Bcl-2, can partially substitute for *Ced-9* in preventing apoptosis in *C. elegans* (Hengartner and Horvitz, 1994, *Cell* 76:665-676). The other homologues are cysteine proteases that are closely related to *Ced-3*, including caspase-3 (Yuan, et al., *Cell* 75:641-652 (1993); Xue, et al., *Genes & Dev.* 10:1073-1083 (1996)). *Ced-4* is the only remaining *C. elegans* general apoptosis gene of which the mammalian counterpart had not been found. This gene is believed to function downstream of

30

Ced-9 but upstream of *Ced-3* in the *C. elegans* apoptosis pathway (Shaham and Horvitz, *Genes & Dev.* 10:578-591, (1996), *Cell* 86:201-208 (1996)).

In mammalian cells, caspase-3 normally exists in the cytosolic fraction as a 32 kDa inactive precursor which is converted proteolytically to a 20 kDa and a 10 kDa active heterodimer when cells are signaled to die (Schlegel, et al., *Biol. Chem.* 271:1841-1844, (1996); Wang, et al., *EMBO J.* 15:1012-1020, (1996)). Bcl-2, located on the outer membrane of mitochondria, prevents the activation of caspase-3. It appears to do this by blocking the mitochondria from releasing cytochrome c, a necessary co-factor for caspase-3 activation (Liu, et al., *Cell* 86:147-157, (1996); Yang, et al., *Science* 275:1129-1132, (1997); Kluck, et al., *Science* 275:1132-1136, (1997)). Deletion of caspase-3 from the mouse genome through homologous recombination results in excessive accumulation of neuronal cells, due to a lack of apoptosis in the brain (Kuida, et al., *Nature* 384:368-372 (1996)). Addition of active caspase-3 to normal cytosol activates the apoptotic program (Enari, et al., *Nature* 380:723-726 (1996)). Thus, caspase-3 is both necessary and sufficient to trigger apoptosis.

Identified substrates for caspase-3 include poly (ADP-ribose) polymerase (PARP), sterol-regulatory element binding proteins (SREBPs), the U1-associated 70 kDa protein, N4-GD1, huntingtin and DNA dependent protein kinase (Casicola-Rosen, et al., *J. Exp. Med.* 183:1957-1964 (1996); Na, et al., *J. Biol. Chem.* 271:11209-11213 (1996); Goldberg, et al., *Nat. Genet.* 13(4):442-449 (1996); Wang, et al. *EMBO J.* 15:1012-1020 (1996); Nicholson, et al., *Nature* 376:37-43, (1995)).

Applicants recently established an *in vitro* apoptosis system to study apoptosis using cytosolic fractions from normally growing HeLa cells. Using this system, two protein factors involved in mammalian apoptosis were identified: cytochrome c (Liu, et al., 1996, *Supra*) and DNA fragmentation factor (DFF), a novel heterodimer of 45 kDa and 40 kDa subunits. DFF functions downstream of caspase-3 to trigger fragmentation of genomic DNA into nucleosomal segments, a hallmark of apoptosis (Wyllie, *Nature* 284:555-556, (1980); Liu, et al., 1997 *Supra*).

Summary of the Invention

Applicants have identified a DNA sequence (SEQ ID NO:1) that encodes a novel protein (SEQ ID NO: 2), designated in the present application as "Apaf-1." It is believed that Apaf-1 is a mammalian homologue of *Ced-4*. The native sequence of human Apaf-1 is a novel 130 kDa polypeptide containing a *Ced-3* homologous domain at the N-terminus, followed by a *Ced-4* homologous domain and multiple WD-40 repeats at the C-terminus. It has been found that Apaf-1 can form a complex with cytochrome-c, and activates caspase-3 in the apoptotic pathway.

10 An alternatively spliced transcript, Apaf-1L has also been discovered. Its novel nucleic acid sequence (SEQ ID NO:15) and encoded protein (SEQ ID NO:16) have similar activities and, for purposes of the invention, included within the invention as an optional Apaf-1 polypeptide.

The invention provides isolated Apaf-1 protein. In particular, the invention provides an isolated native sequence Apaf-1 protein, which includes an amino acid sequence comprising residues 1 to 1194 of Figures 5A-5O and Figure 6 (SEQ ID NO:2). In other embodiments, the isolated Apaf-1 protein comprises at least about 80% amino acid sequence identity with native sequence Apaf-1 protein comprising residues 1 to 1194 of Figures 5A- 5O and Figure 6 (SEQ ID NO:2).

20 In another embodiment, the invention provides an isolated nucleic acid molecule encoding Apaf-1 protein (SEQ ID NO:1). The nucleic acid molecule is an RNA or DNA encoding an Apaf-1 protein or a particular domain of Apaf-1, such as a cytochrome c binding domain or an ATP binding domain, or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. In one embodiment, the nucleic acid sequence is selected from:

(a) the coding region of the nucleic acid sequence of Figures 5A - 5O (SEQ ID NO:1) that codes for amino acid residues 1 to 1194 (i.e., nucleotides 578 through 4159), inclusive;

30 (b) the enlarged cDNA sequence of Figures 5A - 5O (SEQ ID NO:1) encoding amino acids 1 to 1194, and including further nucleic acids positioned 5' of

the translation initiation site and/or 3' of the stop codon sequences (i.e.; nucleotides 1 to 7042);

(c) the coding region of the nucleic acid sequence of Figures 16A-16H (SEQ ID NO:15) that codes for amino acid residues 1 to 1205 (i.e., nucleotides 578 through 4192), inclusive;

(d) the enlarged cDNA sequence of Figures 16A - 16H (SEQ ID NO:15) encoding amino acids 1 to 1205, and including further nucleic acids positioned 5' of the translation initiation site and/or 3' of the stop codon sequences (i.e.; nucleotides 1 to 7075); and

(e) a sequence corresponding to the sequence of (a), (b), (c), or (d) and is within the scope of degeneracy of the genetic code.

In a further embodiment, the invention provides a vector comprising the nucleic acid molecule encoding the Apaf-1 protein or particular domain of Apaf-1. A host cell comprising the vector or the nucleic acid molecule is also provided. A method of producing Apaf-1 is further provided.

In a further embodiment, the invention provides vectors and gene constructs comprising a nucleic acid sequence encoding Apaf-1 or a portion thereof.

In another embodiment, the invention provides an antibody which specifically binds to Apaf-1. The antibody may be an agonistic, antagonistic or neutralizing antibody.

A further embodiment of the invention provides articles of manufacture and kits that include Apaf-1 protein, nucleic acid sequences encoding Apaf-1, or anti-Apaf-1 antibodies.

The novel gene, protein, and antibodies of the invention are particularly useful for screening and identifying cells signaled for apoptosis. In addition, the gene and protein of the invention are particularly useful for screening of candidate compounds to identify agents useful in modulating apoptosis.

Brief Description of the Drawings

Figure 1 shows a schematic outline of the purification procedure leading to the purification of Apaf-1.

Figure 2 shows a combination of agents which triggers apoptosis, as assayed by proteolytic cleavage of caspase-3 precursor into its activated form.

Figure 3 shows Apaf-1 activity of cytosolic fractions of Hela cells supernatant (S100) purified through a Mono Q column.

5 Figure 4 shows Apaf-1 protein of the column fractions described for Figure 3 in a silverstained PAGE gel (8%), having a molecular weight of approximately 130 kDa.

Figures 5A - 5O shows the DNA sequence and deduced amino acid sequence of human Apaf-1.

10 Figure 6 shows the protein sequence of Apaf-1. Previously sequenced 14 tryptic and Lys-C peptides (Table-1) are underlined. The putative WD-repeats are boxed.

Figure 7 is a schematic representation of the Apaf-1 structure shown by a horizontal bar. Numbers corresponding to the amino acid residues of Figures 5A - 5O and 6 are shown. The *Ced-3* homologous region, the *Ced-4* homologous region, and 13 WD repeats are indicated.

Figure 8 shows those amino acid sequences of Apaf-1 that share homology with the N-terminal pro-domain of *Ced-3* (Yuan, et al., *Supra* (1993)). Identical amino acids are shown in bold. Conserved amino acids are boxed.

20 Figure 9 shows those amino acid sequence of Apaf-1 that share homology with amino acids of *Ced-4* (Yuan and Horvitz, *Supra* (1992)).

Figure 10 is a representation of a Northern blot indicating the tissue distribution and relative expression of Apaf-1 mRNA.

25 Figure 11 is a representation of an immunoblot probed with anti-Apaf-1 antibody showing *in vitro* translation of Apaf-1.

Figure 12 is a Western blot probed with anti-Apaf-1 antiserum showing immunoprecipitation of Apaf-1.

30 Figure 13 is a Western blot probed with anti-cytochrome c antibody showing immunoprecipitation of cytochrome c from the Apaf-1/cytochrome c incubation mixture with anti-Apaf-1 antibody.

Figure 14 is a Western blot probed with anti-Apaf-1 antibody and with anti-caspase-3 antibody and demonstrating expression of recombinant Apaf-1 and caspase-3 in mammalian cells.

Figure 15 is a Western blot probed with anti-caspase-3 antibody and demonstrating use of recombinantly expressed Apaf-1 to induce cleavage of caspase 3.

Figures 16A-16H show the nucleic acid sequence encoding Apaf-1L. The additional 33 nucleic acids as compared with Apaf-1 are underlined.

Figure 17 shows the amino acid sequence of Apaf-1L. The additional 11 amino acids are underlined.

Figure 18 is a 2% agarose gel stained with ethidium bromide showing two RT-PCR products generated (Apaf-1 and Apaf-1L) from amplifying the first strand cDNA from HeLa mRNA using two primers designed from both the 5' and 3' sides of the additional 33 nucleotides of Apaf-1L, as described in Example 9.

Figure 19A is a Western Blot probed with anti-Apaf-1 antibody and with anti-caspase-3 antibody, demonstrating expression of recombinant Apaf-1L and caspase-3 in 293 cells. Figure 19B is a Western Blot probed with anti-caspase-3 antibody and demonstrating use of recombinantly expressed Apaf-1L to induce cleavage of caspase-3.

Detailed Description of the Preferred Embodiments

I. Definitions

The terms "Apaf-1 protein" and "Apaf-1" when used herein encompass native sequence Apaf-1, alternative splice variants such as Apaf-1L (described more fully in the Examples below), and other Apaf-1 variants (which are further defined herein). These terms encompass Apaf-1 from a variety of mammals, including humans. The Apaf-1 may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence Apaf-1" comprises a polypeptide having the same amino acid sequence as an Apaf-1 derived from nature. Thus, a native sequence Apaf-1 can have the amino acid sequence of naturally-occurring Apaf-1 from any

mammal. Such native sequence Apaf-1 can be isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence Apaf-1" specifically encompasses naturally-occurring truncated or secreted forms of the Apaf-1, naturally-occurring variant forms (*e.g.*, alternatively spliced forms), and naturally-occurring allelic variants of the Apaf-1.

"Apaf-1 variant" means a biologically active Apaf-1 as defined below having at least about 80% amino acid sequence identity with the Apaf-1 having the deduced amino acid sequence shown in Figures 5A - 5O (SEQ ID NO:2) or in Figure 17 (SEQ ID NO:16). Such Apaf-1 variants include, for instance, Apaf-1 proteins wherein one or more amino acid residues are added or deleted at the N- or C-terminus of the sequence of Figures 5A - 5O (SEQ ID NO:2) or in Figure 17 (SEQ ID NO:16). Ordinarily, an Apaf-1 variant will have at least about 80% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity, and even more preferably at least about 95% amino acid sequence identity with the amino acid sequence of Figures 5A - 5O (SEQ ID NO:2) or in Figure 17 (SEQ ID NO:16).

"Percent (%) amino acid sequence identity" with respect to the Apaf-1 sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the Apaf-1 sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGNTM or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

The term "epitope tagged" when used herein refers to a chimeric polypeptide comprising Apaf-1, or a domain sequence thereof, fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody can be made, yet is short enough such that it does not interfere

with activity of the Apaf-1. The tag polypeptide preferably also is fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 to about 50 amino acid residues (preferably, between about 10 to about 20 residues).

"Isolated", when used to describe the protein disclosed herein, means a polypeptide or protein that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the Apaf-1 natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" Apaf-1 nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the Apaf-1 nucleic acid. An isolated Apaf-1 nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated Apaf-1 nucleic acid molecules therefore are distinguished from the Apaf-1 nucleic acid molecule as it exists in natural cells. However, an isolated Apaf-1 nucleic acid molecule includes Apaf-1 nucleic acid molecules contained in cells that ordinarily express Apaf-1 where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a

promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a
5 presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally,
10 "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adapters or linkers are used in accordance with conventional practice.

15 The term "antibody" is used in the broadest sense and specifically covers single anti-Apaf-1 monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-Apaf-1 antibody compositions with polypeptidic specificity.

The term "monoclonal antibody" as used herein refers to an antibody
20 obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which
25 typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen.

The monoclonal antibodies herein include hybrid and recombinant antibodies produced by splicing a variable (including hypervariable) domain of an
30 anti-Apaf-1 antibody with a constant domain (*e.g.* "humanized" antibodies), or a light chain with a heavy chain, or a chain from one species with a chain from another species, or fusions with heterologous proteins, regardless of species of origin or

immunoglobulin class or subclass designation, as well as antibody fragments (*e.g.*, Fab, F(ab')₂, and Fv), so long as they exhibit the desired biological activity. See, *e.g.* U.S. Pat. No. 4,816,567 and Mage et al., in Monoclonal Antibody Production Techniques and Applications, pp.79-97 (Marcel Dekker, Inc.: New York, 1987).

5 Thus, the modifier "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by
10 Kohler and Milstein, *Nature*, 256:495 (1975), or may be made by recombinant DNA methods such as described in U.S. Pat. No. 4,816,567. The "monoclonal antibodies" may also be isolated from phage libraries generated using the techniques described in McCafferty, et al., *Nature*, 348:552-554 (1990), for example.

 "Humanized" forms of non-human (*e.g.* murine) antibodies are
15 specific chimeric immunoglobulins, immunoglobulin chains, or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are
20 replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat, or rabbit having the desired specificity, affinity, and capacity. In some instances, Fv framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, the humanized antibody may comprise residues which are found neither in the recipient antibody
25 nor in the imported CDR or framework sequences. These modifications are made to further refine and optimize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those
30 of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region or domain (Fc), typically that of a human immunoglobulin.

"Biologically active" and "desired biological activity" for the purposes herein mean having the ability to modulate apoptosis (in an agonistic or stimulating manner) in at least one type of mammalian cell *in vivo* or *ex vivo* or in the cell-free assay system described herein. In particular, the biological activity of Apaf-1 is the activation of caspase-3 to trigger apoptotic events.

The terms "apoptosis" and "apoptotic activity" are used in a broad sense and refer to the orderly or controlled form of cell death in mammals that is typically accompanied by one or more characteristic cell changes, including condensation of cytoplasm, loss of plasma membrane microvilli, segmentation of the nucleus, degradation of chromosomal DNA or loss of mitochondrial function. This activity can be determined and measured, for instance, by cell viability assays, FACS analysis or DNA electrophoresis, all of which are known in the art. In particular, apoptosis can be measured in the cell free system described in the examples below and in Liu, et al., *Supra*.

The terms "treating," "treatment," and "therapy" as used herein refer to curative therapy, prophylactic therapy, and preventative therapy.

The term "mammal" as used herein refers to any mammal classified as a mammal, including humans, cows, horses, dogs and cats. In a preferred embodiment of the invention, the mammal is a human.

II. Compositions and Methods of the Invention

The present invention provides a newly identified and isolated Apaf-1 protein. In particular, Applicants have identified and isolated a human Apaf-1 protein. The properties and characteristics of this Apaf-1 protein are described in further detail in the Examples below. Based upon the properties and characteristics of the Apaf-1 protein disclosed herein, it is Applicants' present belief that Apaf-1 is a mammalian homologue of *Ced-4*.

A. Apaf-1 is a Mammalian Homologue of *Ced-4*

Several lines of evidence suggest that Apaf-1 is the mammalian homologue of *Ced-4*. First, Apaf-1 shows significant sequence homology with *Ced-4* over more than 300 amino acids; second, several important regions of *Ced-4*,

including conserved nucleotide binding regions, and an isoleucine at position 258, are conserved in Apaf-1 (Figure 8); third, the biochemical function of Apaf-1, which is to mediate the activation of caspase-3, is consistent with the function of *Ced-4* as delineated by genetic studies. The genetic studies indicate that *Ced-4* functions downstream of *Ced-9*, but upstream of *Ced-3*, since the killing of the *C. elegans* ALM neurons by overexpression of *Ced-4* is greatly reduced in *Ced-3* mutant, whereas killing of these neurons by overexpression of *Ced-3* is unaffected by mutations in *Ced-4* (Hengartner, et al., *Supra*; Shaham and Horvitz, *Supra*). Similarly, *bcl-2* functions upstream of Apaf-1 by controlling the release of cytochrome c, a co-factor for Apaf-1 activity (Kluck, et al., *Supra*; Yang, et al., *Supra*).

B. Mechanism for Triggering Caspase-3 Activation

Apaf-1 itself does not appear to be a caspase. The conserved active site pentapeptide QACR (or Q/G)G that is present in all identified caspases is not present in Apaf-1. Most of the WD repeat-containing proteins are regulatory in function rather than enzymatic (Neer, et al., 1994, *Nature* 371:297-300). The identification of the regions that are conserved among proteins that utilize ATP is consistent with the requirement of dATP in the caspase-3 activation reaction. In this case, dATP is preferred over ATP.

Apaf-1 appears to be a key factor in the initiation of apoptosis. Apaf-1 binds cytochrome c and dATP, both important factors in triggering apoptosis. Together, Apaf-1, cytochrome c, dATP and Apaf-3 form a unique model system for apoptosis, in a cell-free system or within cells and tissues.

A description follows as to how Apaf-1, as well as Apaf-1 chimeric molecules and anti-Apaf-1 antibodies, may be prepared.

C. Preparation of Apaf-1

The description below relates primarily to production of Apaf-1 by culturing cells transformed or transfected with a vector containing Apaf-1 nucleic acid. It is of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare Apaf-1.

1. Isolation of DNA Encoding Apaf-1

The DNA encoding Apaf-1 may be obtained from any cDNA library prepared from tissue believed to possess the Apaf-1 mRNA and to express it at a detectable level. Accordingly, human Apaf-1 DNA can be conveniently obtained from a cDNA library prepared from human tissues, such as the libraries of human HeLa cell cDNA described in Example 2, and others, particularly DNA libraries produced from tissues known for high apoptotic activity such as spleen, fetal brain, and the like. Such DNA libraries are commercially available, e.g. from CloneTech and Stratagene. The Apaf-1-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to the Apaf-1 or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding Apaf-1 is to use PCR methodology (Sambrook, et al., *Supra*; Dieffenbach, et al., PCR Primer: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)).

A preferred method of screening employs selected oligonucleotide sequences to screen cDNA libraries from various human tissues. Example 2 below describes techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ³²P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook, et al., *Supra*.

Nucleic acid having all the protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook, et al., *Supra*, to detect

precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

Apaf-1 variants can be prepared by introducing appropriate nucleotide changes into the Apaf-1 DNA, or by synthesis of the desired Apaf-1 polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the Apaf-1, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native full-length sequence Apaf-1 or in various domains of the Apaf-1 described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Pat. No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the Apaf-1 that results in a change in the amino acid sequence of the Apaf-1 as compared with the native sequence Apaf-1. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the Apaf-1 molecule. The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis (Carter, et al., *Nucl. Acids Res.*, 13:4331 (1986); Zoller, et al., *Nucl. Acids Res.*, 10:6487 (1987)), cassette mutagenesis (Wells, et al., *Gene*, 34:315 (1985)), restriction selection mutagenesis (Wells, et al., *Philos. Trans. R. Soc. London SerA*, 317:415 (1986)) or other known techniques can be performed on the cloned DNA to produce the Apaf-1 variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence which are involved in the interaction with a particular ligand or receptor. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is the preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant. Alanine is also preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions (Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia,

J. Mol. Biol., 150:1 (1976)). If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

Once selected Apaf-1 variants are produced, they can be screened, for example, by their activity in the activation of caspase-3, for example, in the cell free assay described herein and as described in Example 3 below.

2. Insertion of Nucleic Acid into A Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding Apaf-1 may be inserted into a replicable vector for further cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector components generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence, each of which is described below.

(i) Signal Sequence Component

The Apaf-1 may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the Apaf-1 DNA that is inserted into the vector. The heterologous signal sequence selected preferably is one that is recognized and processed (*i.e.*, cleaved by a signal peptidase) by the host cell. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* -factor leaders, the latter described in U.S. Pat. No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression the native Apaf-1 presequence that normally directs insertion of Apaf-1 in the cell membrane of human cells *in vivo* is satisfactory, although other mammalian signal sequences may be used to direct secretion of the protein, such as

signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders, for example, the herpes simplex glycoprotein D signal.

The DNA for such precursor region is preferably ligated in reading frame to DNA encoding Apaf-1.

5 (ii) Origin of Replication Component

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Generally, in cloning vectors this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or
10 autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2 μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component
15 is not needed for mammalian expression vectors (the SV40 origin may typically be used because it contains the early promoter).

Most expression vectors are "shuttle" vectors, *i.e.*, they are capable of replication in at least one class of organisms but can be transfected into another organism for expression. For example, a vector is cloned in *E. coli* and then the
20 same vector is transfected into yeast or mammalian cells for expression even though it is not capable of replicating independently of the host cell chromosome.

DNA may also be amplified by insertion into the host genome. This is readily accomplished using *Bacillus* species as hosts, for example, by including in the vector a DNA sequence that is complementary to a sequence found in *Bacillus*
25 genomic DNA. Transfection of *Bacillus* with this vector results in homologous recombination with the genome and insertion of Apaf-1 DNA. However, the recovery of genomic DNA encoding Apaf-1 is more complex than that of an exogenously replicated vector because restriction enzyme digestion is required to excise the Apaf-1 DNA.

30 (iii) Selection Gene Component

Expression and cloning vectors typically contain a selection gene, also termed a selectable marker. This gene encodes a protein necessary for the

survival or growth of transformed host cells grown in a selective culture medium.

Host cells not transformed with the vector containing the selection gene will not

survive in the culture medium. Typical selection genes encode proteins that (a)

confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin,

5 methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

One example of a selection scheme utilizes a drug to arrest growth of

a host cell. Those cells that are successfully transformed with a heterologous gene

10 produce a protein conferring drug resistance and thus survive the selection regimen.

Examples of such dominant selection use the drugs neomycin (Southern, et al., *J.*

Molec. Appl. Genet., 1:327 (1982)), mycophenolic acid (Mulligan, et al., *Science*,

209:1422 (1980)) or hygromycin (Sugden, et al., *Mol. Cell. Biol.*, 5:410-413 (1985)).

The three examples given above employ bacterial genes under eukaryotic control to

15 convey resistance to the appropriate drug G418 or neomycin (geneticin), xgpt (mycophenolic acid), or hygromycin, respectively.

Another example of suitable selectable markers for mammalian cells

are those that enable the identification of cells competent to take up the Apaf-1

nucleic acid, such as DHFR or thymidine kinase. The mammalian cell transformants

20 are placed under selection pressure that only the transformants are uniquely adapted

to survive by virtue of having taken up the marker. Selection pressure is imposed by

culturing the transformants under conditions in which the concentration of selection

agent in the medium is successively changed, thereby leading to amplification of

both the selection gene and the DNA that encodes Apaf-1. Amplification is the

25 process by which genes in greater demand for the production of a protein critical for

growth are reiterated in tandem within the chromosomes of successive generations

of recombinant cells. Increased quantities of Apaf-1 are synthesized from the

amplified DNA. Other examples of amplifiable genes include metallothionein-I and

-II, adenosine deaminase, and ornithine decarboxylase.

30 Cells transformed with the DHFR selection gene may first be

identified by culturing all of the transformants in a culture medium that contains

methotrexate (Mtx), a competitive antagonist of DHFR. An appropriate host cell

when wild-type DHFR is employed is the Chinese hamster ovary (CHO) cell line deficient in DHFR activity, prepared and propagated as described by Urlaub, et al., *Proc. Natl. Acad. Sci. USA*, 77:4216 (1980). The transformed cells are then exposed to increased levels of methotrexate. This leads to the synthesis of multiple copies of the DHFR gene, and, concomitantly, multiple copies of other DNA comprising the expression vectors, such as the DNA encoding Apaf-1. This amplification technique can be used with any otherwise suitable host, *e.g.*, ATCC No. CCL61 CHO-K1, notwithstanding the presence of endogenous DHFR if, for example, a mutant DHFR gene that is highly resistant to Mtx is employed (EP 117,060).

Alternatively, host cells (particularly wild-type hosts that contain endogenous DHFR) transformed or co-transformed with DNA sequences encoding Apaf-1, wild-type DHFR protein, and another selectable marker such as aminoglycoside 3'-phosphotransferase (APH) can be selected by cell growth in medium containing a selection agent for the selectable marker such as an aminoglycosidic antibiotic, *e.g.*, kanamycin, neomycin, or G418. See U.S. Patent No. 4,965,199.

A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid YRp7 (Stinchcomb, et al., *Nature*, 282:39 (1979); Kingsman, et al., *Gene*, 7:141 (1979); Tschemper, et al., *Gene*, 10:157 (1980)). The *trp1* gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 (Jones, *Genetics*, 85:12 (1977)). The presence of the *trp1* lesion in the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Similarly, *Leu2*-deficient yeast strains (ATCC 20,622 or 38,626) are complemented by known plasmids bearing the *Leu2* gene.

In addition, vectors derived from the 1.6 μ m circular plasmid pKD1 can be used for transformation of *Kluyveromyces* yeasts (Bianchi, et al., *Curr. Genet.*, 12:185 (1987)). More recently, an expression system for large-scale production of recombinant calf chymosin was reported for *K. lactis* (Van den Berg, *Bio/Technology*, 8:135 (1990)). Stable multi-copy expression vectors for secretion of mature recombinant human serum albumin by industrial strains of *Kluyveromyces* have also been disclosed (Fleer, et al., *Bio/Technology*, 9:968-975 (1991)).

(iv) Promoter Component

Expression and cloning vectors usually contain a promoter that is recognized by the host organism and is operably linked to the Apaf-1 nucleic acid sequence. Promoters are untranslated sequences located upstream (5') to the start codon of a structural gene (generally within about 100 to 1000 bp) that control the transcription and translation of particular nucleic acid sequence, such as the Apaf-1 nucleic acid sequence, to which they are operably linked. Such promoters typically fall into two classes, inducible and constitutive. Inducible promoters are promoters that initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, *e.g.*, the presence or absence of a nutrient or a change in temperature. At this time a large number of promoters recognized by a variety of potential host cells are well known. These promoters are operably linked to Apaf-1 encoding DNA by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native Apaf-1 promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of the Apaf-1 DNA.

Promoters suitable for use with prokaryotic hosts include the β -lactamase and lactose promoter systems (Chang, et al., *Nature*, 275:615 (1978); Goeddel, et al., *Nature*, 281:544 (1979)), alkaline phosphatase, a tryptophan (trp) promoter system (Goeddel, *Nucleic Acids Res.*, 8:4057 (1980); EP 36,776), and hybrid promoters such as the tac promoter (deBoer, et al., *Proc. Natl. Acad. Sci. USA*, 80:21-25 (1983)). However, other known bacterial promoters are suitable. Their nucleotide sequences have been published, thereby enabling a skilled worker operably to ligate them to DNA encoding Apaf-1 (Siebenlist, et al., *Cell*, 20:269 (1980)) using linkers or adapters to supply any required restriction sites. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding Apaf-1.

Promoter sequences are known for eukaryotes. Virtually all eukaryotic genes have an AT-rich region located approximately 25 to 30 bases upstream from the site where transcription is initiated. Another sequence found 70 to 80 bases upstream from the start of transcription of many genes is a CXCAAT region where X may be any nucleotide. At the 3' end of most eukaryotic genes is an

AATAAA sequence that may be the signal for addition of the poly A tail to the 3' end of the coding sequence. All of these sequences are suitably inserted into eukaryotic expression vectors.

Examples of suitable promoting sequences for use with yeast hosts
5 include the promoters for 3-phosphoglycerate kinase (Hitzeman, et al., *J. Biol. Chem.*, 255:2073 (1980)) or other glycolytic enzymes (Hess, et al., *J. Adv. Enzyme Reg.*, 7:149 (1968); Holland, *Biochemistry*, 17:4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase,
10 pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase,
15 degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657. Yeast enhancers also are advantageously used with yeast promoters.

20 Apaf-1 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and most preferably Simian Virus 40
25 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, from heat-shock promoters, and from the promoter normally associated with the Apaf-1 sequence, provided such promoters are compatible with the host cell systems.

The early and late promoters of the SV40 virus are conveniently
30 obtained as an SV40 restriction fragment that also contains the SV40 viral origin of replication (Fiers, et al., *Nature*, 273:113 (1978); Mulligan and Berg, *Science*, 209:1422-1427 (1980); Pavlakis, et al., *Proc. Natl. Acad. Sci. USA*, 78:7398-7402

(1981)). The immediate early promoter of the human cytomegalovirus is conveniently obtained as a HindIII E restriction fragment (Greenaway, et al., *Gene*, 18:355-360 (1982)). A system for expressing DNA in mammalian hosts using the bovine papilloma virus as a vector is disclosed in U.S. Patent No. 4,419,446. A
5 modification of this system is described in U.S. Patent No. 4,601,978 (See also Gray, et al., *Nature*, 295:503-508 (1982) on expressing cDNA encoding immune interferon in monkey cells; Reyes, et al., *Nature*, 297:598-601 (1982) on expression of human γ -interferon cDNA in mouse cells under the control of a thymidine kinase promoter from herpes simplex virus; Canaani and Berg, *Proc. Natl. Acad. Sci. USA*
10 79:5166-5170 (1982) on expression of the human interferon β gene in cultured mouse and rabbit cells; and Gorman, et al., *Proc. Natl. Acad. Sci. USA*, 79:6777-6781 (1982) on expression of bacterial CAT sequences in CV-1 monkey kidney cells, chicken embryo fibroblasts, Chinese hamster ovary cells, HeLa cells, and mouse NIH-3T3 cells using the Rous sarcoma virus long terminal repeat as a
15 promoter).

(v) Enhancer Element Component

Transcription of a DNA encoding the Apaf-1 of this invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp,
20 that act on a promoter to increase its transcription. Enhancers are relatively orientation and position independent, having been found 5' (Laimins, et al., *Proc. Natl. Acad. Sci. USA*, 78:993 (1981) and 3' (Lusky, et al., *Mol. Cell Bio.*, 3:1108 (1983) to the transcription unit, within an intron (Banerji, et al., *Cell*, 33:729 (1983)), as well as within the coding sequence itself (Osborne, et al., *Mol. Cell Bio.*,
25 4:1293 (1984)). Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α -fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin,
30 and adenovirus enhancers. See also Yaniv, *Nature*, 297:17-18 (1982) on enhancing elements for activation of eukaryotic promoters. The enhancer may be spliced into

the vector at a position 5' or 3' to the Apaf-1 coding sequence, but is preferably located at a site 5' from the promoter.

(vi) Transcription Termination Component

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding Apaf-1.

(vii) Construction and Analysis of Vectors

Construction of suitable vectors containing one or more of the above-listed components employs standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids required.

For analysis to confirm correct sequences in plasmids constructed, the ligation mixtures can be used to transform *E. coli* K12 strain 294 (ATCC 31,446) and successful transformants selected by ampicillin or tetracycline resistance where appropriate. Plasmids from the transformants are prepared, analyzed by restriction endonuclease digestion, and/or sequenced by the method of Messing, et al., *Nucleic Acids Res.*, 9:309 (1981) or by the method of Maxam, et al., *Methods in Enzymology*, 65:499 (1980).

(viii) Transient Expression Vectors

Expression vectors that provide for the transient expression in mammalian cells of DNA encoding Apaf-1 may be employed. In general, transient expression involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector and, in turn, synthesizes high levels of a desired polypeptide encoded by the expression vector (Sambrook et al., *supra*). Transient expression systems, comprising a suitable expression vector and a host cell, allow for the convenient positive identification of polypeptides encoded by cloned DNAs, as well as for the rapid screening of such polypeptides for desired biological or

physiological properties. Thus, transient expression systems are particularly useful in the invention for purposes of identifying Apaf-1 variants.

(ix) Suitable Exemplary Vertebrate Cell Vectors

Other methods, vectors, and host cells suitable for adaptation to the synthesis of Apaf-1 in recombinant vertebrate cell culture are described in Gething et al., *Nature*, 293:620-625 (1981); Mantei et al., *Nature*, 281:40-46 (1979); EP 117,060; and EP 117,058.

3. Selection and Transformation of Host Cells

Suitable host cells for cloning or expressing the DNA in the vectors herein are the prokaryote, yeast, or higher eukaryote cells described above. Suitable prokaryotes for this purpose include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *Escherichia*, e.g., *E. coli*, *Enterobacter*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, e.g., *Salmonella typhimurium*, *Serratia*, e.g., *Serratia marcescans*, and *Shigella*, as well as *Bacilli* such as *B. subtilis* and *B. licheniformis* (e.g., *B. licheniformis* 41P disclosed in DD 266,710 published 12 April 1989), *Pseudomonas* such as *P. aeruginosa*, and *Streptomyces*. Preferably, the host cell should secrete minimal amounts of proteolytic enzymes.

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for Apaf-1-encoding vectors. *Saccharomyces cerevisiae*, or common baker's yeast, is the most commonly used among lower eukaryotic host microorganisms. However, a number of other genera, species, and strains are commonly available and useful herein.

Suitable host cells for the expression of glycosylated Apaf-1 are derived from multicellular organisms. Such host cells are capable of complex processing and glycosylation activities. In principle, any higher eukaryotic cell culture is workable, whether from vertebrate or invertebrate culture. Examples of invertebrate cells include plant and insect cells. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts such as *Spodoptera frugiperda* (caterpillar), *Aedes aegypti* (mosquito), *Aedes albopictus* (mosquito), *Drosophila melanogaster* (fruitfly), and *Bombyx mori* have been identified (See, e.g., Luckow, et al., *Bio/Technology*, 6:47-55 (1988); Miller, et al.,

in *Genetic Engineering*, Setlow, et al., eds., Vol. 8 (Plenum Publishing, 1986), pp. 277-279; and Maeda, et al., *Nature*, 315:592-594 (1985)). A variety of viral strains for transfection are publicly available, e.g., the L-1 variant of *Autographa californica* NPV and the Bm-5 strain of *Bombyx mori* NPV.

5 Plant cell cultures of cotton, corn, potato, soybean, petunia, tomato, and tobacco can be utilized as hosts. Typically, plant cells are transfected by incubation with certain strains of the bacterium *Agrobacterium tumefaciens*. During incubation of the plant cell culture with *A. tumefaciens*, the DNA encoding the Apaf-1 can be transferred to the plant cell host such that it is transfected, and will, under
10 appropriate conditions, express the Apaf-1-encoding DNA. In addition, regulatory and signal sequences compatible with plant cells are available, such as the nopaline synthase promoter and polyadenylation signal sequences (Depicker, et al., *J. Mol. Appl. Gen.*, 1:561 (1982)). In addition, DNA segments isolated from the upstream region of the T-DNA 780 gene are capable of activating or increasing transcription
15 levels of plant-expressible genes in recombinant DNA-containing plant tissue (EP 321,196 published 21 June 1989).

 Propagation of vertebrate cells in culture (tissue culture) is also well known in the art (See, e.g., *Tissue Culture*, Academic Press, Kruse and Patterson, editors (1973)). Examples of useful mammalian host cell lines are monkey kidney
20 CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham, et al., *J. Gen Virol.*, 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, *Biol. Reprod.*, 23:243-
25 251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562,
30 ATCC CCL51); TRI cells (Mather, et al., *Annals N.Y. Acad. Sci.*, 383:44-68 (1982)); MRC 5 cells; and FS4 cells.

Host cells are transfected and preferably transformed with the above-described expression or cloning vectors for Apaf-1 production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

5 Transfection refers to the taking up of an expression vector by a host cell whether or not any coding sequences are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, CaPO_4 and electroporation. Successful transfection is generally recognized when any indication of the operation of this vector occurs within the host cell.

10 Transformation means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration. Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook, et al., *Supra*, or electroporation is generally used
15 for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw, et al., *Gene*, 23:315 (1983) and WO 89/05859 published 29 June 1989. In addition, plants may be transfected using ultrasound treatment as described in WO 91/00358 published 10 January 1991.

20 For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, *Virology*, 52:456-457 (1978) is preferred. General aspects of mammalian cell host system transformations have been described in U.S. Pat. No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen, et al., *J. Bact.*, 130:946 (1977)
25 and Hsiao, et al., *Proc. Natl. Acad. Sci. (USA)*, 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, e.g., polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown, et al., *Methods in Enzymology*, 185:527-537 (1990)
30 and Mansour, et al., *Nature*, 336:348-352 (1988).

4. Culturing the Host Cells

Prokaryotic cells used to produce Apaf-1 may be cultured in suitable media as described generally in Sambrook, et al., *Supra*.

5 The mammalian host cells used to produce Apaf-1 may be cultured in a variety of media. Examples of commercially available media include Ham's F10 (Sigma), Minimal Essential Medium ("MEM", Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium ("DMEM", Sigma). Any such media may be supplemented as necessary with hormones and/or other growth factors (such as insulin, transferrin, or epidermal growth factor), salts (such as sodium chloride, calcium, magnesium, and phosphate), buffers (such as HEPES), nucleosides (such as adenosine and thymidine), antibiotics (such as GentamycinTM drug), trace elements (defined as inorganic compounds usually present at final concentrations in the micromolar range), and glucose or an equivalent energy source. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. The culture conditions, such as temperature, pH, and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

In general, principles, protocols, and practical techniques for maximizing the productivity of mammalian cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991).

The host cells referred to in this disclosure encompass cells in culture as well as cells that are within a host animal.

5. Detecting Gene Amplification/Expression

25 Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA (Thomas, *Proc. Natl. Acad. Sci. USA*, 77:5201-5205 (1980)), dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Various labels may be employed, most commonly radioisotopes, and particularly ³²P. However, other techniques may also be employed, such as using biotin-modified nucleotides for introduction into a polynucleotide. The biotin then serves as the site

for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionucleotides, fluorescers or enzymes. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. With immunohistochemical staining techniques, a cell sample is prepared, typically by dehydration and fixation, followed by reaction with labeled antibodies specific for the gene product coupled, where the labels are usually visually detectable, such as enzymatic labels, fluorescent labels, or luminescent labels.

Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence Apaf-1 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to Apaf-1 DNA and encoding a specific antibody epitope.

6. Purification of Apaf-1 Polypeptide

Forms of Apaf-1 may be recovered from culture medium or from host cell lysates. If the Apaf-1 is membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or its extracellular domain may be released by enzymatic cleavage.

When Apaf-1 is produced in a recombinant cell other than one of human origin, the Apaf-1 is free of proteins or polypeptides of human origin. However, it may be desired to purify Apaf-1 from recombinant cell proteins or polypeptides to obtain preparations that are substantially homogeneous as to Apaf-1. As a first step, the culture medium or lysate may be centrifuged to remove particulate cell debris. Apaf-1 thereafter is purified from contaminant soluble proteins and polypeptides, with the following procedures being exemplary of

suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; and protein A
5 Sepharose columns to remove contaminants such as IgG.

Apaf-1 variants in which residues have been deleted, inserted, or substituted can be recovered in the same fashion as native sequence Apaf-1, taking account of changes in properties occasioned by the variation. For example, preparation of an Apaf-1 fusion with another protein or polypeptide, *e.g.*, a bacterial
10 or viral antigen, immunoglobulin sequence, or receptor sequence, may facilitate purification; an immunoaffinity column containing antibody to the sequence can be used to adsorb the fusion polypeptide. Other types of affinity matrices also can be used.

A protease inhibitor such as phenyl methyl sulfonyl fluoride (PMSF)
15 also may be useful to inhibit proteolytic degradation during purification, and antibiotics may be included to prevent the growth of adventitious contaminants. One skilled in the art will appreciate that purification methods suitable for native sequence Apaf-1 may require modification to account for changes in the character of Apaf-1 or its variants upon expression in recombinant cell culture.

20 7. Covalent Modifications of Apaf-1 Polypeptides

Covalent modifications of Apaf-1 are included within the scope of this invention. One type of covalent modification of the Apaf-1 is introduced into the molecule by reacting targeted amino acid residues of the Apaf-1 with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C-
25 terminal residues of the Apaf-1.

Derivatization with bifunctional agents is useful for cross-linking Apaf-1 to a water-insoluble support matrix or surface for use in the method for purifying anti-Apaf-1 antibodies, and vice-versa. Derivatization with one or more bifunctional agents will also be useful for cross-linking Apaf-1 molecules to
30 generate Apaf-1 dimers. Such dimers may increase binding avidity and extend half-life of the molecule *in vivo*. Commonly used cross-linking agents include, *e.g.*, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for

example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-((p-azidophenyl)dithio)propioimide yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include deamidation of glutamyl and asparagyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α -amino groups of lysine, arginine, and histidine side chains (T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group. The modified forms of the residues fall within the scope of the present invention.

Another type of covalent modification of the Apaf-1 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence Apaf-1, and/or adding one or more glycosylation sites that are not present in the native sequence Apaf-1.

Glycosylation of polypeptides is typically either N-linked or O-linked. N-linked refers to the attachment of the carbohydrate moiety to the side chain of an asparagine residue. The tripeptide sequences asparagine-X-serine and asparagine-X-threonine, where X is any amino acid except proline, are the recognition sequences for enzymatic attachment of the carbohydrate moiety to the asparagine side chain. Thus, the presence of either of these tripeptide sequences in a polypeptide creates a potential glycosylation site. O-linked glycosylation refers to the attachment of one of the sugars N-acetylgalactosamine, galactose, or xylose to a

hydroxylamino acid, most commonly serine or threonine, although 5-hydroxyproline or 5-hydroxylysine may also be used.

Addition of glycosylation sites to the Apaf-1 polypeptide may be accomplished by altering the amino acid sequence such that it contains one or more of the above-described tripeptide sequences (for N-linked glycosylation sites). The alteration may also be made by the addition of, or substitution by, one or more serine or threonine residues to the native sequence Apaf-1 (for O-linked glycosylation sites). The Apaf-1 amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the Apaf-1 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids. The DNA mutation(s) may be made using methods described above and in U.S. Pat. No. 5,364,934, *Supra*.

Another means of increasing the number of carbohydrate moieties on the Apaf-1 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Depending on the coupling mode used, the sugar(s) may be attached to (a) arginine and histidine, (b) free carboxyl groups, (c) free sulfhydryl groups such as those of cysteine, (d) free hydroxyl groups such as those of serine, threonine, or hydroxyproline, (e) aromatic residues such as those of phenylalanine, tyrosine, or tryptophan, or (f) the amide group of glutamine. These methods are described in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the Apaf-1 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. For instance, chemical deglycosylation by exposing the polypeptide to the compound trifluoromethanesulfonic acid, or an equivalent compound can result in the cleavage of most or all sugars except the linking sugar (N-acetylglucosamine or N-acetylgalactosamine), while leaving the polypeptide intact. Chemical deglycosylation is described by Hakimuddin, et al., *Arch. Biochem. Biophys.*, 259:52 (1987) and by Edge, et al., *Anal. Biochem.*, 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of

endo- and exo-glycosidases as described by Thotakura, et al., *Meth. Enzymol.*, 138:350 (1987).

Glycosylation at potential glycosylation sites may be prevented by the use of the compound tunicamycin as described by Duskin, et al., *J. Biol. Chem.*, 257:3105 (1982). Tunicamycin blocks the formation of protein-N-glycoside linkages.

Another type of covalent modification of Apaf-1 comprises linking the Apaf-1 polypeptide to one of a variety of nonproteinaceous polymers, e.g., polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

8. Apaf-1 Chimeras

The present invention also provides chimeric molecules comprising Apaf-1 fused to another, heterologous polypeptide or amino acid sequence.

In one embodiment, the chimeric molecule comprises a fusion of the Apaf-1 with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the Apaf-1. The presence of such epitope-tagged forms of the Apaf-1 can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the Apaf-1 to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include the flu HA tag polypeptide and its antibody 12CA5 (Field, et al., *Mol. Cell. Biol.*, 8:2159-2165 (1988)); the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto (Evan, et al., *Molecular and Cellular Biology*, 5:3610-3616 (1985)); and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody (Paborsky, et al., *Protein Engineering*, 3(6):547-553 (1990)). Other tag polypeptides include the Flag-peptide (Hopp, et al., *BioTechnology*, 6:1204-1210 (1988)); the KT3 epitope peptide (Martin, et al., *Science*, 255:192-194 (1992)); an α -tubulin epitope peptide (Skinner, et al., *J. Biol. Chem.*, 266:15163-15166 (1991)); and the T7 gene 10 protein peptide tag (Lutz-

Freyermuth, et al., *Proc. Natl. Acad. Sci. USA*, 87:6393-6397 (1990)). Once the tag polypeptide has been selected, an antibody thereto can be generated using the techniques disclosed herein.

Generally, epitope-tagged Apaf-1 may be constructed and produced according to the methods described above. Apaf-1-tag polypeptide fusions are preferably constructed by fusing the cDNA sequence encoding the Apaf-1 portion in-frame to the tag polypeptide DNA sequence and expressing the resultant DNA fusion construct in appropriate host cells. Ordinarily, when preparing the Apaf-1-tag polypeptide chimeras of the present invention, nucleic acid encoding the Apaf-1 will be fused at its 3' end to nucleic acid encoding the N-terminus of the tag polypeptide, however 5' fusions are also possible. For example, a polyhistidine sequence of about 5 to about 10 histidine residues may be fused at the N-terminus or the C-terminus and used as a purification handle in affinity chromatography.

Epitope-tagged Apaf-1 can be purified by affinity chromatography using the anti-tag antibody. The matrix to which the affinity antibody is attached may include, for instance, agarose, controlled pore glass or poly(styrenedivinyl)benzene. The epitope-tagged Apaf-1 can then be eluted from the affinity column using techniques known in the art.

In another embodiment, the chimeric molecule comprises an Apaf-1 polypeptide fused to an immunoglobulin sequence. The chimeric molecule may also comprise a particular domain sequence of Apaf-1, such as the *Ced-4*-like domain sequence of native Apaf-1 fused to an immunoglobulin sequence. This includes chimeras in monomeric, homo- or heteromultimeric, and particularly homo- or heterodimeric, or -tetrameric forms; optionally, the chimeras may be in dimeric forms or homodimeric heavy chain forms. Generally, these assembled immunoglobulins will have known unit structures as represented by the following diagrams.

X or A
 \ _____ C_H or C_L

X or A
 \ _____ Y _____ C_H or C_L

34

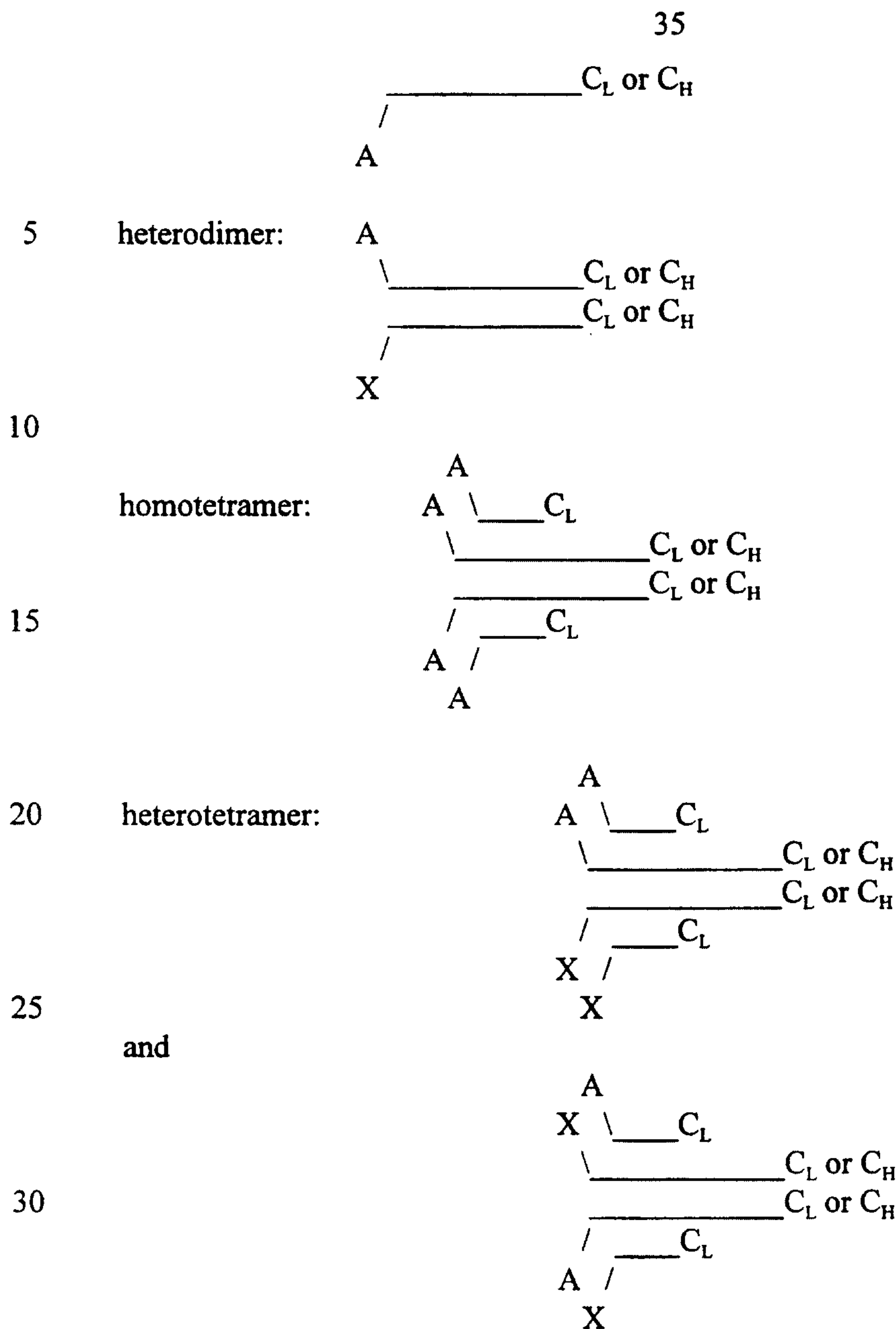


A basic four chain structural unit is the form in which IgG, IgD, and IgE exist. A four chain unit is repeated in the higher molecular weight immunoglobulins; IgM generally exists as a pentamer of basic four-chain units held together by disulfide bonds. IgA globulin, and occasionally IgG globulin, may also exist in a multimeric form in serum. In the case of multimers, each four chain unit may be the same or different.

The following diagrams depict some exemplary monomer, homo- and heterodimer and homo- and heteromultimer structures. These diagrams are merely illustrative, and the chains of the multimers are believed to be disulfide bonded in the same fashion as native immunoglobulins.

monomer: A _____ C_L or C_H

homodimer: A _____ C_L or C_H



In the foregoing diagrams, "A" means an Apaf-1 sequence or an Apaf-1 sequence fused to a heterologous sequence; X is an additional agent, which may be the same as A or different, a portion of an immunoglobulin superfamily member such as a variable region or a variable region-like domain, including a native or chimeric immunoglobulin variable region, a toxin such a pseudomonas exotoxin or ricin, or a sequence functionally binding to another protein, such as other cytokines (i.e., IL-1, interferon- α) or cell surface molecules (i.e., NGFR, CD40, OX40, Fas antigen, T2 proteins of Shope and myxoma poxviruses), or a polypeptide therapeutic agent not otherwise normally associated with a constant

domain; Y is a linker or another receptor sequence; and V_L , V_H , C_L and C_H represent light or heavy chain variable or constant domains of an immunoglobulin. Structures comprising at least one CRD of an Apaf-1 sequence as "A" and another cell-surface protein having a repetitive pattern of CRDs (such as TNFR) as "X" are specifically included.

It will be understood that the above diagrams are merely exemplary of the possible structures of the chimeras of the present invention, and do not encompass all possibilities. For example, there might desirably be several different "A"s, "X"s, or "Y"s in any of these constructs. Also, the heavy or light chain constant domains may be originated from the same or different immunoglobulins. All possible permutations of the illustrated and similar structures are all within the scope of the invention herein.

In general, the chimeric molecules can be constructed in a fashion similar to chimeric antibodies in which a variable domain from an antibody of one species is substituted for the variable domain of another species. See, for example, EP 0 125 023; EP 173,494; Munro, *Nature*, 312:597 (13 December 1984); Neuberger, et al., *Nature*, 312:604-608 (1984); Sharon, et al., *Nature*, 309:364-367 (1984); Morrison et al., *Proc. Nat'l. Acad. Sci. USA*, 81:6851-6855 (1984); Morrison, et al., *Science*, 229:1202-1207 (1985); Boulianne, et al., *Nature*, 312:643-646 (1984); Capon, et al., *Nature*, 337:525-531 (1989); Traunecker, et al., *Nature*, 339:68-70 (1989).

Alternatively, the chimeric molecules may be constructed as follows. The DNA including a region encoding the desired sequence, such as an Apaf-1 and/or TNFR sequence, is cleaved by a restriction enzyme at or proximal to the 3' end of the DNA encoding the immunoglobulin-like domain(s) and at a point at or near the DNA encoding the N-terminal end of the Apaf-1 or TNFR polypeptide (where use of a different leader is contemplated) or at or proximal to the N-terminal coding region for TNFR (where the native signal is employed). This DNA fragment then is readily inserted proximal to DNA encoding an immunoglobulin light or heavy chain constant region and, if necessary, the resulting construct tailored by deletional mutagenesis. Preferably, the Ig is a human immunoglobulin when the chimeric molecule is intended for *in vivo* therapy for humans. DNA encoding

immunoglobulin light or heavy chain constant regions is known or readily available from cDNA libraries or is synthesized. See for example, Adams, et al., *Biochemistry*, 19:2711-2719 (1980); Gough, et al., *Biochemistry*, 19:2702-2710 (1980); Dolby, et al., *Proc. Natl. Acad. Sci. USA*, 77:6027-6031 (1980); Rice, et al., *Proc. Natl. Acad. Sci.*, 79:7862-7865 (1982); Falkner, et al., *Nature*, 298:286-288 (1982); and Morrison, et al., *Ann. Rev. Immunol.*, 2:239-256 (1984).

Further details of how to prepare such fusions are found in publications concerning the preparation of immunoadhesins. Immunoadhesins in general, and CD4-Ig fusion molecules specifically are disclosed in WO 89/02922, published 6 April 1989. Molecules comprising the extracellular portion of CD4, the receptor for human immunodeficiency virus (HIV), linked to IgG heavy chain constant region are known in the art and have been found to have a markedly longer half-life and lower clearance than the soluble extracellular portion of CD4 (Capon, et al., *Supra*; Byrn, et al., *Nature*, 344:667 (1990)). The construction of specific chimeric TNFR-IgG molecules is also described in Ashkenazi, et al. *Proc. Natl. Acad. Sci.*, 88:10535-10539 (1991); Lesslauer, et al. (*J. Cell. Biochem. Supplement* 15F, 1991, p. 115 (P 432); and Peppel and Beutler, *J. Cell. Biochem. Supplement* 15F, 1991, p. 118 (P 439)).

20 B. Therapeutic and Non-therapeutic Uses for Apaf-1

Apaf-1, as disclosed in the present specification, can be employed to induce apoptosis. This induction of apoptosis may be in a cell free system or in mammalian cells, as described in the examples below.

Apaf-1 of the invention also has utility in non-therapeutic applications. Nucleic acid sequences encoding the Apaf-1 may be used as a diagnostic for tissue-specific typing. For example, procedures like *in situ* hybridization, Northern and Southern blotting, and PCR analysis may be used to determine whether DNA and/or RNA encoding Apaf-1 is present in the cell type(s) being evaluated, and what amount of the protein or RNA is produced in the cell. This embodiment provides a method for identifying cells that are destined to die. The presence or amount of Apaf-1 protein or RNA is assessed in cells. A high amount in the cell (as compared with a non-apoptotic control) correlates with cells

destined to die. A lessened amount correlates with a reduced likelihood cells will die. Expression of Apaf-1 in apoptotic cells is expected to be at least 2-10X greater than expression in normal, non-apoptotic cells.

5 Nucleic acid sequences encoding Apaf-1 are also useful for the preparation of Apaf-1 by the recombinant techniques described herein.

Isolated Apaf-1 may be used in quantitative diagnostic assays as a control against which samples containing unknown quantities of Apaf-1 may be prepared. Apaf-1 preparations are also useful in generating antibodies, as standards in assays for Apaf-1 (*e.g.*, by labeling Apaf-1 for use as a standard in a
10 radioimmunoassay, radioreceptor assay, or enzyme-linked immunoassay), in affinity purification techniques, and in competitive-type receptor binding assays when labeled with, for instance, radioiodine, enzymes, or fluorophores. Modified forms of the Apaf-1, such as the Apaf-1-IgG chimeric molecules (immunoadhesins) described above, can be used as immunogens in producing anti-Apaf-1 antibodies.

15 Nucleic acids which encode Apaf-1 or its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (*e.g.*, a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the
20 animal at a prenatal, *e.g.*, an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding Apaf-1 or an appropriate sequence thereof (such as Apaf-1-IgG) can be used to clone genomic DNA encoding Apaf-1 in accordance with established techniques and the genomic sequences used to generate transgenic
25 animals that contain cells which express DNA encoding Apaf-1. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for Apaf-1 transgene incorporation with tissue-specific enhancers. Transgenic animals that
30 include a copy of a transgene encoding Apaf-1 introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding Apaf-1. Such animals can be used as tester animals for

reagents thought to confer protection from, for example, pathological conditions associated with excessive apoptosis. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of Apaf-1 can be used to construct an Apaf-1 "knock out" animal which has a defective or altered gene encoding Apaf-1 as a result of homologous recombination between the endogenous gene encoding Apaf-1 and altered genomic DNA encoding Apaf-1 introduced into an embryonic cell of the animal. For example, cDNA encoding Apaf-1 can be used to clone genomic DNA encoding Apaf-1 in accordance with established techniques. A portion of the genomic DNA encoding Apaf-1 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas and Capecchi, *Cell*, 51:503 (1987) for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected (see e.g., Li et al., *Cell*, 69:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras (see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed., IRL, Oxford, 1987, pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the Apaf-1 polypeptide, including for example, non-regulated growth of cells and/or development of tumors.

The Apaf-1 protein is useful in assays for identifying therapeutically active molecules that modulate apoptosis. Specifically, compounds that either inhibit the initiation of apoptosis or enhance such initiation can be conveniently identified by these screening methods. Molecules inhibiting apoptosis are useful to prevent cell death, for example in degenerative disease or to extend the life of cultured cells. Molecules enhancing or promoting the initiation of apoptosis are useful for example in therapy of cancer or in promoting death of particular cells *in vitro*.

Assay of candidate compounds able to competitively compete with cytochrome c for specific binding to Apaf-1 provides for high-throughput screening of chemical libraries, and is particularly useful for screening small molecule drug candidates. Small molecules, usually less than 10K molecules weight, are desirable as therapeutics since they are more likely to be permeable to cells, are less susceptible to degradation by cells, and are not as apt to elicit an immune response as proteins. Small molecules include, but are not limited to, synthetic organic or inorganic compounds. Many pharmaceutical companies have extensive libraries of such molecules, which can be conveniently screened by assessing binding to Apaf-1.

Assay of candidate compounds able to replace cytochrome c in the cell free apoptotic assay described above further provides a useful screening method. Combining Apaf-1, Apaf-3, dATP, and the candidate drug, for example, in the presence of caspase 3 precursor, induction of apoptosis is measured by proteolytic cleavage of the enzyme or by DNA fragmentation.

C. Anti-Apaf-1 Antibody Preparation

The present invention further provides anti-Apaf-1 antibodies. Antibodies against Apaf-1 may be prepared as follows. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The Apaf-1 antibodies may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent

and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the Apaf-1 polypeptide or a fusion protein thereof. An example of a suitable immunizing agent is a Apaf-1-IgG fusion protein or chimeric molecule (including an Apaf-1 ECD-IgG fusion protein). Cells expressing Apaf-1 at their surface may also be employed. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins which may be employed include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. An aggregating agent such as alum may also be employed to enhance the mammal's immune response. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation. The mammal can then be bled, and the serum assayed for antibody titer. If desired, the mammal can be boosted until the antibody titer increases or plateaus.

2. Monoclonal Antibodies

The anti-Apaf-1 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Supra*. In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized (such as described above) with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the Apaf-1 polypeptide or a fusion protein thereof. An example of a suitable immunizing agent is a Apaf-1-IgG fusion protein or chimeric molecule. Cells expressing Apaf-1 at their surface may also be employed. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice,

Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur, et al., *Monoclonal Antibody Production Techniques and Applications*, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against Apaf-1. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods (Goding, *Supra*). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, et al., *Supra*) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent cross-linking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art. For instance, digestion can be performed using papain. Examples of papain digestion are described in WO 94/29348 published 12/22/94 and U.S. Patent No. 4,342,566. Papain digestion of antibodies typically produces two identical antigen binding fragments, called Fab fragments, each with a single antigen binding site, and a residual Fc fragment. Pepsin treatment yields an F(ab')₂ fragment that has two antigen combining sites and is still capable of cross-linking antigen.

The Fab fragments produced in the antibody digestion also contain the constant domains of the light chain and the first constant domain (CH₁) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH₁ domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. F(ab')₂ antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

3. Humanized Antibodies

The Apaf-1 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the

humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized
5 antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones, et al., *Nature*, 321:522-525 (1986); Reichmann, et al., *Nature*, 332:323-329 (1988); and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

Methods for humanizing non-human antibodies are well known in the
10 art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers (Jones, et al., *Nature*, 321:522-525 (1986);
15 Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeyen, et al., *Science*, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the
20 corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

The choice of human variable domains, both light and heavy, to be
25 used in making the humanized antibodies is very important in order to reduce antigenicity. According to the "best-fit" method, the sequence of the variable domain of a rodent antibody is screened against the entire library of known human variable domain sequences. The human sequence which is closest to that of the rodent is then accepted as the human framework (FR) for the humanized antibody
30 (Sims, et al., *J. Immunol.*, 151:2296 (1993); Chothia and Lesk, *J. Mol. Biol.*, 196:901 (1987)). Another method uses a particular framework derived from the consensus sequence of all human antibodies of a particular subgroup of light or

heavy chains. The same framework may be used for several different humanized antibodies (Carter, et al., *Proc. Natl. Acad. Sci. USA*, 89:4285 (1992); Presta, et al., *J. Immunol.*, 151:2623 (1993)).

It is further important that antibodies be humanized with retention of high affinity for the antigen and other favorable biological properties. To achieve this goal, according to a preferred method, humanized antibodies are prepared by a process of analysis of the parental sequences and various conceptual humanized products using three dimensional models of the parental and humanized sequences. Three dimensional immunoglobulin models are commonly available and are familiar to those skilled in the art. Computer programs are available which illustrate and display probable three-dimensional conformational structures of selected candidate immunoglobulin sequences. Inspection of these displays permits analysis of the likely role of the residues in the functioning of the candidate immunoglobulin sequence, i.e., the analysis of residues that influence the ability of the candidate immunoglobulin to bind its antigen. In this way, FR residues can be selected and combined from the consensus and import sequence so that the desired antibody characteristic, such as increased affinity for the target antigen(s), is achieved. In general, the CDR residues are directly and most substantially involved in influencing antigen binding (see, WO 94/04679 published 3 March 1994).

Transgenic animals (e.g., mice) that are capable, upon immunization, of producing a full repertoire of human antibodies in the absence of endogenous immunoglobulin production can be employed. For example, it has been described that the homozygous deletion of the antibody heavy chain joining region (J_H) gene in chimeric and germ-line mutant mice results in complete inhibition of endogenous antibody production. Transfer of the human germ-line immunoglobulin gene array in such germ-line mutant mice will result in the production of human antibodies upon antigen challenge (see, e.g., Jakobovits, et al., *Proc. Natl. Acad. Sci. USA*, 90:2551-255 (1993); Jakobovits, et al., *Nature*, 362:255-258 (1993); Bruggermann, et al., *Year in Immuno.*, 7:33 (1993)). Human antibodies can also be produced in phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks, et al., *J. Mol. Biol.*, 222:581 (1991)). The techniques of Cole, et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies

(Cole, et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, p. 77 (1985) and Boerner, et al., *J. Immunol.*, 147(1):86-95 (1991)).

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for the Apaf-1, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art.

Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Millstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker, et al., *EMBO J.*, 10:3655-3659 (1991).

According to a different and more preferred approach, antibody variable domains with the desired binding specificities (antibody-antigen combining sites) are fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. This provides for great flexibility in adjusting the mutual proportions of the three polypeptide fragments in embodiments when unequal ratios of the three polypeptide chains used in the construction provide the optimum yields. It is, however, possible to insert the coding sequences for two or all three polypeptide chains in one expression vector when the expression of at least two polypeptide chains in equal

ratios results in high yields or when the ratios are of no particular significance. In a preferred embodiment of this approach, the bispecific antibodies are composed of a hybrid immunoglobulin heavy chain with a first binding specificity in one arm, and a hybrid immunoglobulin heavy-chain/light-chain pair (providing a second binding specificity) in the other arm. It was found that this asymmetric structure facilitates the separation of the desired bispecific compound from unwanted immunoglobulin chain combinations, as the presence of an immunoglobulin light chain in only one half of the bispecific molecule provides for a facile way of separation. This approach is disclosed in WO 94/04690 published 3 March 1994. For further details of generating bispecific antibodies see, for example, Suresh, et al., *Methods in Enzymology*, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (US Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving cross-linking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Pat. No. 4,676,980.

D. Therapeutic and Non-therapeutic Uses for Apaf-1 Antibodies

The anti-Apaf-1 antibodies of the invention have therapeutic utility. Agonistic anti-Apaf-1 antibodies, for instance, may be employed to activate or stimulate apoptosis in cancer cells. Alternatively, antagonistic antibodies may be used to block excessive apoptosis (for instance in neurodegenerative disease) or to block potential autoimmune/inflammatory effects of Apaf-1 resulting from caspase-3 activation.

Anti-Apaf-1 antibodies may further be used in diagnostic assays for anti-Apaf-1, *e.g.*, detecting its expression in specific cells, tissues, or serum.

Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases (Zola, *Monoclonal Antibodies: A Manual of Techniques*, CRC Press, Inc. (1987) pp. 147-158). The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^3H , ^{14}C , ^{32}P , ^{35}S , or ^{125}I , a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter, et al., *Nature*, 144:945 (1962); David, et al., *Biochemistry*, 13:1014 (1974); Pain, et al., *J. Immunol. Meth.*, 40:219 (1981); and Nygren, *J. Histochem. and Cytochem.*, 30:407 (1982).

Anti-Apaf-1 antibodies also are useful for the affinity purification of Apaf-1 from recombinant cell culture or natural sources. In this process, the antibodies against Apaf-1 are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the Apaf-1 to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the Apaf-1, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the Apaf-1 from the antibody.

Anti-Apaf-1 antibodies are also useful for identifying Apaf-1 expression in cells, providing a useful marker for cells that are destined to die, as discussed above.

E. Kits Containing Apaf-1 or Anti-Apaf-1 Antibodies

In a further embodiment of the invention, there are provided articles of manufacture and kits containing Apaf-1 or anti-Apaf-1 antibodies which can be used, for instance, for the therapeutic or non-therapeutic applications described above. The article of manufacture comprises a container with a label. Suitable

containers include, for example, bottles, vials, and test tubes. The containers may be formed from a variety of materials such as glass or plastic. The container holds a composition which includes an active agent that is effective for therapeutic or non-therapeutic applications, such as described above. The active agent in the composition is Apaf-1 or an anti-Apaf-1 antibody. The label on the container indicates that the composition is used for a specific therapy or non-therapeutic application, and may also indicate directions for either *in vivo* or *in vitro* use, such as those described above.

The kit of the invention will typically comprise the container described above and one or more other containers comprising materials desirable from a commercial and user standpoint, including buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

F. Model System for Apoptosis

A model system for apoptosis is provided by combining Apaf-1, Apaf-3, dATP and cytochrome c. This combination of agents, both in a cell-free system and in mammalian cells, is sufficient to trigger apoptosis. The apoptotic activity is confirmed by the proteolytic cleavage of caspase 3 and by the fragmentation of DNA. Using this model system, novel agents are screened for their ability to modulate apoptosis.

EXAMPLES

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

25 Experimental Procedures

General Methods and Materials

dATP and other nucleotides were obtained from Pharmacia; Radioactive materials from Amersham; molecular weight standards for SDS-PAGE and gel filtration chromatography from Bio-Rad. Protein concentrations were determined by the Bradford assay method. General molecular biology methods were performed as described in Sambrook, et al., *Supra*.

Example 1**Fractionation of Cytosol and Reconstitution of Caspase-3 Activation**

The cell-free *in vitro* system as described in Liu, et al. (*Cell* 86:147-157, (1996)) in which the apoptotic program is initiated by the addition of deoxyATP (dATP) was used in these studies. This system permits fractionation and purification of the biochemical components that trigger activation of the apoptotic proteases and DNA fragmentation. Initiation of the apoptotic program leads to activation of caspase-3 and ultimately to fragmentation of DNA in nuclei. The cell free system combined fractions of a 100,000Xg cytosolic supernatant prepared from suspension cultures of Hela cells (S100) with cytochrome c and activating dATP, as described in Liu, et al., *Supra*. Caspase-3 activation was monitored by cleavage of *in vitro* translated, ³⁵S-labeled, affinity purified caspase-3 precursor. The proteolytic fragments were visualized by SDS-polyacrylamide gel electrophoresis (PAGE) followed by phosphorimaging.

Figure 1 shows the fractionation scheme that was used for the separation of cytochrome c (Apaf-2) from two other required factors, Apaf-1 and Apaf-3. The initial step was SP-Sepharose chromatography of an S100 fraction of Hela cell Cytosol, which separated cytochrome c (bound fraction) from Apaf-1 and Apaf-3 (flow-through). To separate Apaf-1 and Apaf-3, the flow-through fraction was loaded onto a hydroxylapatite column. The flow-through and bound fractions were collected.

To reconstitute the apoptotic initiation program, aliquots of flow-through (Apaf-3) and bound (Apaf-1) fractions (4μl each) and an aliquot of cytochrome c (0.2 μg in 1 μl) were incubated alone or in combination and in the presence or absence of 1 mM dATP. To each sample was added an aliquot (3μl) of *in vitro* translated, ³⁵S-labeled caspase 3, and the samples were incubated at 30°C for one hour in a final volume of 25 μl buffer A (20 mM Hepes-KOH, pH 7.5, 10 mM KCl, 1.5 mM MgCl₂, 1 mM sodium EDTA, 1 mM sodium EGTA, 1 mM dithiothreitol, and 0.1 mM PMSF). At the end of the incubation, 7 μl of 4X SDS sample buffer was added to each reaction. After boiling for three minutes, each sample was then subjected to 15% SDS PAGE and the gel transferred to a

nitrocellulose filter. The filter was exposed to a Phosphorimaging plate for 16 hours at room temperature and visualized in a Fuji BAS-1000 Phosphorimager.

As shown in Figure 2, neither fraction alone was competent to activate caspase-3 when incubated with cytochrome c and dATP (lanes 1-6).

5 However, when the two fractions (Apaf-1 and Apaf-3) were mixed in the presence of cytochrome c and dATP, caspase-3 activation was restored (lane 7). No activity was detected when dATP and cytochrome c were omitted from the reaction (lanes 5 and 8).

10

Example 2

Purification of Apaf-1

For purposes of further purification, Apaf-1 activity was assayed by incubating various cytosolic fractions with cytochrome c, dATP, the crude Apaf-3 fraction, and the substrate caspase-3, as described above for Example 1. As shown
15 above for Example 1, Apaf-1 activity results in the cleavage of caspase-3. Purification of Apaf-1 was achieved through a six step procedure.

Purification of Apaf-1 from HeLa S-100

All purification steps were carried out at 4°C. All chromatography
20 steps except the SP-Sepharose column (Pharmacia) and first Hydroxylapatite column (Bio-Rad) were carried out using an automatic fast protein liquid chromatography (FPLC) station (Pharmacia).

700 ml (4.9 g of protein) of HeLa S-100 from 100 liter of suspension
cultured HeLa cells were prepared as described in Liu, et al., 1996, *Supra*, and
25 applied to a SP-Sepharose column (200 ml bed volume) equilibrated with buffer A. Buffer A was prepared as described above for Example 1. An 800 ml flow-through fraction (3,648 mg of protein) was collected and loaded directly to a
Hydroxylapaptite column (50 ml bed volume) equilibrated with buffer A. The
column was washed with three column volumes of buffer A containing 1 M NaCl
30 followed by two column volumes of buffer A. The bound material was eluted with
200 ml 0.3 M KPO₄, pH 7.5. The protein peak was eluted from the column (115 ml,
287 mg of protein), was dialyzed against buffer A, and then loaded onto a second

hydroxylapatite column (10 ml bed volume) equilibrated with buffer A. The column was eluted with 200 ml buffer A to 250 mM KPO_4 , pH 7.5 linear gradient.

Fractions of 10 ml each were collected and assayed for Apaf-1 activity (caspase-3 cleavage), as described above for Example 1. Active fractions (50 ml, 11 mg of protein) were pooled, dialyzed against buffer A, and then loaded onto a 5 ml Heparin-Sepharose column (Pharmacia) equilibrated with buffer A. The column was washed with 20 ml of buffer A containing 100 mM NaCl and eluted with 50 ml buffer A containing 100 mM NaCl to buffer A containing 400 mM NaCl linear gradient. Fractions of 4 ml were collected and assayed for Apaf-1 activity. The active fractions (8 ml, 0.86 mg of protein) were pooled and loaded directly onto a Superdex 200 16/60 gel-filtration column equilibrated with buffer A containing 100 mM NaCl. The column was eluted with the same buffer (4 runs) and fractions of 4 ml were collected starting from 30 ml of elution. The fractions were assayed for Apaf-1 activity and the active fractions were pooled (16 ml, 32 μ g of protein) and loaded directly onto a Mono Q 5/5 column (Pharmacia) equilibrated with buffer A containing 100 mM NaCl. The column was eluted with 20 ml buffer A containing 100 mM NaCl to buffer A containing 300 mM NaCl linear gradient. Fractions of 1 ml were collected and assayed for Apaf-1 activity.

The results of the last step of purification, a Mono Q column chromatography, are shown in Figure 3. Apaf-1 activity was eluted from the Mono Q column at 250 mM NaCl (fractions 14-16). The same fractions were subjected to SDS-PAGE followed by silver staining and a single polypeptide band of approximately 130 kDa was observed to co-elute with the Apaf-1 activity (Figure 4). No other proteins were detected by silver staining in the peak fractions (Figure 4, fractions 14-16). Apaf-1 activity eluted from gel-filtration columns at about 130 kDa, indicating that Apaf-1 exists as a monomer in solution (data not shown). About 10 microgram of pure Apaf-1 was obtained from the cytosol from 100 liters of HeLa cells.

Example 3**Sequencing and cDNA Cloning of Apaf-1**

The 130 kDa Apaf-1 protein was excised from an SDS gel and subjected to trypsin and Lys-C digestion. The resulting peptides were separated by capillary reverse-phase high pressure liquid chromatography (HPLC).

Protein Sequencing of Apaf-1

The 130 kDa Apaf-1 protein band produced as described for Example 2 and shown in Figure 4 (4-8 pmol) was electroeluted from the SDS polyacrylamide gel onto PVDF membrane (ProBlott, Applied Biosystem) and excised. The band was reduced and alkylated with isopropylacetamide followed by digestion in 20 μ l of 0.05 M ammonium bicarbonate, 20% acetonitrile with 0.2 μ g of trypsin (Promega) or Lysine-C (Wako) at 37°C for 17 hours, as described previously (Henzel, et al., *Proc. Natl. Acad. Sci. USA* 90:5011-5015 (1993)). The digested solution was then directly injected onto a 0.32 x 150 mm C18 capillary column (LC Packing, Inc.). Solvent A was 0.1% aqueous TFA and solvent B was acetonitrile containing 0.07% TFA. The peptides were eluted with a linear gradient of 0-80% solvent B in 120 minutes. Peptide peaks were detected at 195 nm and hand collected into 0.5 ml Eppendorf tubes.

An aliquot (0.2 μ l) of each of the isolated HPLC fractions was applied to a pre-made spot of matrix (0.5 μ l of 20 mg/ml α -cyano-4-hydroxycinnamic acid plus 5 mg/ml nitrocellulose in 50% acetone/50% 2-propanol) on the target plate (Shevchenko, et al., *Anal. Chem.*, 68:850-858, (1996)). Ions were formed by matrix-assisted laser desorption/ionization with a 337 nm nitrogen laser. Spectra were acquired with a Perseptive Biosystems Voyager Elite time-of-flight mass spectrometer, operated in linear delayed extraction mode. Subsequently, fragment ions for selected precursor masses were obtained from post-source decay (PSD) experiments (Kaufmann, et al., *International J. Mass Spectro. and Ion Processes*, 131:355-385, (1994)). In order to enhance the ion abundances at low mass, collision gas (air) was introduced to the collision cell during the acquisition of the lower portion (<200 μ) of the fragment ion spectrum. Each peptide mass and its

associated fragment ion masses were used to search an in-house sequence database with an enhanced version of the FRAGFIT program (Henzel, et al., *Supra*).

Automated protein sequencing was performed on models Procise 494A and 494CL, Applied Biosystems sequencers equipped with an on-line PTH analyzer. Peaks were integrated with Justice Innovation software using Nelson
5 Analytical 760 interfaces. Sequence interpretation was performed on a DEC Alpha (Henzel, et al., 1987, *J. Chromatogr.* 404:41-52).

The sequences of 14 peptides were determined by Mass Spectrometry and Edman degradation (Table 1). Protein data base searches revealed that no
10 proteins in the data bases were identical to Apaf-1.

56

Table 1**Peptide Analyzed by MALDI MS Analysis and Edman Sequencing**

Peptide #	Fraction #	MH+ Measured	MH+ Calculated	Residues in cDNA
1	Lys-C33a	1780.8	1780.0	5- 18
2	Lys-C6	1255.7	1255.6	43- 52
3	Lys-C43	N.D.	4103.82	82- 100
4	T9	913.5	913.5	121- 128
5	Lys-C46	3153.6 ^{ab}	3153.6	150- 178
6	Lys-C42	3160.8 ^a	3161.8	187- 212
7	Lys-C31	1177.8	1177.7	266- 276
8	Lys-C29	1450.0	1449.8	291- 303
9	Lys-C53	2747.1 ^b	2745.4	502- 524
10	T24	817.6	817.5	590- 596
11	Lys-C33b	735.5	735.9	627- 632
12	Lys-C31	615.3	615.4	638- 642
13	Lys-C32	1278	1277.7	746- 757
14	T17	650.5	650.4	941- 945
15	Lys-C28	942.7	942.5	1169-1176

^a Masses that represent average isotopic mass.

5 ^b The peptide contained a oxidized tryptophan resulting in an additional 32 daltons.

Degenerate oligonucleotides encoding peptides 2 and 4 (see Table 1) were used to prime polymerase chain reactions (PCRs) with a HeLa cDNA library as
10 template.

cDNA Cloning of Apaf-1

An aliquot of 1 μ l (10^8 pfu) of a λ Elox Hela cDNA library (Yokoyama, et al., Cell, 75:187-197, (1993)) was heated at 99°C for 15 minutes to release the DNA contents. The DNA was directly amplified with 300 pmol of the primer-1 degenerate oligonucleotide (SEQ ID NO:3) and 20 pmol SP6 primer (SEQ ID NO:4) using PCR reaction with 5 cycles of 94°C for 30 seconds, 60°C for 30 seconds; and 72°C for 1 minute followed by 25 cycles of 94°C for 30 seconds; 55°C for 30 seconds; and 72°C for 1 minute. The PCR product was purified by passing through a PCR Purification column (Qiagen). A portion (1/50) of the purified product was further amplified using 300 pmol of primer-2 (SEQ ID NO:5) and pmol of primer-3 (SEQ ID NO:6) in a PCR reaction as described above.

Primer	Sequence	SEQ ID NO.
1	5' -AAAGTAAGAAATGAACCAAC-3' G TC T C G T C C C G G G	3
SP6	5' -ATTTAGGTGACACTATAGAA-3'	4
2	5' -AATGAACCAACACAACAACA-3' C G T T C C G G	5
3	5' -TTTTGTTGTATAGCATTCAC-3' C C A T G G C G	6

A 285 bp PCR product was obtained and subsequently sequenced after subcloning into a PCR II vector using the TA cloning kit (Invitrogen).

The sequencing yielded a 285 base pair DNA fragment that also encoded peptide 3. Probing of the HeLa cDNA library with this PCR fragment identified two overlapping cDNAs with an open reading frame encoding 1194 amino acids with a calculated molecular mass of 136,088 daltons (Figures 5A - 5O). Multiple in-frame stop codons were identified in the 5' untranslated region of the

cDNA, indicating that these cDNA clones encode full length Apaf-1 (data not shown).

The 285 bp PCR product was labeled with α -³²P-dCTP using *red*i prime RANDOM Primer Labelling Kit (Amersham) and used to screen the Hela
5 λ Exlox cDNA library by hybridizing duplicate filters at 42°C overnight in Rapid-hyb buffer (Amersham). The filters were washed twice with 1x saline citrate (SSC)/0.1% SDS for 15 minutes at room temperature and once with 0.5 x SSC/0.1% SDS for 10 minutes at 65°C. Out of 6x10⁵ plaques that were screened, 4 positive clones were identified and a 1.6 kb partial clone was characterized. The 1.6 kb insert
10 was excised and labeled with α -³²P-dCTP, as described above. The Hela cell cDNA library was re-screened with this 1.6 kb cDNA fragment using the procedures described above. Positive clones (45) were identified and characterized. The longest clone (3kb) containing the 5' portion of Apaf-1 was sequenced. An aliquot of 40 ng of this plasmid was amplified by two PCR primers designated APPN
15 (5'ACATCACGAATCTTTCCCGC) (SEQ ID NO:7) and APPC (5'AACACTTCACTATCACTTCC3') (SEQ ID NO:8) according the 3' end of the 3kb insert. A 600 bp PCR fragment was generated and labeled with α -³²P-dCTP, as above. The same filters were rescreened with this PCR fragment and 35 positive clones were identified and characterized. The longest one, having an insert of 4.5 kb
20 was sequenced. This clone contained the 3' portion of Apaf-1 which overlaps with the 3 kb 5' clone by 500 bp. The full length cDNA was obtained by ligating the two clones at an EcoR I site located within the 500 bp overlapping region, and is shown in Figures 5A-O (SEQ ID NO: 1).

Figures 5A-O also shows the predicted amino acid sequence (SEQ ID
25 NO:2) encoded by human Apaf-1 (SEQ ID NO:1). All 14 previously sequenced peptides were encoded by the open reading frame of the cDNA (indicated by underlines) and distributed throughout the protein. The measured molecular masses of these peptides by Mass Spectrometry were consistent with the calculated molecular masses, indicating that there is no post-translational modification of these
30 peptides.

Example 4**Domain Structure of Apaf-1**

A search of protein data bases (GenBank and Prosite) revealed that the COOH-terminal segment of Apaf-1 contains 13 WD-40 repeats. This loosely conserved set of sequences is found in many regulatory proteins, including β -subunits of heterotrimeric G proteins (Neer, et al., *Nature*, 371:297-300, (1994); Wall, et al., *Cell*, 83:1047-1058, (1995); Sondek, et al., *Nature*, 379:369-374, (1996)) the LIS-1 gene for Miller-Dieker Lissencephaly (Hattori, et al., 1994, *Nature* 370:216-218), and the SREBP cleavage-activating protein (SCAP) (Hua, et al., *Cell*, 87:415-426, (1996)).

The searches of the protein data base (GenBank) also revealed significant similarities between Apaf-1, and *Ced-3* and *Ced-4*, two proteins that are required for the apoptotic program in *C. elegans* (Yuan and Horvitz, *Supra*) (see Figures 8 and 9). The NH₂-terminal 85 amino acids of Apaf-1 shows 21% identity and 53% similarity with the NH₂-terminal pro-domain of *Ced-3* (Figure 9). The amino acid sequence of Apaf-1 and the N-terminal prodomain of *Ced 3* were aligned to the Lipman-Pearson method of DNASTAR program. This domain is followed by a stretch of 310 amino acids that shows 22% identity and 48% similarity with *Ced-4* aligned, as described above (Figure 9). The two longest stretches of amino acids conserved with *Ced-4* lie at positions 142-157 and 227-234 of Apaf-1 (Figure 7, underlined). These two regions correspond to Walker's A- and B-box consensus sequences for nucleotide binding proteins, respectively (Walker, et al., 1982, *EMBO J.* 1:945-951). Several amino acids that are known to be required for *Ced-4* activity, including two aspartic acids residues at positions 250-251 and the isoleucine at position 258 (Yuan and Horvitz, *Supra*; Chinnaiyan, et al., *Supra*), are conserved in Apaf-1 (Figure 9, asterisks).

Example 5**Tissue Distribution of Apaf-1****Northern Blotting Analysis**

Poly(A)+ RNA blots containing 2 μ g of poly(A)+ RNA per lane from multiple human adult and fetal tissues were purchased from Clontech. Blots were

60

hybridized with 2×10^6 cpm/ml of a random primed 607 bp Apaf-1 PCR fragment corresponding to amino acid 590-792 (amplified with 5'ACATCACGAATCTTTCCCGC 3' (SEQ ID NO:9) and 5' CAACACTTCACTATCACTTCC 3') (SEQ ID NO:10) in Rapid-hyb buffer (Amersham) at 65°C overnight. The filters were washed twice with 1 x SSC/0.5% SDS for 15 minutes at 65°C followed by 0.5% SSC/0.5% SDS for 20 minutes at 65°C.

The same filters were also hybridized at 65°C for 2 hours with a 2.0 kb β -actin cDNA probe and the filters were then washed, as above. The left filter was exposed to an X-ray film with an intensifying screen at -80°C for 8 days and the right two filters were exposed to an X-ray film with an intensifying screen at -80°C for 4 days. The same filters were subsequently hybridized with human β -actin and exposed to film for 2 hours (left filter) or 30 minutes (right two filters) at -80°C with an intensifying screen. The samples from human adult or fetal tissue are indicated.

Figure 10 shows the tissue distribution of Apaf-1 mRNA analyzed by Northern blotting of human adult and fetal tissues by the methods described above. In all tissues examined, including adult brain, heart, liver, spleen, skeletal muscle, lung, pancreas, thymus, small intestine, colon, peripheral blood leukocytes, kidney, testis, ovary, and fetal brain, liver, kidney and lung, a predominant mRNA of about 8,000 nucleotides was detected, indicating ubiquitous expression of Apaf-1. Expression was highest in adult spleen and peripheral blood leukocytes and fetal brain, kidney and lung, all of which have a high level of apoptosis.

Example 6

Production of Recombinant Apaf-1

Apaf-1 was translated *in vitro* from a 6.0 kb cDNA fragment containing the entire coding region of Apaf-1 plus the 577 bp of the 5' untranslated region and about 1.5kb of the 3' untranslated region cloned into Not I and EcoRI sites of a PCDNA 3.1(-) vector (Invitrogene). The 3' region was truncated at the Eco RI site located at nucleotide 5727 of SEQ ID NO:1. The TNT T7 coupled reciculocyte lysate system (Promega) was used according to the manufacturing instructions.

The translation products were analyzed by Western blotting. Lane 1 contained 30 μ l of the Apaf-1 *in vitro* translation product. Lane 2 contained 30 μ l of a translation mixture control, where the vector alone was used as a template in the same translation reaction. Lane 3 contained 5 μ l of Apaf-1 purified as described for Example 1.

After electrophoresis, the samples were transferred to a nitrocellulose filter and the filter was probed with an antiserum (1:5000 dilution) against Apaf-1 raised against a recombinant fusion protein containing amino acids 10-254 of Apaf-1 [SEQ ID NO:1]. The fusion protein (0.5mg) was used to immunize rabbits via subcutaneous injection of the protein in Freund's Complete adjuvant (1:1). Two boosters of protein in Freund's Incomplete adjuvant (1:1) were given in intervals of two weeks. Antigen-antibody complexes formed in the blot were visualized by an ECL method as described above. The filter was exposed to a Kodak X-OMAT X-ray film for 5 seconds.

As shown in Figure 11, the anti-Apaf-1 polyclonal antibody reacted with the purified protein (lane 3). The same antibody also reacted with Apaf-1 synthesized by *in vitro* translation of the Apaf-1 cDNA in a rabbit reticulocyte lysate system (lane 1). X denotes a cross reactive band in the reticulocyte lysate. The *in vitro* translated Apaf-1 migrated in an identical position as purified Apaf-1, confirming that the cDNA encoded full length Apaf-1.

Example 7

Cytochrome c Binds to Apaf-1

As noted above, the activation of caspase-3 by Apaf-1 and Apaf-3 requires cytochrome c as well as dATP. To determine whether Apaf-1 interacts directly with cytochrome c, binding studies were performed.

Aliquots of 100 μ l of partially purified Apaf-1 were incubated alone or with 2 μ g of cytochrome c in the absence or presence of 1 mM dATP, at 30°C for twenty minutes. The Apaf-1 was partially purified from HeLa cell cytosol through a SP Sepharose column as well as a hydroxylapatite column, as described for Example 2.

After the incubation period, the samples were immunoprecipitated with an anti-Apaf-1 antiserum. The anti-Apaf-1 anti-serum was

generated by immunizing rabbits with a recombinant Apaf-1 fusion protein produced as described below.

Primers 5'-

GCAAAAGCTCGAAATCATATGCTTCAACATAGAG-3' (SEQ ID NO:11) and
5'-TCGCGGCCGCCTCGAGGGCTCTGGTTGTAAG-3' (SEQ ID NO:12) were
designed to PCR-amplify the 1.6 kb plasmid Apaf-1 cDNA open reading frame.
The amplified 800 bp fragment encoding amino acid 10-254 of Apaf-1 was
subcloned inframe into the NdeI/XhoI sites of the bacterial expression vector pET-
15b (Novagen). The expression plasmid was transformed into bacteria BL21(DE3).
In a typical Apaf-1 preparation, a 10 ml overnight cultured bacteria containing Apaf-
1 expression vector was added into a 500 ml LB broth, cultured for 3 hours by
shaking at 220 rpm at 37°C. Isopropyl-1-thio-B-D-galactopyranoside (IPTG) was
then added to a final concentration of 1 mM and the mixture cultured for another 2
hours. The bacterial pellet was resuspended in 10 ml of buffer B (6 M GuHCl, 0.1
M Na-phosphate, 0.01 M Tris-HCl, pH 8.0). After centrifugation at 10,000 g for 15
minutes, the supernatant was loaded onto a nickel affinity column (6 ml). The
column was washed with 30 ml buffer B followed by 30 ml buffer C (8 M urea, 0.1
M Na-phosphate, 0.01 M Tris-HCl, pH 8.0). The column was eluted with Buffer C
containing 250 mM imidazole. About 10 mg Apaf-1 protein was purified from a
500 ml culture.

Western blot analysis of cytochrome c was performed as described in
Liu, et al., 1996, *Supra*. Anti-Apaf-1 anti-serum was generated by immunizing
rabbits with a recombinant Apaf-1 fusion protein produced as described above.
Immunoblot analysis was performed with horseradish peroxidase conjugated goat
anti-mouse (cytochrome c) or goat anti-rabbit (Apaf-1) immunoglobulin G using
Enhanced Chemiluminescence western blotting detection reagents (Amersham).

As shown in Figure 12, antiserum against Apaf-1 precipitated Apaf-1,
while the preimmune serum from the same animal did not. The antiserum against
Apaf-1 also precipitated cytochrome c (Figure 13), indicating that cytochrome c
forms a complex with Apaf-1. Preimmune serum caused no such precipitation. The
binding of cytochrome c to Apaf-1 was not influenced by the presence or absence of
dATP (Figure 13, lanes 3-6).

To confirm that cytochrome c interacts with Apaf-1 in the crude cytosolic extracts, the immunoprecipitation experiment was performed using unfractionated HeLa cell S-100. As shown in Figure 12, lanes 7-8, Apaf-1 in the S-100 fraction was precipitated by the anti-Apaf-1 antiserum. Cytochrome c, which is released to the cytosol during homogenization (see Liu et. al., 1996, *Supra*) coprecipitated with Apaf-1 (Figure 12, lane 8). The preimmune serum caused neither protein to precipitate (lane 7).

Example 8

Expression of Apaf-1 in 293 Cells

A 5.7 kb cDNA containing the entire coding region of Apaf-1 plus 577 bp of the 5' untranslated region and about 1.5 kb of the 3' untranslated region (truncated at the Eco RI site at nucleotide 5727, as described above for Example 6) was subcloned into NotI and EcoRI sites of a pcDNA 3.1(-) vector (Invitrogen) and the plasmid (pApaf-1) was prepared using a Qiagen Mega plasmid kit. Human embryonic kidney 293 cells were plated at 1×10^6 cells per 100 mm dish in medium A (Dulbecco's modified Eagle's Medium containing 100 U/ml of penicillin and 100 μ g/ml streptomycin sulfate) supplemented with 10% fetal calf serum, and grown in a monolayer at 37°C in an atmosphere of 6-7% CO₂. After incubation for 24 hours, cells were transfected with 15 μ g vector alone; 10 μ g pApaf-1 plus 5 μ g vector; 10 μ g of vector plus 5 μ g of pCPP32 (hamster caspase-3) (Wang, et.al., 1996, *EMBO J.* 15:1012-1020); or 10 μ g of pApaf-1 plus 5 μ g of pCPP32, using the MBS Transfection Kit (Stratagene) as described in (Hua, et al., *Cell* 87:415-426 (1996)). After 36 hours of further incubation, the cells were harvested and the cytosolic S-100 fractions were prepared as described above for Example 2.

Aliquots (30 μ g) of cytosol from control (lane 1), pApaf-1 transfected (lanes 2,4), and pCPP32 transfected (lanes 3,4) cells were analyzed in an 8% SDS-PAGE (upper panel of Figure 14) or in a 15% SDS PAGE (lower panel), and blotted onto nitrocellulose filters. The filters were probed with either a rabbit anti-Apaf-1 antiserum (1:2000, upper panel) or a rabbit anti-hamster caspase-3 (1:2000, lower panel) antibody, respectively. The antigen-antibody complexes were

visualized with horseradish peroxidase conjugated antibodies and Enhanced Chemiluminescence Western blotting reagents (Amersham) as described above.

The films were exposed for 30 seconds (upper panel) or 40 seconds (lower panel). Lane 5 contained 3 μ l of Apaf-1 protein purified to the MonoQ column step described for Example 2. The results are shown in Figure 14, and show expression of pApaf-1 and pCaspase-3 in mammalian cells.

The cytosol of the transfected cells was further analyzed for its ability to induce apoptosis, using the cell-free assay described above for Example 1. Aliquots of 40 μ g of cytosol prepared from 293 cells transfected with the indicated plasmids were incubated with buffer A; in the presence of 1mM dATP; 0.2 μ g of cytochrome c; or 1mM dATP plus 0.2 μ g cytochrome c at 30°C for thirty minutes, in a final volume of 20 μ l adjusted with buffer A. The samples were subjected to 15% SDS PAGE followed by Western blotting analysis using a rabbit anti-caspase 3 antibody (1:2000). The antigen-antibody complexes were visualized by an ECL method as described above.

As shown in Figure 15, recombinant Apaf-1 expressed in mammalian cells was effective in combination with dATP and cytochrome c. This mixture induced cleavage of the caspase 3 precursor to its active cleavage products.

20

Example 9

Alternatively spliced transcript: Apaf-1L

A non-matched peptide fragment:

To confirm the sequence and identity of the Apaf-1 gene described above, protein expressed from the gene was purified, digested, separated by capillary reverse-phase HPLC, and the peptide fragments analyzed by matrix-assisted laser desorption/ionization and mass spectroscopy as described for Example 3. The spectra obtained for the fragments of the expressed protein was compared with the spectra obtained from the purified protein of Example 3. Each of the peptide masses from the expressed protein appeared to match the peptide masses from the purified protein but for one fragment.

30

A comparison of the nucleotide sequence of a 1.4 kb partial clone encoding the 5' region of Apaf-1 (Ap2) with the 5' region of the Apaf-1 cDNA of

Example 3 and Figures 5 A-O revealed that Ap2 included 33 nucleotides encoding 11 amino acids at the beginning of the ced-4 homologous regions that was not seen in the cloned Apaf-1.

5 **Alternatively spliced transcript:**

To confirm the existence of an alternatively spliced transcript, RT-PCR was performed using specific primers designed to amplify the alternative spliced region of Apaf-1.

HeLa poly (A)+ mRNA was purified using Rapid mRNA purification
10 kit (Pharmacia). The first strand cDNA was carried using a First-Strand cDNA synthesis kit (Pharmacia) with the specific primer AppC 5'-AACACTTCACTATCACTTCC-3' [SEQ ID NO:8], designed from the 3' end of the 3kb clone described above for Example 3. An aliquot of 400 ng of this first strand cDNA mixture was amplified by two PCR primers: Apn 5'
15 TAATGATTCCTACGTATCATTCTACAATGC - 3' [SEQ ID NO:13] and Aps10 5'- GAATGATCTCTAACAGCTTC -3' [SEQ ID NO:14], designed from the 5' side and the 3' side of the additional 33 nucleotides, respectively.

Both 316 bp and 283 bp PCR products (shown in Figure 17) were subcloned into the PCR II vector using TA cloning kit (Invitrogen) and sequenced.
20 The resulting nucleotide sequence of Apaf-1L is shown below [SEQ ID NO:15]. Its imputed amino acid sequence is also shown [SEQ ID NO:16].

Differences between Apaf-1 and Apaf-1L:

The amino acid sequence of Apaf-1L [SEQ ID NO:16] differs from
25 the Apaf-1 amino acid sequence [SEQ ID NO:2] by having the additional 11 amino acids (GKDSVSGITSY) inserted after amino acid residue 108, shown underlined in the amino acid sequence of Apaf-1L (see Figure 17). The nucleic acid sequence of Apaf-1L [SEQ ID NO:15], with the additional 33 nucleic acids (TGGTAAAGATTCAGTTAGTGGAATAACTTCGTA) underlined, is shown in
30 Figures 16A-16H.

Two PCR products were observed from the RT-PCR reaction, as shown in Figure 18, indicating the presence of alternatively spliced transcripts from

across this region. Direct sequence analysis of the two PCR products confirmed that the longer sequence (Apaf-1L) encodes the sequence of Ap2, having the inserted 11 amino acids. The shorter product had the identical sequence of Apaf-1.

Because the additional eleven amino acids are located at the junction
5 of the Ced-3 homologous region and the Ced-4 homologous region, the overall homology of Apaf-1L with Ced-3 and Ced-4 remains the same as that of Apaf-1. The calculated mass of amino acid 82-100 of Apaf-1L also correlated with the mass measured for peptide K43 of the purified Apaf-1 described above in Example 3, indicating that Apaf-1L exists in the Apaf-1 purified from HeLa cells.

10

Functional expression of Apaf-1L:

To test the function of recombinant Apaf-1L, a plasmid containing a cDNA insert encoding hamster caspase-3 was co-transfected into human embryonic kidney 293 cells together with a plasmid containing a cDNA insert encoding Apaf-
15 1L (entire coding region of Apaf-1L plus 577 base pairs of 5' untranslated region and 1.5 kb 3' untranslated region) or with the same region of Apaf-1 was inserted into pcDNA3.1. The transfection methods were those described above for Example 8.

The transformed cells were assayed for caspase-3 activating activity
20 as described above for Example 3. As shown in Figures 19A and 19B, Apaf-1L functioned similarly to Apaf-1, as a higher expression of caspase-3 was observed when cotransfected with Apaf-1L. In addition, cell extracts from Apaf-1L overexpressing cells showed higher caspase activating activity.

25 All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

WO 98/55615

PCT/US98/11773

67

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Genentech, Inc. and
University of Texas Board of Regents
- (ii) TITLE OF THE INVENTION: APAF-1, AN ACTIVATOR OF C ASPASE-3
- (iii) NUMBER OF SEQUENCES: 16
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
 - (B) STREET: 3100 Norwest Center, 90 South Seventh St
 - (C) CITY: Minneapolis
 - (D) STATE: MN
 - (E) COUNTRY: USA
 - (F) ZIP: 55402
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: Unassigned
 - (B) FILING DATE: 05-JUN-1998
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 60/048,807
 - (B) FILING DATE: 05-JUN-1997
 - (A) APPLICATION NUMBER: 60/055,258
 - (B) FILING DATE: 07-AUG-1997
- viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Kettelberger, Ph.D., Denise M
 - (B) REGISTRATION NUMBER: 33,924
 - (C) REFERENCE/DOCKET NUMBER: 11669.6WO01
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 612-332-5300
 - (B) TELEFAX: 612-332-9081
 - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7042 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: Coding Sequence
 - (B) LOCATION: 578...4159
 - (D) OTHER INFORMATION:

WO 98/55615

PCT/US98/11773

68

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

AAGAAGAGGT	AGCGAGTGGA	CGTGACTGCT	CTATCCCGGG	CAAAAGGGAT	AGAACCAGAG	60
GTGGGGAGTC	TGGGCAGTCG	GCGACCCGCG	AAGACTTGAG	GTGCCGCAGC	GGCATCCGGA	120
GTAGCGCCGG	GCTCCCTCCG	GGGTGCAGCC	GCCGTCGGGG	GAAGGGCGCC	ACAGGCCGGG	180
AAGACCTCCT	CCCTTTGTGT	CCAGTAGTGG	GGTCCACCGG	AGGGCGGCCC	GTGGGCCGGG	240
CCTCACCGCG	GCGCTCCGGG	ACTGTGGGGT	CAGGCTGCGT	TGGGTGGACG	CCCACCTCGC	300
CAACCTTCGG	AGGTCCCTGG	GGGTCTTCGT	GCGCCCCGGG	GCTGCAGAGA	TCCAGGGGAG	360
GCGCCTGTGA	GGCCCGGACC	TGCCCCGGGG	CGAAGGGTAT	GTGGCGAGAC	AGAGCCCTGC	420
ACCCCTAATT	CCCGGTGGAA	AACTCCTGTT	GCCGTTTCCC	TCCACCGGCC	TGGAGTCTCC	480
CAGTCTTGTC	CCGGCAGTGC	CGCCCTCCCC	ACTAAGACCT	AGGCGCAAAG	GCTTGGCTCA	540
TGGTTGACAG	CTCAGAGAGA	GAAAGATCTG	AGGGAAG	ATG GAT GCA AAA	GCT CGA	595
				Met Asp Ala Lys Ala Arg		
				1	5	
AAT TGT TTG CTT CAA CAT AGA GAA GCT CTG GAA AAG GAC ATC AAG ACA						643
Asn Cys Leu Leu Gln His Arg Glu Ala Leu Glu Lys Asp Ile Lys Thr						
	10		15		20	
TCC TAC ATC ATG GAT CAC ATG ATT AGT GAT GGA TTT TTA ACA ATA TCA						691
Ser Tyr Ile Met Asp His Met Ile Ser Asp Gly Phe Leu Thr Ile Ser						
	25		30		35	
GAA GAG GAA AAA GTA AGA AAT GAG CCC ACT CAA CAG CAA AGA GCA GCT						739
Glu Glu Glu Lys Val Arg Asn Glu Pro Thr Gln Gln Gln Arg Ala Ala						
	40		45		50	
ATG CTG ATT AAA ATG ATA CTT AAA AAA GAT AAT GAT TCC TAC GTA TCA						787
Met Leu Ile Lys Met Ile Leu Lys Lys Asp Asn Asp Ser Tyr Val Ser						
	55		60		65	70
TTC TAC AAT GCT CTA CTA CAT GAA GGA TAT AAA GAT CTT GCT GCC CTT						835
Phe Tyr Asn Ala Leu Leu His Glu Gly Tyr Lys Asp Leu Ala Ala Leu						
		75		80		85
CTC CAT GAT GGC ATT CCT GTT GTC TCT TCT TCC AGT GTA AGG ACA GTC						883
Leu His Asp Gly Ile Pro Val Val Ser Ser Ser Ser Val Arg Thr Val						
		90		95		100
CTG TGT GAA GGT GGA GTA CCA CAG AGG CCA GTT GTT TTT GTC ACA AGG						931
Leu Cys Glu Gly Gly Val Pro Gln Arg Pro Val Val Phe Val Thr Arg						
	105		110		115	
AAG AAG CTG GTG AAT GCA ATT CAG CAG AAG CTC TCC AAA TTG AAA GGT						979
Lys Lys Leu Val Asn Ala Ile Gln Gln Lys Leu Ser Lys Leu Lys Gly						
	120		125		130	
GAA CCA GGA TGG GTC ACC ATA CAT GGA ATG GCA GGC TGT GGG AAG TCT						1027
Glu Pro Gly Trp Val Thr Ile His Gly Met Ala Gly Cys Gly Lys Ser						
	135		140		145	150
GTA TTA GCT GCA GAA GCT GTT AGA GAT CAT TCC CTT TTA GAA GGT TGT						1075
Val Leu Ala Ala Glu Ala Val Arg Asp His Ser Leu Leu Glu Gly Cys						
		155		160		165
TTC CCA GGG GGA GTG CAT TGG GTT TCA GTT GGG AAA CAA GAC AAA TCT						1123
Phe Pro Gly Gly Val His Trp Val Ser Val Gly Lys Gln Asp Lys Ser						
		170		175		180
GGG CTT CTG ATG AAA CTG CAG AAT CTT TGC ACA CGG TTG GAT CAG GAT						1171
Gly Leu Leu Met Lys Leu Gln Asn Leu Cys Thr Arg Leu Asp Gln Asp						
	185		190		195	

WO 98/55615

PCT/US98/11773

69

GAG AGT TTT TCC CAG AGG CTT CCA CTT AAT ATT GAA GAG GCT AAA GAC Glu Ser Phe Ser Gln Arg Leu Pro Leu Asn Ile Glu Glu Ala Lys Asp 200 205 210	1219
CGT CTC CGC ATT CTG ATG CTT CGC AAA CAC CCA AGG TCT CTC TTG ATC Arg Leu Arg Ile Leu Met Leu Arg Lys His Pro Arg Ser Leu Leu Ile 215 220 225 230	1267
TTG GAT GAT GTT TGG GAC TCT TGG GTG TTG AAA GCT TTT GAC AGT CAG Leu Asp Asp Val Trp Asp Ser Trp Val Leu Lys Ala Phe Asp Ser Gln 235 240 245	1315
TGT CAG ATT CTT CTT ACA ACC AGA GAC AAG AGT GTT ACA GAT TCA GTA Cys Gln Ile Leu Leu Thr Thr Arg Asp Lys Ser Val Thr Asp Ser Val 250 255 260	1363
ATG GGT CCT AAA TAT GTA GTC CCT GTG GAG AGT TCC TTA GGA AAG GAA Met Gly Pro Lys Tyr Val Val Pro Val Glu Ser Ser Leu Gly Lys Glu 265 270 275	1411
AAA GGA CTT GAA ATT TTA TCC CTT TTT GTT AAT ATG AAG AAG GCA GAT Lys Gly Leu Glu Ile Leu Ser Leu Phe Val Asn Met Lys Lys Ala Asp 280 285 290	1459
TTG CCA GAA CAA GCT CAT AGT ATT ATA AAA GAA TGT AAA GGC TCT CCC Leu Pro Glu Gln Ala His Ser Ile Ile Lys Glu Cys Lys Gly Ser Pro 295 300 305 310	1507
CTT GTA GTA TCT TTA ATT GGT GCA CTT TTA CGT GAT TTT CCC AAT CGC Leu Val Val Ser Leu Ile Gly Ala Leu Leu Arg Asp Phe Pro Asn Arg 315 320 325	1555
TGG GAG TAC TAC CTC AAA CAG CTT CAG AAT AAG CAG TTT AAG AGA ATA Trp Glu Tyr Tyr Leu Lys Gln Leu Gln Asn Lys Gln Phe Lys Arg Ile 330 335 340	1603
AGG AAA TCT TCG TCT TAT GAT TAT GAG GCT CTA GAT GAA GCC ATG TCT Arg Lys Ser Ser Ser Tyr Asp Tyr Glu Ala Leu Asp Glu Ala Met Ser 345 350 355	1651
ATA AGT GTT GAA ATG CTC AGA GAA GAC ATC AAA GAT TAT TAC ACA GAT Ile Ser Val Glu Met Leu Arg Glu Asp Ile Lys Asp Tyr Tyr Thr Asp 360 365 370	1699
CTT TCC ATC CTT CAG AAG GAC GTT AAG GTG CCT ACA AAG GTG TTA TGT Leu Ser Ile Leu Gln Lys Asp Val Lys Val Pro Thr Lys Val Leu Cys 375 380 385 390	1747
ATT CTC TGG GAC ATG GAA ACT GAA GAA GTT GAA GAC ATA CTG CAG GAG Ile Leu Trp Asp Met Glu Thr Glu Glu Val Glu Asp Ile Leu Gln Glu 395 400 405	1795
TTT GTA AAT AAG TCT CTT TTA TTC TGT GAT CGG AAT GGA AAG TCG TTT Phe Val Asn Lys Ser Leu Leu Phe Cys Asp Arg Asn Gly Lys Ser Phe 410 415 420	1843
CGT TAT TAT TTA CAT GAT CTT CAA GTA GAT TTT CTT ACA GAG AAG AAT Arg Tyr Tyr Leu His Asp Leu Gln Val Asp Phe Leu Thr Glu Lys Asn 425 430 435	1891
TGC AGC CAG CTT CAG GAT CTA CAT AAG AAG ATA ATC ACT CAG TTT CAG Cys Ser Gln Leu Gln Asp Leu His Lys Lys Ile Ile Thr Gln Phe Gln 440 445 450	1939

WO 98/55615

PCT/US98/11773

70

AGA Arg 455	TAT Tyr	CAC His	CAG Gln	CCG Pro	CAT His 460	ACT Thr	CTT Leu	TCA Ser	CCA Pro	GAT Asp 465	CAG Gln	GAA Glu	GAC Asp	TGT Cys	ATG Met 470	1987
TAT Tyr	TGG Trp	TAC Tyr	AAC Asn 475	TTT Phe 475	CTG Leu	GCC Ala	TAT Tyr	CAC His	ATG Met 480	GCC Ala	AGT Ser	GCC Ala	AAG Lys	ATG Met 485	CAC His	2035
AAG Lys	GAA Glu	CTT Leu	TGT Cys 490	GCT Ala	TTA Leu	ATG Met	TTT Phe	TCC Ser 495	CTG Leu	GAT Asp	TGG Trp	ATT Ile	AAA Lys 500	GCA Ala	AAA Lys	2083
ACA Thr	GAA Glu	CTT Leu 505	GTA Val	GGC Gly	CCT Pro	GCT Ala	CAT His 510	CTG Leu	ATT Ile	CAT His	GAA Glu	TTT Phe 515	GTG Val	GAA Glu	TAC Tyr	2131
AGA Arg 520	CAT His	ATA Ile	CTA Leu	GAT Asp	GAA Glu	AAG Lys 525	GAT Asp	TGT Cys	GCA Ala	GTC Val	AGT Ser 530	GAG Glu	AAT Asn	TTT Phe	CAG Gln	2179
GAG Glu 535	TTT Phe	TTA Leu	TCT Ser	TTA Leu	AAT Asn 540	GGA Gly	CAC His	CTT Leu	CTT Leu	GGA Gly 545	CGA Arg	CAG Gln	CCA Pro	TTT Phe	CCT Pro 550	2227
AAT Asn	ATT Ile	GTA Val	CAA Gln 555	CTG Leu	GGT Gly	CTC Leu	TGT Cys	GAG Glu	CCG Pro 560	GAA Glu	ACT Thr	TCA Ser	GAA Glu	GTT Val 565	TAT Tyr	2275
CAG Gln	CAA Gln	GCT Ala	AAG Lys 570	CTG Leu	CAG Gln	GCC Ala	AAG Lys 575	CAG Gln	GAG Glu	GTC Val	GAT Asp	AAT Asn	GGA Gly 580	ATG Met	CTT Leu	2323
TAC Tyr	CTG Leu 585	GAA Glu	TGG Trp	ATA Ile	AAC Asn	AAA Lys	AAA Lys 590	AAC Asn	ATC Ile	ACG Thr	AAT Asn 595	CTT Leu	TCC Ser	CGC Arg	TTA Leu	2371
GTT Val 600	GTC Val	CGC Arg	CCC Pro	CAC His	ACA Thr	GAT Asp 605	GCT Ala	GTT Val	TAC Tyr	CAT His	GCC Ala 610	TGC Cys	TTT Phe	TCT Ser	GAG Glu	2419
GAT Asp 615	GGT Gly	CAG Gln	AGA Arg	ATA Ile	GCT Ala 620	TCT Ser	TGT Cys	GGA Gly	GCT Ala	GAT Asp 625	AAA Lys	ACC Thr	TTA Leu	CAG Gln	GTG Val 630	2467
TTC Phe	AAA Lys	GCT Ala	GAA Glu 635	ACA Thr	GGA Gly	GAG Glu	AAA Lys	CTT Leu	CTA Leu 640	GAA Glu	ATC Ile	AAG Lys	GCT Ala 645	CAT His	GAG Glu	2515
GAT Asp	GAA Glu	GTG Val	CTT Leu 650	TGT Cys	TGT Cys	GCA Ala	TTC Phe	TCT Ser 655	ACA Thr	GAT Asp	GAC Asp	AGA Arg	TTT Phe 660	ATA Ile	GCA Ala	2563
ACC Thr	TGC Cys 665	TCA Ser	GTG Val	GAT Asp	AAA Lys	AAA Lys	GTG Val 670	AAG Lys	ATT Ile	TGG Trp	AAT Asn 675	TCT Ser	ATG Met	ACT Thr	GGG Gly	2611
GAA Glu 680	CTA Leu	GTA Val	CAC His	ACC Thr	TAT Tyr	GAT Asp 685	GAG Glu	CAC His	TCA Ser	GAG Glu	CAA Gln 690	GTC Val	AAT Asn	TGC Cys	TGC Cys	2659
CAT His 695	TTC Phe	ACC Thr	AAC Asn	AGT Ser	AGT Ser 700	CAT His	CAT His	CTT Leu	CTC Leu	TTA Leu 705	GCC Ala	ACT Thr	GGG Gly	TCA Ser	AGT Ser 710	2707

WO 98/55615

PCT/US98/11773

71

GAC TGC TTC CTC AAA CTT TGG GAT TTG AAT CAA AAA GAA TGT CGA AAT	2755
Asp Cys Phe Leu Lys Leu Trp Asp Leu Asn Gln Lys Glu Cys Arg Asn	
715 720 725	
ACC ATG TTT GGT CAT ACA AAT TCA GTC AAT CAC TGC AGA TTT TCA CCA	2803
Thr Met Phe Gly His Thr Asn Ser Val Asn His Cys Arg Phe Ser Pro	
730 735 740	
GAT GAT AAG CTT TTG GCT AGT TGT TCA GCT GAT GGA ACC TTA AAG CTT	2851
Asp Asp Lys Leu Leu Ala Ser Cys Ser Ala Asp Gly Thr Leu Lys Leu	
745 750 755	
TGG GAT GCG ACA TCA GCA AAT GAG AGG AAA AGC ATT AAT GTG AAA CAG	2899
Trp Asp Ala Thr Ser Ala Asn Glu Arg Lys Ser Ile Asn Val Lys Gln	
760 765 770	
TTC TTC CTA AAT TTG GAG GAC CCT CAA GAG GAT ATG GAA GTG ATA GTG	2947
Phe Phe Leu Asn Leu Glu Asp Pro Gln Glu Asp Met Glu Val Ile Val	
775 780 785 790	
AAG TGT TGT TCG TGG TCT GCT GAT GGT GCA AGG ATA ATG GTG GCA GCA	2995
Lys Cys Cys Ser Trp Ser Ala Asp Gly Ala Arg Ile Met Val Ala Ala	
795 800 805	
AAA AAT AAA ATC TTT TTG TGG AAT ACA GAC TCA CGT TCA AAG GTG GCT	3043
Lys Asn Lys Ile Phe Leu Trp Asn Thr Asp Ser Arg Ser Lys Val Ala	
810 815 820	
GAT TGC AGA GGA CAT TTA AGT TGG GTT CAT GGT GTG ATG TTT TCT CCT	3091
Asp Cys Arg Gly His Leu Ser Trp Val His Gly Val Met Phe Ser Pro	
825 830 835	
GAT GGA TCA TCA TTT TTG ACA TCT TCT GAT GAC CAG ACA ATC AGG CTC	3139
Asp Gly Ser Ser Phe Leu Thr Ser Ser Asp Asp Gln Thr Ile Arg Leu	
840 845 850	
TGG GAG ACA AAG AAA GTA TGT AAG AAC TCT GCT GTA ATG TTA AAG CAA	3187
Trp Glu Thr Lys Lys Val Cys Lys Asn Ser Ala Val Met Leu Lys Gln	
855 860 865 870	
GAA GTA GAT GTT GTG TTT CAA GAA AAT GAA GTG ATG GTC CTT GCA GTT	3235
Glu Val Asp Val Val Phe Gln Glu Asn Glu Val Met Val Leu Ala Val	
875 880 885	
GAC CAT ATA AGA CGT CTG CAA CTC ATT AAT GGA AGA ACA GGT CAG ATT	3283
Asp His Ile Arg Arg Leu Gln Leu Ile Asn Gly Arg Thr Gly Gln Ile	
890 895 900	
GAT TAT CTG ACT GAA GCT CAA GTT AGC TGC TGT TGC TTA AGT CCA CAT	3331
Asp Tyr Leu Thr Glu Ala Gln Val Ser Cys Cys Cys Leu Ser Pro His	
905 910 915	
CTT CAG TAC ATT GCA TTT GGA GAT GAA AAT GGA GCC ATT GAG ATT TTA	3379
Leu Gln Tyr Ile Ala Phe Gly Asp Glu Asn Gly Ala Ile Glu Ile Leu	
920 925 930	
GAA CTT GTA AAC AAT AGA ATC TTC CAG TCC AGG TTT CAG CAC AAG AAA	3427
Glu Leu Val Asn Asn Arg Ile Phe Gln Ser Arg Phe Gln His Lys Lys	
935 940 945 950	
ACT GTA TGG CAC ATC CAG TTC ACA GCC GAT GAG AAG ACT CTT ATT TCA	3475
Thr Val Trp His Ile Gln Phe Thr Ala Asp Glu Lys Thr Leu Ile Ser	
955 960 965	

WO 98/55615

PCT/US98/11773

72

AGT TCT GAT GAT GCT GAA ATT CAG GTA TGG AAT TGG CAA TTG GAC AAA	3523
Ser Ser Asp Asp Ala Glu Ile Gln Val Trp Asn Trp Gln Leu Asp Lys	
970 975 980	
TGT ATC TTT CTA CGA GGC CAT CAG GAA ACA GTG AAA GAC TTT AGA CTC	3571
Cys Ile Phe Leu Arg Gly His Gln Glu Thr Val Lys Asp Phe Arg Leu	
985 990 995	
TTG AAA AAT TCA AGA CTG CTT TCT TGG TCA TTT GAT GGA ACA GTG AAG	3619
Leu Lys Asn Ser Arg Leu Leu Ser Trp Ser Phe Asp Gly Thr Val Lys	
1000 1005 1010	
GTA TGG AAT ATT ATT ACT GGA AAT AAA GAA AAA GAC TTT GTC TGT CAC	3667
Val Trp Asn Ile Ile Thr Gly Asn Lys Glu Lys Asp Phe Val Cys His	
1015 1020 1025 1030	
CAG GGT ACA GTA CTT TCT TGT GAC ATT TCT CAC GAT GCT ACC AAG TTT	3715
Gln Gly Thr Val Leu Ser Cys Asp Ile Ser His Asp Ala Thr Lys Phe	
1035 1040 1045	
TCA TCT ACC TCT GCT GAC AAG ACT GCA AAG ATC TGG AGT TTT GAT CTC	3763
Ser Ser Thr Ser Ala Asp Lys Thr Ala Lys Ile Trp Ser Phe Asp Leu	
1050 1055 1060	
CTT TTG CCA CTT CAT GAA TTG AGG GGC CAC AAC GGC TGT GTG CGC TGC	3811
Leu Leu Pro Leu His Glu Leu Arg Gly His Asn Gly Cys Val Arg Cys	
1065 1070 1075	
TCT GCC TTC TCT GTG GAC AGT ACC CTG CTG GCA ACG GGA GAT GAC AAT	3859
Ser Ala Phe Ser Val Asp Ser Thr Leu Leu Ala Thr Gly Asp Asp Asn	
1080 1085 1090	
GGA GAA ATC AGG ATA TGG AAT GTC TCA AAC GGT GAG CTT CTT CAT TTG	3907
Gly Glu Ile Arg Ile Trp Asn Val Ser Asn Gly Glu Leu Leu His Leu	
1095 1100 1105 1110	
TGT GCT CCG CTT TCA GAA GAA GGA GCT GCT ACC CAT GGA GGC TGG GTG	3955
Cys Ala Pro Leu Ser Glu Glu Gly Ala Ala Thr His Gly Gly Trp Val	
1115 1120 1125	
ACT GAC CTT TGC TTT TCT CCA GAT GGC AAA ATG CTT ATC TCT GCT GGA	4003
Thr Asp Leu Cys Phe Ser Pro Asp Gly Lys Met Leu Ile Ser Ala Gly	
1130 1135 1140	
GGA TAT ATT AAG TGG TGG AAC GTT GTC ACT GGG GAA TCC TCA CAG ACC	4051
Gly Tyr Ile Lys Trp Trp Asn Val Val Thr Gly Glu Ser Ser Gln Thr	
1145 1150 1155	
TTC TAC ACA AAT GGA ACC AAT CTT AAG AAA ATA CAC GTG TCC CCT GAC	4099
Phe Tyr Thr Asn Gly Thr Asn Leu Lys Lys Ile His Val Ser Pro Asp	
1160 1165 1170	
TTC AAA ACA TAT GTG ACT GTG GAT AAT CTT GGT ATT TTA TAT ATT TTA	4147
Phe Lys Thr Tyr Val Thr Val Asp Asn Leu Gly Ile Leu Tyr Ile Leu	
1175 1180 1185 1190	
CAG ACT TTA GAA TAAAATAGTT AAGCATTAAT GTAGTTGAAC TTTTAAATT TTTGA	4204
Gln Thr Leu Glu	

WO 98/55615

PCT/US98/11773

73

ATTGGAAAAA	AATTCTAATG	AAACCCTGAT	ATCAACTTTT	TATAAAGCTC	TTAATTGTTG	4264
TGCAGTATTG	CATTCATTAC	AAAAGTGTTT	GTGGTTGGAT	GAATAATATT	AATGTAGCTT	4324
TTTCCCAAAT	GAACATACCT	TTAATCTTGT	TTTTCATGAT	CATCATTAAAC	AGTTTGTCTT	4384
TAGGATGCAA	ATGAAAATGT	GAATACATAC	CTTGTTGTAC	TGTTGGTAAA	ATTCTGTCTT	4444
GATGCATTCA	AAATGGTTGA	CATAATTAAT	GAGAAGAATT	TGGAAGAAAT	TGGTATTTTA	4504
ATACTGTCTG	TATTTATTAC	TGTTATGCAG	GCTGTGCCTC	AGGGTAGCAG	TGGCCTGCTT	4564
TTTGAACCAC	ACTTACCCCA	AGGGGGTTTT	GTTCTCCTAA	ATACAATCTT	AGAGGTTTTT	4624
TGCACTCTTT	AAATTTGCTT	TAAAAATATT	GTGTCTGTGT	GCATAGTCTG	CAGCATTTCC	4684
TTTAATTGAC	TCAATAAGTG	AGTCTTGGAT	TTAGCAGGCC	CCCCCACCTT	TTTTTTTTGT	4744
TTTTGGAGAC	AGAGTCTTGC	TTTGTTGCCA	GGCTGGAGTG	CAGTGGCGCG	ATCTCGGCTC	4804
ACCACAATCG	CTGCCTCCTG	GGTTCAAGCA	ATTCTCCTGC	CTCAGCCTCC	CGAGTAGCTG	4864
GGACTACAGG	TGTGCGCACA	TGCCAGGCTA	ATTTTTGTAT	TTTTAGTAGA	GACGGGGTTT	4924
CACCATGTTG	GCCGGGATGG	TCTCGATCTC	TTGACCTCAT	GATCTACCCG	CCTTGGCCTC	4984
CCAAAGTGCT	GAGATTACAG	GCGTGAGCCA	CCGTGCCTGG	CCAGGCCCTT	TCTCTTTTAA	5044
TGGAGACAGG	GTCTTGCACT	ATCACCCAGG	CTGGAGTGCA	GTGGCATAAT	CATACCTCAT	5104
TGCAGCCTCA	GACTCCTGGG	TTCAAGCAAT	CCTCCTGCCT	CAGCCTCCCA	AGTAGCTGAG	5164
ACTGCAGGCA	CGAGCCACCA	CACCCAGCTA	ATTTTTAAGT	TTTCTTGTAG	AGACAGGGTC	5224
TCACTATGTT	GTCTAGGCTG	GTCTTGAACT	CTTGGCCTCA	AGTAATCCTC	CTGCCTCAGC	5284
CTCCCAAAGT	GTTGGGATTG	CAGATATGAG	CCACTGGCCT	GGCCTTCAGC	AGTTCTTTTT	5344
GTGAAGTAAA	ACTTGTATGT	TGGAAAGAGT	AGATTTTATT	GGTCTACCCT	TTTCTCACTG	5404
TAGCTGCTGG	CAGCCCTGTG	CCATATCTGG	ACTCTAGTTG	TCAGTATCTG	AGTTGGACAC	5464
TATTCCTGCT	CCCTCTTGTT	TCTTACATAT	CAGACTTCTT	ACTTGAATGA	AACCTGATCT	5524
TTCCTAATCC	TCACTTTTTT	CTTTTTTAAA	AAGCAGTTTC	TCCACTGCTA	AATGTTAGTC	5584
ATTGAGGTGG	GGCCAATTTT	AATCATAAGC	CTTAATAAGA	TTTTTCTAAG	AAATGTGAAA	5644
TAGAACAATT	TTCATCTAAT	TCCATTTACT	TTTAGATGAA	TGGCATTGTG	AATGCCATTC	5704
TTTTAATGAA	TTTCAAGAGA	ATTCTCTGGT	TTTCTGTGTA	ATTCCAGATG	AGTCACTGTA	5764
ACTCTAGAAG	ATTAACCTTC	CAGCCAACCT	ATTTTCCTTT	CCCTTGCTCT	TCTCATCCTC	5824
TTTTCTTCC	TTCTTTCCTT	TCTCTTCTTT	TATCTCCAAG	GTTAATCAGG	AAAAATAGCT	5884
TTTGACAGGG	GAAAAAACTC	AATAACTAGC	TATTTTTGAC	CTCCTGATCA	GGAACTTTAG	5944
TTGAAGCGTA	AATCTAAAGA	AACATTTTCT	CTGAAATATA	TTATTAAGGG	CAATGGAGAT	6004
AAATTAATAG	TAGATGTGGT	TCCCAGAAAA	TATAATCAAA	ATTCAAAGAT	TTTTTTTTGTT	6064
TCTGTAACTG	GAACTAAATC	AAATGATTAC	TAGTGTTAAT	AGTAGATAAC	TTGTTTTTTAT	6124
TGTTGGTGCA	TATTAGTATA	ACTGTGGGGT	AGGTCGGGGA	GAGGGTAAGG	GAATAGATCA	6184
CTCAGATGTA	TTTTAGATAA	GCTATTTAGC	CTTTGATGGA	ATCATAAATA	CAGTGAATAC	6244
AATCCTTTGC	ATTGTTAAGG	AGGTTTTTTG	TTTTTAAATG	GTGGGTCAAG	GAGCTAGTTT	6304
ACAGGCTTAC	TGTGATTTAA	GCAAATGTGA	AAAGTGAAAC	CTTAATTTTA	TCAAAAGAAA	6364
TTTCTGTAAA	TGGTATGTCT	CCTTAGAATA	CCCAAATCAT	AATTTTATTT	GTACACACTG	6424
TTAGGGGCTC	ATCTCATGTA	GGCAGAGTAT	AAAGTATTAC	CTTTTGGAAT	TAAAAGCCAC	6484
TGACTGTTAT	AAAGTATAAC	AACACACATC	AGGTTTTTAA	AAGCCTTGAA	TGGCCCTTGT	6544
CTTAAAAAGA	AATTAGGAGC	CAGGTGCGGT	GGCACGTGCC	TGTAGTCCCA	GCTCCTTGGG	6604
AGGCTGAGAC	AGGAGGATTC	CTTGAGCCCT	GGAGTTTGAG	TCCAGCCTGG	GTGACATAGC	6664
AAGACCCTGT	CTTAAAAGAA	AAATGGGAAG	AAAGACAAGG	TAACATGAAG	AAAGAAGAGA	6724
TACCTAGTAT	GATGGAGCTG	CAAATTTTCAT	GGCAGTTCAT	GCAGTCGGTC	AAGAGGAGGA	6784
TTTTGTTTTG	TAGTTTGCAG	ATGAGCATTT	CTAAAGCATT	TTCCCTTGCT	GTATTTTTTT	6844
GTATTATAAA	TTACATTGGA	CTTCATATAT	ATAATTTTTT	TTTACATTAT	ATGTCTCTTG	6904
TATGTTTTGA	AACTCTTGTA	TTTATGATAT	AGCTTATATG	ATTTTTTTGC	CTTGGTATAC	6964
ATTTTAAAAT	ATGAATTTAA	AAAATTTTGT	TAAAAATAAA	ATTCACAAAA	TTGTTTTGAA	7024
AAACAAAAAA	AAAAAAA					7042

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1194 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

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

WO 98/55615

PCT/US98/11773

74

Met	Asp	Ala	Lys	Ala	Arg	Asn	Cys	Leu	Leu	Gln	His	Arg	Glu	Ala	Leu
1				5					10					15	
Glu	Lys	Asp	Ile	Lys	Thr	Ser	Tyr	Ile	Met	Asp	His	Met	Ile	Ser	Asp
			20					25					30		
Gly	Phe	Leu	Thr	Ile	Ser	Glu	Glu	Glu	Lys	Val	Arg	Asn	Glu	Pro	Thr
		35					40					45			
Gln	Gln	Gln	Arg	Ala	Ala	Met	Leu	Ile	Lys	Met	Ile	Leu	Lys	Lys	Asp
	50					55					60				
Asn	Asp	Ser	Tyr	Val	Ser	Phe	Tyr	Asn	Ala	Leu	His	Glu	Gly	Tyr	
65				70					75					80	
Lys	Asp	Leu	Ala	Ala	Leu	Leu	His	Asp	Gly	Ile	Pro	Val	Val	Ser	Ser
			85						90					95	
Ser	Ser	Val	Arg	Thr	Val	Leu	Cys	Glu	Gly	Gly	Val	Pro	Gln	Arg	Pro
			100					105					110		
Val	Val	Phe	Val	Thr	Arg	Lys	Lys	Leu	Val	Asn	Ala	Ile	Gln	Gln	Lys
		115					120					125			
Leu	Ser	Lys	Leu	Lys	Gly	Glu	Pro	Gly	Trp	Val	Thr	Ile	His	Gly	Met
	130				135						140				
Ala	Gly	Cys	Gly	Lys	Ser	Val	Leu	Ala	Ala	Glu	Ala	Val	Arg	Asp	His
145					150					155					160
Ser	Leu	Leu	Glu	Gly	Cys	Phe	Pro	Gly	Gly	Val	His	Trp	Val	Ser	Val
			165						170					175	
Gly	Lys	Gln	Asp	Lys	Ser	Gly	Leu	Leu	Met	Lys	Leu	Gln	Asn	Leu	Cys
		180						185					190		
Thr	Arg	Leu	Asp	Gln	Asp	Glu	Ser	Phe	Ser	Gln	Arg	Leu	Pro	Leu	Asn
		195					200					205			
Ile	Glu	Glu	Ala	Lys	Asp	Arg	Leu	Arg	Ile	Leu	Met	Leu	Arg	Lys	His
	210					215					220				
Pro	Arg	Ser	Leu	Leu	Ile	Leu	Asp	Asp	Val	Trp	Asp	Ser	Trp	Val	Leu
225					230					235					240
Lys	Ala	Phe	Asp	Ser	Gln	Cys	Gln	Ile	Leu	Leu	Thr	Thr	Arg	Asp	Lys
			245						250					255	
Ser	Val	Thr	Asp	Ser	Val	Met	Gly	Pro	Lys	Tyr	Val	Val	Pro	Val	Glu
		260						265					270		
Ser	Ser	Leu	Gly	Lys	Glu	Lys	Gly	Leu	Glu	Ile	Leu	Ser	Leu	Phe	Val
		275					280					285			
Asn	Met	Lys	Lys	Ala	Asp	Leu	Pro	Glu	Gln	Ala	His	Ser	Ile	Ile	Lys
	290					295					300				
Glu	Cys	Lys	Gly	Ser	Pro	Leu	Val	Val	Ser	Leu	Ile	Gly	Ala	Leu	Leu
305					310					315					320
Arg	Asp	Phe	Pro	Asn	Arg	Trp	Glu	Tyr	Tyr	Leu	Lys	Gln	Leu	Gln	Asn
			325						330					335	
Lys	Gln	Phe	Lys	Arg	Ile	Arg	Lys	Ser	Ser	Ser	Tyr	Asp	Tyr	Glu	Ala
		340					345					350			
Leu	Asp	Glu	Ala	Met	Ser	Ile	Ser	Val	Glu	Met	Leu	Arg	Glu	Asp	Ile
		355					360					365			
Lys	Asp	Tyr	Tyr	Thr	Asp	Leu	Ser	Ile	Leu	Gln	Lys	Asp	Val	Lys	Val
	370					375					380				
Pro	Thr	Lys	Val	Leu	Cys	Ile	Leu	Trp	Asp	Met	Glu	Thr	Glu	Glu	Val
385					390					395					400
Glu	Asp	Ile	Leu	Gln	Glu	Phe	Val	Asn	Lys	Ser	Leu	Leu	Phe	Cys	Asp
			405					410						415	
Arg	Asn	Gly	Lys	Ser	Phe	Arg	Tyr	Tyr	Leu	His	Asp	Leu	Gln	Val	Asp
		420					425						430		
Phe	Leu	Thr	Glu	Lys	Asn	Cys	Ser	Gln	Leu	Gln	Asp	Leu	His	Lys	Lys
		435					440					445			
Ile	Ile	Thr	Gln	Phe	Gln	Arg	Tyr	His	Gln	Pro	His	Thr	Leu	Ser	Pro
	450					455					460				
Asp	Gln	Glu	Asp	Cys	Met	Tyr	Trp	Tyr	Asn	Phe	Leu	Ala	Tyr	His	Met
465					470					475					480
Ala	Ser	Ala	Lys	Met	His	Lys	Glu	Leu	Cys	Ala	Leu	Met	Phe	Ser	Leu
			485						490					495	
Asp	Trp	Ile	Lys	Ala	Lys	Thr	Glu	Leu	Val	Gly	Pro	Ala	His	Leu	Ile
			500					505					510		

WO 98/55615

PCT/US98/11773

75

His	Glu	Phe	Val	Glu	Tyr	Arg	His	Ile	Leu	Asp	Glu	Lys	Asp	Cys	Ala
		515					520					525			
Val	Ser	Glu	Asn	Phe	Gln	Glu	Phe	Leu	Ser	Leu	Asn	Gly	His	Leu	Leu
	530					535					540				
Gly	Arg	Gln	Pro	Phe	Pro	Asn	Ile	Val	Gln	Leu	Gly	Leu	Cys	Glu	Pro
545					550					555					560
Glu	Thr	Ser	Glu	Val	Tyr	Gln	Gln	Ala	Lys	Leu	Gln	Ala	Lys	Gln	Glu
			565						570					575	
Val	Asp	Asn	Gly	Met	Leu	Tyr	Leu	Glu	Trp	Ile	Asn	Lys	Lys	Asn	Ile
			580					585					590		
Thr	Asn	Leu	Ser	Arg	Leu	Val	Val	Arg	Pro	His	Thr	Asp	Ala	Val	Tyr
	595						600					605			
His	Ala	Cys	Phe	Ser	Glu	Asp	Gly	Gln	Arg	Ile	Ala	Ser	Cys	Gly	Ala
	610					615					620				
Asp	Lys	Thr	Leu	Gln	Val	Phe	Lys	Ala	Glu	Thr	Gly	Glu	Lys	Leu	Leu
625					630					635					640
Glu	Ile	Lys	Ala	His	Glu	Asp	Glu	Val	Leu	Cys	Cys	Ala	Phe	Ser	Thr
			645						650					655	
Asp	Asp	Arg	Phe	Ile	Ala	Thr	Cys	Ser	Val	Asp	Lys	Lys	Val	Lys	Ile
			660					665					670		
Trp	Asn	Ser	Met	Thr	Gly	Glu	Leu	Val	His	Thr	Tyr	Asp	Glu	His	Ser
	675						680					685			
Glu	Gln	Val	Asn	Cys	Cys	His	Phe	Thr	Asn	Ser	Ser	His	His	Leu	Leu
	690					695					700				
Leu	Ala	Thr	Gly	Ser	Ser	Asp	Cys	Phe	Leu	Lys	Leu	Trp	Asp	Leu	Asn
705					710					715					720
Gln	Lys	Glu	Cys	Arg	Asn	Thr	Met	Phe	Gly	His	Thr	Asn	Ser	Val	Asn
			725						730					735	
His	Cys	Arg	Phe	Ser	Pro	Asp	Asp	Lys	Leu	Leu	Ala	Ser	Cys	Ser	Ala
			740					745					750		
Asp	Gly	Thr	Leu	Lys	Leu	Trp	Asp	Ala	Thr	Ser	Ala	Asn	Glu	Arg	Lys
	755						760					765			
Ser	Ile	Asn	Val	Lys	Gln	Phe	Phe	Leu	Asn	Leu	Glu	Asp	Pro	Gln	Glu
	770					775					780				
Asp	Met	Glu	Val	Ile	Val	Lys	Cys	Cys	Ser	Trp	Ser	Ala	Asp	Gly	Ala
785					790					795					800
Arg	Ile	Met	Val	Ala	Ala	Lys	Asn	Lys	Ile	Phe	Leu	Trp	Asn	Thr	Asp
			805						810					815	
Ser	Arg	Ser	Lys	Val	Ala	Asp	Cys	Arg	Gly	His	Leu	Ser	Trp	Val	His
			820					825					830		
Gly	Val	Met	Phe	Ser	Pro	Asp	Gly	Ser	Ser	Phe	Leu	Thr	Ser	Ser	Asp
	835						840					845			
Asp	Gln	Thr	Ile	Arg	Leu	Trp	Glu	Thr	Lys	Lys	Val	Cys	Lys	Asn	Ser
	850					855					860				
Ala	Val	Met	Leu	Lys	Gln	Glu	Val	Asp	Val	Val	Phe	Gln	Glu	Asn	Glu
865					870					875					880
Val	Met	Val	Leu	Ala	Val	Asp	His	Ile	Arg	Arg	Leu	Gln	Leu	Ile	Asn
			885					890						895	
Gly	Arg	Thr	Gly	Gln	Ile	Asp	Tyr	Leu	Thr	Glu	Ala	Gln	Val	Ser	Cys
		900						905					910		
Cys	Cys	Leu	Ser	Pro	His	Leu	Gln	Tyr	Ile	Ala	Phe	Gly	Asp	Glu	Asn
		915					920					925			
Gly	Ala	Ile	Glu	Ile	Leu	Glu	Leu	Val	Asn	Asn	Arg	Ile	Phe	Gln	Ser
	930					935					940				
Arg	Phe	Gln	His	Lys	Lys	Thr	Val	Trp	His	Ile	Gln	Phe	Thr	Ala	Asp
945					950					955					960
Glu	Lys	Thr	Leu	Ile	Ser	Ser	Ser	Asp	Asp	Ala	Glu	Ile	Gln	Val	Trp
			965						970					975	
Asn	Trp	Gln	Leu	Asp	Lys	Cys	Ile	Phe	Leu	Arg	Gly	His	Gln	Glu	Thr
		980						985					990		
Val	Lys	Asp	Phe	Arg	Leu	Leu	Lys	Asn	Ser	Arg	Leu	Leu	Ser	Trp	Ser
	995						1000					1005			
Phe	Asp	Gly	Thr	Val	Lys	Val	Trp	Asn	Ile	Ile	Thr	Gly	Asn	Lys	Glu
1010						1015					1020				

WO 98/55615

PCT/US98/11773

76

Lys Asp Phe Val Cys His Gln Gly Thr Val Leu Ser Cys Asp Ile Ser
 025 1030 1035 1040
 His Asp Ala Thr Lys Phe Ser Ser Thr Ser Ala Asp Lys Thr Ala Lys
 1045 1050 1055
 Ile Trp Ser Phe Asp Leu Leu Leu Pro Leu His Glu Leu Arg Gly His
 1060 1065 1070
 Asn Gly Cys Val Arg Cys Ser Ala Phe Ser Val Asp Ser Thr Leu Leu
 1075 1080 1085
 Ala Thr Gly Asp Asp Asn Gly Glu Ile Arg Ile Trp Asn Val Ser Asn
 1090 1095 1100
 Gly Glu Leu Leu His Leu Cys Ala Pro Leu Ser Glu Glu Gly Ala Ala
 1105 1110 1115 1120
 Thr His Gly Gly Trp Val Thr Asp Leu Cys Phe Ser Pro Asp Gly Lys
 1125 1130 1135
 Met Leu Ile Ser Ala Gly Gly Tyr Ile Lys Trp Trp Asn Val Val Thr
 1140 1145 1150
 Gly Glu Ser Ser Gln Thr Phe Tyr Thr Asn Gly Thr Asn Leu Lys Lys
 1155 1160 1165
 Ile His Val Ser Pro Asp Phe Lys Thr Tyr Val Thr Val Asp Asn Leu
 1170 1175 1180
 Gly Ile Leu Tyr Ile Leu Gln Thr Leu Glu
 185 1190 1

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

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

AARGTNMGNA AYGARCCNAC

20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

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

ATTTAGGTGA CACTATAGAA

20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

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

AAYGARCCNA CNCAACAACA

20

WO 98/55615

PCT/US98/11773

77

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

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

TTYTGyTGDA TNGCRtTCAC

20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACATCAGAA TCTTTCCCGC

20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AACACTTCAC TATCACTTCC

20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACATCAGAA TCTTTCCCGC

20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

WO 98/55615

PCT/US98/11773

78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CAACACTTCA CTATCACTTC C

21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCAAAAGCTC GAAATCATAT GCTTCAACAT AGAG

34

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCGCGGCCGC CTCGAGGGCT CTGGTTGTAA G

31

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TAATGATTCC TACGTATCAT TCTACAATGC

30

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAATGATCTC TAACAGCTTC

20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7075 base pairs

WO 98/55615

PCT/US98/11773

79

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
 (ix) FEATURE:

(A) NAME/KEY: Coding Sequence
 (B) LOCATION: 578...4192
 (D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAGAAGAGGT	AGCGAGTGG	CGTGACTGCT	CTATCCCGGG	CAAAAGGGAT	AGAACCAGAG	60
GTGGGGAGTC	TGGGCAGTCG	GCGACCCGCG	AAGACTTGAG	GTGCCGCAGC	GGCATCCGGA	120
GTAGCGCCGG	GCTCCCTCCG	GGGTGCAGCC	GCCGTCGGGG	GAAGGGCGCC	ACAGGCCGGG	180
AAGACCTCCT	CCCTTTGTGT	CCAGTAGTGG	GGTCCACCGG	AGGGCGGCCC	GTGGGCCGGG	240
CCTCACCGCG	GCGCTCCGGG	ACTGTGGGGT	CAGGCTGCGT	TGGGTGGACG	CCCACCTCGC	300
CAACCTTCGG	AGGTCCCTGG	GGGTCTTCGT	GCGCCCCGGG	GCTGCAGAGA	TCCAGGGGAG	360
GCGCCTGTGA	GGCCCCGGAC	TGCCCCGGGG	CGAAGGGTAT	GTGGCGAGAC	AGAGCCCTGC	420
ACCCCTAATT	CCCGGTGGAA	AACTCCTGTT	GCCGTTTCCC	TCCACCGGCC	TGGAGTCTCC	480
CAGTCTTGTC	CCGGCAGTGC	CGCCCTCCCC	ACTAAGACCT	AGGCGCAAAG	GCTTGGCTCA	540
TGGTTGACAG	CTCAGAGAGA	GAAAGATCTG	AGGGAAG	ATG GAT GCA	AAA GCT CGA	595
				Met Asp Ala Lys Ala Arg		
				1	5	
AAT TGT TTG CTT CAA CAT AGA GAA GCT CTG GAA AAG GAC ATC AAG ACA						643
Asn Cys Leu Leu Gln His Arg Glu Ala Leu Glu Lys Asp Ile Lys Thr						
	10		15		20	
TCC TAC ATC ATG GAT CAC ATG ATT AGT GAT GGA TTT TTA ACA ATA TCA						691
Ser Tyr Ile Met Asp His Met Ile Ser Asp Gly Phe Leu Thr Ile Ser						
	25		30		35	
GAA GAG GAA AAA GTA AGA AAT GAG CCC ACT CAA CAG CAA AGA GCA GCT						739
Glu Glu Glu Lys Val Arg Asn Glu Pro Thr Gln Gln Gln Arg Ala Ala						
	40		45		50	
ATG CTG ATT AAA ATG ATA CTT AAA AAA GAT AAT GAT TCC TAC GTA TCA						787
Met Leu Ile Lys Met Ile Leu Lys Lys Asp Asn Asp Ser Tyr Val Ser						
	55		60		65	70
TTC TAC AAT GCT CTA CTA CAT GAA GGA TAT AAA GAT CTT GCT GCC CTT						835
Phe Tyr Asn Ala Leu Leu His Glu Gly Tyr Lys Asp Leu Ala Ala Leu						
	75		80		85	
CTC CAT GAT GGC ATT CCT GTT GTC TCT TCT TCC AGT GGT AAA GAT TCA						883
Leu His Asp Gly Ile Pro Val Val Ser Ser Ser Ser Gly Lys Asp Ser						
	90		95		100	
GTT AGT GGA ATA ACT TCG TAT GTA AGG ACA GTC CTG TGT GAA GGT GGA						931
Val Ser Gly Ile Thr Ser Tyr Val Arg Thr Val Leu Cys Glu Gly Gly						
	105		110		115	
GTA CCA CAG AGG CCA GTT GTT TTT GTC ACA AGG AAG AAG CTG GTG AAT						979
Val Pro Gln Arg Pro Val Val Phe Val Thr Arg Lys Lys Leu Val Asn						
	120		125		130	
GCA ATT CAG CAG AAG CTC TCC AAA TTG AAA GGT GAA CCA GGA TGG GTC						1027
Ala Ile Gln Gln Lys Leu Ser Lys Leu Lys Gly Glu Pro Gly Trp Val						
	135		140		145	150
ACC ATA CAT GGA ATG GCA GGC TGT GGG AAG TCT GTA TTA GCT GCA GAA						1075
Thr Ile His Gly Met Ala Gly Cys Gly Lys Ser Val Leu Ala Ala Glu						
	155		160		165	

WO 98/55615

PCT/US98/11773

80

GCT	GTT	AGA	GAT	CAT	TCC	CTT	TTA	GAA	GGT	TGT	TTC	CCA	GGG	GGA	GTG	1123
Ala	Val	Arg	Asp	His	Ser	Leu	Leu	Glu	Gly	Cys	Phe	Pro	Gly	Gly	Val	
			170					175					180			
CAT	TGG	GTT	TCA	GTT	GGG	AAA	CAA	GAC	AAA	TCT	GGG	CTT	CTG	ATG	AAA	1171
His	Trp	Val	Ser	Val	Gly	Lys	Gln	Asp	Lys	Ser	Gly	Leu	Leu	Met	Lys	
		185					190					195				
CTG	CAG	AAT	CTT	TGC	ACA	CGG	TTG	GAT	CAG	GAT	GAG	AGT	TTT	TCC	CAG	1219
Leu	Gln	Asn	Leu	Cys	Thr	Arg	Leu	Asp	Gln	Asp	Glu	Ser	Phe	Ser	Gln	
	200					205					210					
AGG	CTT	CCA	CTT	AAT	ATT	GAA	GAG	GCT	AAA	GAC	CGT	CTC	CGC	ATT	CTG	1267
Arg	Leu	Pro	Leu	Asn	Ile	Glu	Glu	Ala	Lys	Asp	Arg	Leu	Arg	Ile	Leu	
215					220					225					230	
ATG	CTT	CGC	AAA	CAC	CCA	AGG	TCT	CTC	TTG	ATC	TTG	GAT	GAT	GTT	TGG	1315
Met	Leu	Arg	Lys	His	Pro	Arg	Ser	Leu	Leu	Ile	Leu	Asp	Asp	Val	Trp	
				235					240					245		
GAC	TCT	TGG	GTG	TTG	AAA	GCT	TTT	GAC	AGT	CAG	TGT	CAG	ATT	CTT	CTT	1363
Asp	Ser	Trp	Val	Leu	Lys	Ala	Phe	Asp	Ser	Gln	Cys	Gln	Ile	Leu	Leu	
			250					255					260			
ACA	ACC	AGA	GAC	AAG	AGT	GTT	ACA	GAT	TCA	GTA	ATG	GGT	CCT	AAA	TAT	1411
Thr	Thr	Arg	Asp	Lys	Ser	Val	Thr	Asp	Ser	Val	Met	Gly	Pro	Lys	Tyr	
		265					270					275				
GTA	GTC	CCT	GTG	GAG	AGT	TCC	TTA	GGA	AAG	GAA	AAA	GGA	CTT	GAA	ATT	1459
Val	Val	Pro	Val	Glu	Ser	Ser	Leu	Gly	Lys	Glu	Lys	Gly	Leu	Glu	Ile	
	280					285					290					
TTA	TCC	CTT	TTT	GTT	AAT	ATG	AAG	AAG	GCA	GAT	TTG	CCA	GAA	CAA	GCT	1507
Leu	Ser	Leu	Phe	Val	Asn	Met	Lys	Lys	Ala	Asp	Leu	Pro	Glu	Gln	Ala	
295					300					305					310	
CAT	AGT	ATT	ATA	AAA	GAA	TGT	AAA	GGC	TCT	CCC	CTT	GTA	GTA	TCT	TTA	1555
His	Ser	Ile	Ile	Lys	Glu	Cys	Lys	Gly	Ser	Pro	Leu	Val	Val	Ser	Leu	
				315				320						325		
ATT	GGT	GCA	CTT	TTA	CGT	GAT	TTT	CCC	AAT	CGC	TGG	GAG	TAC	TAC	CTC	1603
Ile	Gly	Ala	Leu	Leu	Arg	Asp	Phe	Pro	Asn	Arg	Trp	Glu	Tyr	Tyr	Leu	
			330					335					340			
AAA	CAG	CTT	CAG	AAT	AAG	CAG	TTT	AAG	AGA	ATA	AGG	AAA	TCT	TCG	TCT	1651
Lys	Gln	Leu	Gln	Asn	Lys	Gln	Phe	Lys	Arg	Ile	Arg	Lys	Ser	Ser	Ser	
		345					350					355				
TAT	GAT	TAT	GAG	GCT	CTA	GAT	GAA	GCC	ATG	TCT	ATA	AGT	GTT	GAA	ATG	1699
Tyr	Asp	Tyr	Glu	Ala	Leu	Asp	Glu	Ala	Met	Ser	Ile	Ser	Val	Glu	Met	
	360					365					370					
CTC	AGA	GAA	GAC	ATC	AAA	GAT	TAT	TAC	ACA	GAT	CTT	TCC	ATC	CTT	CAG	1747
Leu	Arg	Glu	Asp	Ile	Lys	Asp	Tyr	Tyr	Thr	Asp	Leu	Ser	Ile	Leu	Gln	
375					380					385					390	
AAG	GAC	GTT	AAG	GTG	CCT	ACA	AAG	GTG	TTA	TGT	ATT	CTC	TGG	GAC	ATG	1795
Lys	Asp	Val	Lys	Val	Pro	Thr	Lys	Val	Leu	Cys	Ile	Leu	Trp	Asp	Met	
				395				400						405		
GAA	ACT	GAA	GAA	GTT	GAA	GAC	ATA	CTG	CAG	GAG	TTT	GTA	AAT	AAG	TCT	1843
Glu	Thr	Glu	Glu	Val	Glu	Asp	Ile	Leu	Gln	Glu	Phe	Val	Asn	Lys	Ser	
			410					415					420			

WO 98/55615

PCT/US98/11773

81

CTT	TTA	TTC	TGT	GAT	CGG	AAT	GGA	AAG	TCG	TTT	CGT	TAT	TAT	TTA	CAT	1891
Leu	Leu	Phe	Cys	Asp	Arg	Asn	Gly	Lys	Ser	Phe	Arg	Tyr	Tyr	Leu	His	
		425					430					435				
GAT	CTT	CAA	GTA	GAT	TTT	CTT	ACA	GAG	AAG	AAT	TGC	AGC	CAG	CTT	CAG	1939
Asp	Leu	Gln	Val	Asp	Phe	Leu	Thr	Glu	Lys	Asn	Cys	Ser	Gln	Leu	Gln	
	440					445					450					
GAT	CTA	CAT	AAG	AAG	ATA	ATC	ACT	CAG	TTT	CAG	AGA	TAT	CAC	CAG	CCG	1987
Asp	Leu	His	Lys	Lys	Ile	Ile	Thr	Gln	Phe	Gln	Arg	Tyr	His	Gln	Pro	
455					460					465					470	
CAT	ACT	CTT	TCA	CCA	GAT	CAG	GAA	GAC	TGT	ATG	TAT	TGG	TAC	AAC	TTT	2035
His	Thr	Leu	Ser	Pro	Asp	Gln	Glu	Asp	Cys	Met	Tyr	Trp	Tyr	Asn	Phe	
				475					480					485		
CTG	GCC	TAT	CAC	ATG	GCC	AGT	GCC	AAG	ATG	CAC	AAG	GAA	CTT	TGT	GCT	2083
Leu	Ala	Tyr	His	Met	Ala	Ser	Ala	Lys	Met	His	Lys	Glu	Leu	Cys	Ala	
			490					495					500			
TTA	ATG	TTT	TCC	CTG	GAT	TGG	ATT	AAA	GCA	AAA	ACA	GAA	CTT	GTA	GGC	2131
Leu	Met	Phe	Ser	Leu	Asp	Trp	Ile	Lys	Ala	Lys	Thr	Glu	Leu	Val	Gly	
		505					510					515				
CCT	GCT	CAT	CTG	ATT	CAT	GAA	TTT	GTG	GAA	TAC	AGA	CAT	ATA	CTA	GAT	2179
Pro	Ala	His	Leu	Ile	His	Glu	Phe	Val	Glu	Tyr	Arg	His	Ile	Leu	Asp	
	520					525					530					
GAA	AAG	GAT	TGT	GCA	GTC	AGT	GAG	AAT	TTT	CAG	GAG	TTT	TTA	TCT	TTA	2227
Glu	Lys	Asp	Cys	Ala	Val	Ser	Glu	Asn	Phe	Gln	Glu	Phe	Leu	Ser	Leu	
535					540					545					550	
AAT	GGA	CAC	CTT	CTT	GGA	CGA	CAG	CCA	TTT	CCT	AAT	ATT	GTA	CAA	CTG	2275
Asn	Gly	His	Leu	Leu	Gly	Arg	Gln	Pro	Phe	Pro	Asn	Ile	Val	Gln	Leu	
				555					560					565		
GGT	CTC	TGT	GAG	CCG	GAA	ACT	TCA	GAA	GTT	TAT	CAG	CAA	GCT	AAG	CTG	2323
Gly	Leu	Cys	Glu	Pro	Glu	Thr	Ser	Glu	Val	Tyr	Gln	Gln	Ala	Lys	Leu	
			570					575					580			
CAG	GCC	AAG	CAG	GAG	GTC	GAT	AAT	GGA	ATG	CTT	TAC	CTG	GAA	TGG	ATA	2371
Gln	Ala	Lys	Gln	Glu	Val	Asp	Asn	Gly	Met	Leu	Tyr	Leu	Glu	Trp	Ile	
		585					590					595				
AAC	AAA	AAA	AAC	ATC	ACG	AAT	CTT	TCC	CGC	TTA	GTT	GTC	CGC	CCC	CAC	2419
Asn	Lys	Lys	Asn	Ile	Thr	Asn	Leu	Ser	Arg	Leu	Val	Val	Arg	Pro	His	
	600					605					610					
ACA	GAT	GCT	GTT	TAC	CAT	GCC	TGC	TTT	TCT	GAG	GAT	GGT	CAG	AGA	ATA	2467
Thr	Asp	Ala	Val	Tyr	His	Ala	Cys	Phe	Ser	Glu	Asp	Gly	Gln	Arg	Ile	
615					620					625					630	
GCT	TCT	TGT	GGA	GCT	GAT	AAA	ACC	TTA	CAG	GTG	TTC	AAA	GCT	GAA	ACA	2515
Ala	Ser	Cys	Gly	Ala	Asp	Lys	Thr	Leu	Gln	Val	Phe	Lys	Ala	Glu	Thr	
				635					640					645		
GGA	GAG	AAA	CTT	CTA	GAA	ATC	AAG	GCT	CAT	GAG	GAT	GAA	GTG	CTT	TGT	2563
Gly	Glu	Lys	Leu	Leu	Glu	Ile	Lys	Ala	His	Glu	Asp	Glu	Val	Leu	Cys	
			650					655					660			
TGT	GCA	TTC	TCT	ACA	GAT	GAC	AGA	TTT	ATA	GCA	ACC	TGC	TCA	GTG	GAT	2611
Cys	Ala	Phe	Ser	Thr	Asp	Asp	Arg	Phe	Ile	Ala	Thr	Cys	Ser	Val	Asp	
		665					670					675				

WO 98/55615

PCT/US98/11773

82

AAA	AAA	GTG	AAG	ATT	TGG	AAT	TCT	ATG	ACT	GGG	GAA	CTA	GTA	CAC	ACC	2659
Lys	Lys	Val	Lys	Ile	Trp	Asn	Ser	Met	Thr	Gly	Glu	Leu	Val	His	Thr	
	680					685					690					
TAT	GAT	GAG	CAC	TCA	GAG	CAA	GTC	AAT	TGC	TGC	CAT	TTC	ACC	AAC	AGT	2707
Tyr	Asp	Glu	His	Ser	Glu	Gln	Val	Asn	Cys	Cys	His	Phe	Thr	Asn	Ser	
	695				700					705					710	
AGT	CAT	CAT	CTT	CTC	TTA	GCC	ACT	GGG	TCA	AGT	GAC	TGC	TTC	CTC	AAA	2755
Ser	His	His	Leu	Leu	Leu	Ala	Thr	Gly	Ser	Ser	Asp	Cys	Phe	Leu	Lys	
				715					720					725		
CTT	TGG	GAT	TTG	AAT	CAA	AAA	GAA	TGT	CGA	AAT	ACC	ATG	TTT	GGT	CAT	2803
Leu	Trp	Asp	Leu	Asn	Gln	Lys	Glu	Cys	Arg	Asn	Thr	Met	Phe	Gly	His	
			730					735					740			
ACA	AAT	TCA	GTC	AAT	CAC	TGC	AGA	TTT	TCA	CCA	GAT	GAT	AAG	CTT	TTG	2851
Thr	Asn	Ser	Val	Asn	His	Cys	Arg	Phe	Ser	Pro	Asp	Asp	Lys	Leu	Leu	
		745					750					755				
GCT	AGT	TGT	TCA	GCT	GAT	GGA	ACC	TTA	AAG	CTT	TGG	GAT	GCG	ACA	TCA	2899
Ala	Ser	Cys	Ser	Ala	Asp	Gly	Thr	Leu	Lys	Leu	Trp	Asp	Ala	Thr	Ser	
	760					765					770					
GCA	AAT	GAG	AGG	AAA	AGC	ATT	AAT	GTG	AAA	CAG	TTC	TTC	CTA	AAT	TTG	2947
Ala	Asn	Glu	Arg	Lys	Ser	Ile	Asn	Val	Lys	Gln	Phe	Phe	Leu	Asn	Leu	
	775				780					785					790	
GAG	GAC	CCT	CAA	GAG	GAT	ATG	GAA	GTG	ATA	GTG	AAG	TGT	TGT	TCG	TGG	2995
Glu	Asp	Pro	Gln	Glu	Asp	Met	Glu	Val	Ile	Val	Lys	Cys	Cys	Ser	Trp	
				795					800					805		
TCT	GCT	GAT	GGT	GCA	AGG	ATA	ATG	GTG	GCA	GCA	AAA	AAT	AAA	ATC	TTT	3043
Ser	Ala	Asp	Gly	Ala	Arg	Ile	Met	Val	Ala	Ala	Lys	Asn	Lys	Ile	Phe	
			810					815					820			
TTG	TGG	AAT	ACA	GAC	TCA	CGT	TCA	AAG	GTG	GCT	GAT	TGC	AGA	GGA	CAT	3091
Leu	Trp	Asn	Thr	Asp	Ser	Arg	Ser	Lys	Val	Ala	Asp	Cys	Arg	Gly	His	
		825					830					835				
TTA	AGT	TGG	GTT	CAT	GGT	GTG	ATG	TTT	TCT	CCT	GAT	GGA	TCA	TCA	TTT	3139
Leu	Ser	Trp	Val	His	Gly	Val	Met	Phe	Ser	Pro	Asp	Gly	Ser	Ser	Phe	
	840					845					850					
TTG	ACA	TCT	TCT	GAT	GAC	CAG	ACA	ATC	AGG	CTC	TGG	GAG	ACA	AAG	AAA	3187
Leu	Thr	Ser	Ser	Asp	Asp	Gln	Thr	Ile	Arg	Leu	Trp	Glu	Thr	Lys	Lys	
	855				860					865					870	
GTA	TGT	AAG	AAC	TCT	GCT	GTA	ATG	TTA	AAG	CAA	GAA	GTA	GAT	GTT	GTG	3235
Val	Cys	Lys	Asn	Ser	Ala	Val	Met	Leu	Lys	Gln	Glu	Val	Asp	Val	Val	
				875					880					885		
TTT	CAA	GAA	AAT	GAA	GTG	ATG	GTC	CTT	GCA	GTT	GAC	CAT	ATA	AGA	CGT	3283
Phe	Gln	Glu	Asn	Glu	Val	Met	Val	Leu	Ala	Val	Asp	His	Ile	Arg	Arg	
			890					895					900			
CTG	CAA	CTC	ATT	AAT	GGA	AGA	ACA	GGT	CAG	ATT	GAT	TAT	CTG	ACT	GAA	3331
Leu	Gln	Leu	Ile	Asn	Gly	Arg	Thr	Gly	Gln	Ile	Asp	Tyr	Leu	Thr	Glu	
		905					910					915				
GCT	CAA	GTT	AGC	TGC	TGT	TGC	TTA	AGT	CCA	CAT	CTT	CAG	TAC	ATT	GCA	3379
Ala	Gln	Val	Ser	Cys	Cys	Cys	Leu	Ser	Pro	His	Leu	Gln	Tyr	Ile	Ala	
	920					925					930					

WO 98/55615

PCT/US98/11773

83

TTT GGA GAT GAA AAT GGA GCC ATT GAG ATT TTA GAA CTT GTA AAC AAT Phe Gly Asp Glu Asn Gly Ala Ile Glu Ile Leu Glu Leu Val Asn Asn 935 940 945 950	3427
AGA ATC TTC CAG TCC AGG TTT CAG CAC AAG AAA ACT GTA TGG CAC ATC Arg Ile Phe Gln Ser Arg Phe Gln His Lys Lys Thr Val Trp His Ile 955 960 965	3475
CAG TTC ACA GCC GAT GAG AAG ACT CTT ATT TCA AGT TCT GAT GAT GCT Gln Phe Thr Ala Asp Glu Lys Thr Leu Ile Ser Ser Ser Asp Asp Ala 970 975 980	3523
GAA ATT CAG GTA TGG AAT TGG CAA TTG GAC AAA TGT ATC TTT CTA CGA Glu Ile Gln Val Trp Asn Trp Gln Leu Asp Lys Cys Ile Phe Leu Arg 985 990 995	3571
GGC CAT CAG GAA ACA GTG AAA GAC TTT AGA CTC TTG AAA AAT TCA AGA Gly His Gln Glu Thr Val Lys Asp Phe Arg Leu Leu Lys Asn Ser Arg 1000 1005 1010	3619
CTG CTT TCT TGG TCA TTT GAT GGA ACA GTG AAG GTA TGG AAT ATT ATT Leu Leu Ser Trp Ser Phe Asp Gly Thr Val Lys Val Trp Asn Ile Ile 1015 1020 1025 1030	3667
ACT GGA AAT AAA GAA AAA GAC TTT GTC TGT CAC CAG GGT ACA GTA CTT Thr Gly Asn Lys Glu Lys Asp Phe Val Cys His Gln Gly Thr Val Leu 1035 1040 1045	3715
TCT TGT GAC ATT TCT CAC GAT GCT ACC AAG TTT TCA TCT ACC TCT GCT Ser Cys Asp Ile Ser His Asp Ala Thr Lys Phe Ser Ser Thr Ser Ala 1050 1055 1060	3763
GAC AAG ACT GCA AAG ATC TGG AGT TTT GAT CTC CTT TTG CCA CTT CAT Asp Lys Thr Ala Lys Ile Trp Ser Phe Asp Leu Leu Leu Pro Leu His 1065 1070 1075	3811
GAA TTG AGG GGC CAC AAC GGC TGT GTG CGC TGC TCT GCC TTC TCT GTG Glu Leu Arg Gly His Asn Gly Cys Val Arg Cys Ser Ala Phe Ser Val 1080 1085 1090	3859
GAC AGT ACC CTG CTG GCA ACG GGA GAT GAC AAT GGA GAA ATC AGG ATA Asp Ser Thr Leu Leu Ala Thr Gly Asp Asp Asn Gly Glu Ile Arg Ile 1095 1100 1105 1110	3907
TGG AAT GTC TCA AAC GGT GAG CTT CTT CAT TTG TGT GCT CCG CTT TCA Trp Asn Val Ser Asn Gly Glu Leu Leu His Leu Cys Ala Pro Leu Ser 1115 1120 1125	3955
GAA GAA GGA GCT GCT ACC CAT GGA GGC TGG GTG ACT GAC CTT TGC TTT Glu Glu Gly Ala Ala Thr His Gly Gly Trp Val Thr Asp Leu Cys Phe 1130 1135 1140	4003
TCT CCA GAT GGC AAA ATG CTT ATC TCT GCT GGA GGA TAT ATT AAG TGG Ser Pro Asp Gly Lys Met Leu Ile Ser Ala Gly Gly Tyr Ile Lys Trp 1145 1150 1155	4051
TGG AAC GTT GTC ACT GGG GAA TCC TCA CAG ACC TTC TAC ACA AAT GGA Trp Asn Val Val Thr Gly Glu Ser Ser Gln Thr Phe Tyr Thr Asn Gly 1160 1165 1170	4099
ACC AAT CTT AAG AAA ATA CAC GTG TCC CCT GAC TTC AAA ACA TAT GTG Thr Asn Leu Lys Lys Ile His Val Ser Pro Asp Phe Lys Thr Tyr Val 1175 1180 1185 1190	4147

WO 98/55615

PCT/US98/11773

84

ACT GTG GAT AAT CTT GGT ATT TTA TAT ATT TTA CAG ACT TTA GAA TAAAA 4197
 Thr Val Asp Asn Leu Gly Ile Leu Tyr Ile Leu Gln Thr Leu Glu
 1195 1200 1205

TAGTTAAGCA	TTAATGTAGT	TGAACTTTTT	AAATTTTTGA	ATTGGAAAAA	AATTCTAATG	4257
AAACCCTGAT	ATCAACTTTT	TATAAAGCTC	TTAATTGTTG	TGCAGTATTG	CATTCATTAC	4317
AAAAGTGTTT	GTGGTTGGAT	GAATAATATT	AATGTAGCTT	TTTCCCAAAT	GAACATACCT	4377
TTAATCTTGT	TTTTCATGAT	CATCATTAAAC	AGTTTGTCTT	TAGGATGCAA	ATGAAAATGT	4437
GAATACATAC	CTTGTTGTAC	TGTTGGTAAA	ATTCTGTCTT	GATGCATTCA	AAATGGTTGA	4497
CATAATTAAT	GAGAAGAATT	TGGAAGAAAT	TGGTATTTTA	ATACTGTCTG	TATTTATTAC	4557
TGTTATGCAG	GCTGTGCCTC	AGGGTAGCAG	TGGCCTGCTT	TTTGAACCAC	ACTTACCCCA	4617
AGGGGGTTTT	GTTCTCCTAA	ATACAATCTT	AGAGGTTTTT	TGCACTCTTT	AAATTTGCTT	4677
TAAAAATATT	GTGTCTGTGT	GCATAGTCTG	CAGCATTTCC	TTTAATTGAC	TCAATAAGTG	4737
AGTCTTGGAT	TTAGCAGGCC	CCCCCACCTT	TTTTTTTTGT	TTTTGGAGAC	AGAGTCTTGC	4797
TTTGTTGCCA	GGCTGGAGTG	CAGTGGCGCG	ATCTCGGCTC	ACCACAATCG	CTGCCTCCTG	4857
GGTTCAAGCA	ATTCTCCTGC	CTCAGCCTCC	CGAGTAGCTG	GGACTACAGG	TGTGCGCACA	4917
TGCCAGGCTA	ATTTTTGTAT	TTTTAGTAGA	GACGGGGTTT	CACCATGTTG	GCCGGGATGG	4977
TCTCGATCTC	TTGACCTCAT	GATCTACCCG	CCTTGGCCTC	CCAAAGTGCT	GAGATTACAG	5037
GCGTGAGCCA	CCGTGCCTGG	CCAGGCCCTT	TCTCTTTTAA	TGGAGACAGG	GTCTTGCACT	5097
ATCACCCAGG	CTGGAGTGCA	GTGGCATAAT	CATACCTCAT	TGCAGCCTCA	GACTCCTGGG	5157
TTCAAGCAAT	CCTCCTGCCT	CAGCCTCCCA	AGTAGCTGAG	ACTGCAGGCA	CGAGCCACCA	5217
CACCCAGCTA	ATTTTTAAGT	TTTCTTGTAG	AGACAGGGTC	TCACTATGTT	GTCTAGGCTG	5277
GTCTTGAAct	CTTGGCCTCA	AGTAATCCTC	CTGCCTCAGC	CTCCCAAAGT	GTTGGGATTG	5337
CAGATATGAG	CCACTGGCCT	GGCCTTCAGC	AGTTCTTTTT	GTGAAGTAAA	ACTTGTATGT	5397
TGGAAAGAGT	AGATTTTATT	GGTCTACCCT	TTTCTCACTG	TAGCTGCTGG	CAGCCCTGTG	5457
CCATATCTGG	ACTCTAGTTG	TCAGTATCTG	AGTTGGACAC	TATTCCTGCT	CCCTCTTGTT	5517
TCTTACATAT	CAGACTTCTT	ACTTGAATGA	AACCTGATCT	TTCTTAATCC	TCACTTTTTT	5577
CTTTTTTAAA	AAGCAGTTTC	TCCACTGCTA	AATGTTAGTC	ATTGAGGTGG	GGCCAATTTT	5637
AATCATAAGC	CTTAATAAGA	TTTTTCTAAG	AAATGTGAAA	TAGAACAATT	TTCATCTAAT	5697
TCCATTTACT	TTTAGATGAA	TGGCATTGTG	AATGCCATTC	TTTTAATGAA	TTTCAAGAGA	5757
ATTCTCTGGT	TTTCTGTGTA	ATTCCAGATG	AGTCACTGTA	ACTCTAGAAG	ATTAACCTTC	5817
CAGCCAACCT	ATTTTCCTTT	CCCTTGTCTC	TCTCATCCTC	TTTTCCTTCC	TTCTTTCCTT	5877
TCTCTTCTTT	TATCTCCAAG	GTTAATCAGG	AAAAATAGCT	TTTGACAGGG	GAAAAAACTC	5937
AATAACTAGC	TATTTTTGAC	CTCCTGATCA	GGAACCTTAG	TTGAAGCGTA	AATCTAAAGA	5997
AACATTTTCT	CTGAAATATA	TTATTAAGGG	CAATGGAGAT	AAATTAATAG	TAGATGTGGT	6057
TCCCAGAAAA	TATAATCAAA	ATTCAAAGAT	TTTTTTTTGTT	TCTGTAAGTG	GAAGTAAATC	6117
AAATGATTAC	TAGTGTTAAT	AGTAGATAAC	TTGTTTTTAT	TGTTGGTGCA	TATTAGTATA	6177
ACTGTGGGGT	AGGTCGGGGA	GAGGGTAAGG	GAATAGATCA	CTCAGATGTA	TTTLAGATAA	6237
GCTATTTAGC	CTTTGATGGA	ATCATAAATA	CAGTGAATAC	AATCCTTTGC	ATTGTTAAGG	6297
AGGTTTTTTG	TTTTTAAATG	GTGGGTCAAG	GAGCTAGTTT	ACAGGCTTAC	TGTGATTTAA	6357
GCAAATGTGA	AAAGTGAAAC	CTTAATTTTA	TCAAAAGAAA	TTTCTGTAAA	TGGTATGTCT	6417
CCTTAGAATA	CCCAAATCAT	AATTTTATTT	GTACACACTG	TTAGGGGCTC	ATCTCATGTA	6477
GGCAGAGTAT	AAAGTATTAC	CTTTTGGAAT	TAAAAGCCAC	TGACTGTTAT	AAAGTATAAC	6537
AACACACATC	AGGTTTTTAA	AAGCCTTGAA	TGGCCCTTGT	CTTAAAAAGA	AATTAGGAGC	6597
CAGGTGCGGT	GGCACGTGCC	TGTAGTCCCA	GCTCCTTGGG	AGGCTGAGAC	AGGAGGATTC	6657
CTTGAGCCCT	GGAGTTTGAG	TCCAGCCTGG	GTGACATAGC	AAGACCCTGT	CTTAAAAGAA	6717
AAATGGGAAG	AAAGACAAGG	TAACATGAAG	AAAGAAGAGA	TACCTAGTAT	GATGGAGCTG	6777
CAAATTTTCAT	GGCAGTTCAT	GCAGTCGGTC	AAGAGGAGGA	TTTTGTTTTG	TAGTTTGCAG	6837
ATGAGCATTT	CTAAAGCATT	TTCCCTTGCT	GTATTTTTTT	GTATTATAAA	TTACATTGGA	6897
CTTCATATAT	ATAATTTTTT	TTTACATTAT	ATGTCTCTTG	TATGTTTTGA	AACTCTTGTA	6957
TTTATGATAT	AGCTTATATG	ATTTTTTTGC	CTTGGTATAC	ATTTTAAAAT	ATGAATTTAA	7017
AAAATTTTTG	TAAAAATAAA	ATTCACAAAA	TTGTTTTGAA	AAACAAAAAA	AAAAAAA	7075

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1205 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
 (v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

WO 98/55615

PCT/US98/11773

85

Met	Asp	Ala	Lys	Ala	Arg	Asn	Cys	Leu	Leu	Gln	His	Arg	Glu	Ala	Leu
1				5				10						15	
Glu	Lys	Asp	Ile	Lys	Thr	Ser	Tyr	Ile	Met	Asp	His	Met	Ile	Ser	Asp
		20						25					30		
Gly	Phe	Leu	Thr	Ile	Ser	Glu	Glu	Glu	Lys	Val	Arg	Asn	Glu	Pro	Thr
	35						40					45			
Gln	Gln	Gln	Arg	Ala	Ala	Met	Leu	Ile	Lys	Met	Ile	Leu	Lys	Lys	Asp
	50					55					60				
Asn	Asp	Ser	Tyr	Val	Ser	Phe	Tyr	Asn	Ala	Leu	Leu	His	Glu	Gly	Tyr
65					70					75					80
Lys	Asp	Leu	Ala	Ala	Leu	Leu	His	Asp	Gly	Ile	Pro	Val	Val	Ser	Ser
			85						90					95	
Ser	Ser	Gly	Lys	Asp	Ser	Val	Ser	Gly	Ile	Thr	Ser	Tyr	Val	Arg	Thr
		100						105					110		
Val	Leu	Cys	Glu	Gly	Gly	Val	Pro	Gln	Arg	Pro	Val	Val	Phe	Val	Thr
	115					120						125			
Arg	Lys	Lys	Leu	Val	Asn	Ala	Ile	Gln	Gln	Lys	Leu	Ser	Lys	Leu	Lys
	130				135						140				
Gly	Glu	Pro	Gly	Trp	Val	Thr	Ile	His	Gly	Met	Ala	Gly	Cys	Gly	Lys
145					150					155					160
Ser	Val	Leu	Ala	Ala	Glu	Ala	Val	Arg	Asp	His	Ser	Leu	Leu	Glu	Gly
			165						170					175	
Cys	Phe	Pro	Gly	Gly	Val	His	Trp	Val	Ser	Val	Gly	Lys	Gln	Asp	Lys
		180						185					190		
Ser	Gly	Leu	Leu	Met	Lys	Leu	Gln	Asn	Leu	Cys	Thr	Arg	Leu	Asp	Gln
	195						200					205			
Asp	Glu	Ser	Phe	Ser	Gln	Arg	Leu	Pro	Leu	Asn	Ile	Glu	Glu	Ala	Lys
	210				215						220				
Asp	Arg	Leu	Arg	Ile	Leu	Met	Leu	Arg	Lys	His	Pro	Arg	Ser	Leu	Leu
225					230					235					240
Ile	Leu	Asp	Asp	Val	Trp	Asp	Ser	Trp	Val	Leu	Lys	Ala	Phe	Asp	Ser
			245						250					255	
Gln	Cys	Gln	Ile	Leu	Leu	Thr	Thr	Arg	Asp	Lys	Ser	Val	Thr	Asp	Ser
		260						265					270		
Val	Met	Gly	Pro	Lys	Tyr	Val	Val	Pro	Val	Glu	Ser	Ser	Leu	Gly	Lys
	275						280					285			
Glu	Lys	Gly	Leu	Glu	Ile	Leu	Ser	Leu	Phe	Val	Asn	Met	Lys	Lys	Ala
	290				295						300				
Asp	Leu	Pro	Glu	Gln	Ala	His	Ser	Ile	Ile	Lys	Glu	Cys	Lys	Gly	Ser
305					310					315					320
Pro	Leu	Val	Val	Ser	Leu	Ile	Gly	Ala	Leu	Leu	Arg	Asp	Phe	Pro	Asn
			325						330					335	
Arg	Trp	Glu	Tyr	Tyr	Leu	Lys	Gln	Leu	Gln	Asn	Lys	Gln	Phe	Lys	Arg
		340						345					350		
Ile	Arg	Lys	Ser	Ser	Ser	Tyr	Asp	Tyr	Glu	Ala	Leu	Asp	Glu	Ala	Met
	355						360					365			
Ser	Ile	Ser	Val	Glu	Met	Leu	Arg	Glu	Asp	Ile	Lys	Asp	Tyr	Tyr	Thr
	370					375					380				
Asp	Leu	Ser	Ile	Leu	Gln	Lys	Asp	Val	Lys	Val	Pro	Thr	Lys	Val	Leu
385					390					395					400
Cys	Ile	Leu	Trp	Asp	Met	Glu	Thr	Glu	Glu	Val	Glu	Asp	Ile	Leu	Gln
			405					410						415	
Glu	Phe	Val	Asn	Lys	Ser	Leu	Leu	Phe	Cys	Asp	Arg	Asn	Gly	Lys	Ser
		420						425					430		
Phe	Arg	Tyr	Tyr	Leu	His	Asp	Leu	Gln	Val	Asp	Phe	Leu	Thr	Glu	Lys
	435					440						445			
Asn	Cys	Ser	Gln	Leu	Gln	Asp	Leu	His	Lys	Lys	Ile	Ile	Thr	Gln	Phe
	450					455					460				
Gln	Arg	Tyr	His	Gln	Pro	His	Thr	Leu	Ser	Pro	Asp	Gln	Glu	Asp	Cys
465					470					475					480
Met	Tyr	Trp	Tyr	Asn	Phe	Leu	Ala	Tyr	His	Met	Ala	Ser	Ala	Lys	Met
			485					490						495	
His	Lys	Glu	Leu	Cys	Ala	Leu	Met	Phe	Ser	Leu	Asp	Trp	Ile	Lys	Ala
			500					505					510		

WO 98/55615

PCT/US98/11773

86

Lys	Thr	Glu	Leu	Val	Gly	Pro	Ala	His	Leu	Ile	His	Glu	Phe	Val	Glu
		515					520					525			
Tyr	Arg	His	Ile	Leu	Asp	Glu	Lys	Asp	Cys	Ala	Val	Ser	Glu	Asn	Phe
	530					535					540				
Gln	Glu	Phe	Leu	Ser	Leu	Asn	Gly	His	Leu	Leu	Gly	Arg	Gln	Pro	Phe
545					550					555					560
Pro	Asn	Ile	Val	Gln	Leu	Gly	Leu	Cys	Glu	Pro	Glu	Thr	Ser	Glu	Val
				565					570					575	
Tyr	Gln	Gln	Ala	Lys	Leu	Gln	Ala	Lys	Gln	Glu	Val	Asp	Asn	Gly	Met
			580					585					590		
Leu	Tyr	Leu	Glu	Trp	Ile	Asn	Lys	Lys	Asn	Ile	Thr	Asn	Leu	Ser	Arg
	595						600					605			
Leu	Val	Val	Arg	Pro	His	Thr	Asp	Ala	Val	Tyr	His	Ala	Cys	Phe	Ser
	610					615					620				
Glu	Asp	Gly	Gln	Arg	Ile	Ala	Ser	Cys	Gly	Ala	Asp	Lys	Thr	Leu	Gln
625					630					635					640
Val	Phe	Lys	Ala	Glu	Thr	Gly	Glu	Lys	Leu	Leu	Glu	Ile	Lys	Ala	His
				645					650					655	
Glu	Asp	Glu	Val	Leu	Cys	Cys	Ala	Phe	Ser	Thr	Asp	Asp	Arg	Phe	Ile
			660					665					670		
Ala	Thr	Cys	Ser	Val	Asp	Lys	Lys	Val	Lys	Ile	Trp	Asn	Ser	Met	Thr
	675						680					685			
Gly	Glu	Leu	Val	His	Thr	Tyr	Asp	Glu	His	Ser	Glu	Gln	Val	Asn	Cys
	690					695					700				
Cys	His	Phe	Thr	Asn	Ser	Ser	His	His	Leu	Leu	Leu	Ala	Thr	Gly	Ser
705				710						715					720
Ser	Asp	Cys	Phe	Leu	Lys	Leu	Trp	Asp	Leu	Asn	Gln	Lys	Glu	Cys	Arg
				725					730					735	
Asn	Thr	Met	Phe	Gly	His	Thr	Asn	Ser	Val	Asn	His	Cys	Arg	Phe	Ser
			740				745						750		
Pro	Asp	Asp	Lys	Leu	Leu	Ala	Ser	Cys	Ser	Ala	Asp	Gly	Thr	Leu	Lys
	755					760						765			
Leu	Trp	Asp	Ala	Thr	Ser	Ala	Asn	Glu	Arg	Lys	Ser	Ile	Asn	Val	Lys
	770					775					780				
Gln	Phe	Phe	Leu	Asn	Leu	Glu	Asp	Pro	Gln	Glu	Asp	Met	Glu	Val	Ile
785				790						795					800
Val	Lys	Cys	Cys	Ser	Trp	Ser	Ala	Asp	Gly	Ala	Arg	Ile	Met	Val	Ala
				805					810					815	
Ala	Lys	Asn	Lys	Ile	Phe	Leu	Trp	Asn	Thr	Asp	Ser	Arg	Ser	Lys	Val
			820					825					830		
Ala	Asp	Cys	Arg	Gly	His	Leu	Ser	Trp	Val	His	Gly	Val	Met	Phe	Ser
	835					840						845			
Pro	Asp	Gly	Ser	Ser	Phe	Leu	Thr	Ser	Ser	Asp	Asp	Gln	Thr	Ile	Arg
	850					855					860				
Leu	Trp	Glu	Thr	Lys	Lys	Val	Cys	Lys	Asn	Ser	Ala	Val	Met	Leu	Lys
865				870						875					880
Gln	Glu	Val	Asp	Val	Val	Phe	Gln	Glu	Asn	Glu	Val	Met	Val	Leu	Ala
				885					890					895	
Val	Asp	His	Ile	Arg	Arg	Leu	Gln	Leu	Ile	Asn	Gly	Arg	Thr	Gly	Gln
			900					905					910		
Ile	Asp	Tyr	Leu	Thr	Glu	Ala	Gln	Val	Ser	Cys	Cys	Cys	Leu	Ser	Pro
	915					920						925			
His	Leu	Gln	Tyr	Ile	Ala	Phe	Gly	Asp	Glu	Asn	Gly	Ala	Ile	Glu	Ile
	930					935					940				
Leu	Glu	Leu	Val	Asn	Asn	Arg	Ile	Phe	Gln	Ser	Arg	Phe	Gln	His	Lys
945				950						955					960
Lys	Thr	Val	Trp	His	Ile	Gln	Phe	Thr	Ala	Asp	Glu	Lys	Thr	Leu	Ile
				965					970					975	
Ser	Ser	Ser	Asp	Asp	Ala	Glu	Ile	Gln	Val	Trp	Asn	Trp	Gln	Leu	Asp
			980					985					990		
Lys	Cys	Ile	Phe	Leu	Arg	Gly	His	Gln	Glu	Thr	Val	Lys	Asp	Phe	Arg
	995					1000						1005			
Leu	Leu	Lys	Asn	Ser	Arg	Leu	Leu	Ser	Trp	Ser	Phe	Asp	Gly	Thr	Val
1010						1015					1020				

WO 98/55615

PCT/US98/11773

87

Lys Val Trp Asn Ile Ile Thr Gly Asn Lys Glu Lys Asp Phe Val Cys
 025 1030 1035 1040
 His Gln Gly Thr Val Leu Ser Cys Asp Ile Ser His Asp Ala Thr Lys
 1045 1050 1055
 Phe Ser Ser Thr Ser Ala Asp Lys Thr Ala Lys Ile Trp Ser Phe Asp
 1060 1065 1070
 Leu Leu Leu Pro Leu His Glu Leu Arg Gly His Asn Gly Cys Val Arg
 1075 1080 1085
 Cys Ser Ala Phe Ser Val Asp Ser Thr Leu Leu Ala Thr Gly Asp Asp
 1090 1095 1100
 Asn Gly Glu Ile Arg Ile Trp Asn Val Ser Asn Gly Glu Leu Leu His
 105 1110 1115 1120
 Leu Cys Ala Pro Leu Ser Glu Glu Gly Ala Ala Thr His Gly Gly Trp
 1125 1130 1135
 Val Thr Asp Leu Cys Phe Ser Pro Asp Gly Lys Met Leu Ile Ser Ala
 1140 1145 1150
 Gly Gly Tyr Ile Lys Trp Trp Asn Val Val Thr Gly Glu Ser Ser Gln
 1155 1160 1165
 Thr Phe Tyr Thr Asn Gly Thr Asn Leu Lys Lys Ile His Val Ser Pro
 1170 1175 1180
 Asp Phe Lys Thr Tyr Val Thr Val Asp Asn Leu Gly Ile Leu Tyr Ile
 185 1190 1195 1200
 Leu Gln Thr Leu Glu
 1205

WE CLAIM:

1. An isolated Apaf-1 protein having at least about 80% amino acid sequence identity with the sequence of SEQ ID NO:2 or the sequence of SEQ ID NO:16.
2. The Apaf-1 protein of claim 1 having at least about 90% amino acid sequence identity with the sequence of SEQ ID NO:2 or the sequence of SEQ ID NO:16.
3. The Apaf-1 protein of claim 2 having at least about 95% amino acid sequence identity with the sequence of SEQ ID NO:2 or the sequence of SEQ ID NO:16.
4. An isolated Apaf-1 protein comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:16.
5. A chimeric molecule comprising the Apaf-1 protein of claim 1 fused to a heterologous amino acid sequence.
6. The chimeric molecule of claim 5, wherein said heterologous amino acid sequence is an epitope tag sequence.
7. The chimeric molecule of claim 5, wherein said heterologous amino acid sequence is an immunoglobulin sequence.
8. The chimeric molecule of claim 7, wherein said immunoglobulin sequence is an IgG.
9. An antibody which binds to the Apaf-1 protein of claim 1.
10. The antibody of claim 9, wherein said antibody is a monoclonal antibody.
11. The antibody of claim 9, which is an agonist antibody.
12. An isolated nucleic acid comprising a nucleotide sequence encoding the Apaf-1 protein of claim 1.

13. The isolated nucleic acid of claim 12, comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:15.
14. The nucleic acid of claim 12, wherein said nucleotide sequence encodes an Apaf-1 protein comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:16.
15. An isolated nucleic acid comprising a nucleotide sequence which hybridizes under high stringency conditions to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:15.
16. A vector comprising the nucleic acid of claim 12.
17. The vector of claim 16 operably linked to control sequences recognized by a host cell transformed with the vector.
18. A host cell comprising the vector of claim 16.
19. A process for producing Apaf-1 protein comprising:
culturing the host cell of claim 18 to express Apaf-1 protein in the host cell.
20. A non-human, transgenic animal containing cells that expresses an inserted nucleotide sequence encoding Apaf-1 protein.
21. The animal of claim 20 which is a mouse or rat.
22. A non-human, knockout animal which contains cells having an altered gene encoding Apaf-1 protein.
23. The animal of claim 22 which is a mouse or rat.
24. An article of manufacture, comprising a container and a composition contained within said container, wherein the composition includes Apaf-1 protein or anti-Apaf-1 antibodies.

25. The article of manufacture of claim 24 further comprising instructions for using the Apaf-1 protein or anti-Apaf-1 antibodies *in vivo* or *ex vivo*.
26. A method for screening compounds for agents which modulate apoptosis, the method comprising:
- analyzing the ability of a candidate compound to competitively displace cytochrome c binding to an Apaf-1 protein; and
 - correlating competitive displacement of cytochrome c binding to said Apaf-1 protein by the candidate compound with the ability of the candidate compound to modulate apoptosis.
27. A method for identifying compounds which modulate apoptosis, the method comprising,
- incubating a candidate compound in a mixture containing Apaf-1 protein, Apaf-3 protein, dATP, and cytochrome c;
 - adding a caspase precursor to the incubating mixture; and
 - correlating a change in cleavage of the caspase precursor in the presence of the candidate compound as compared with cleavage in the absence of the compound with the candidate compound's ability to modulate apoptosis.

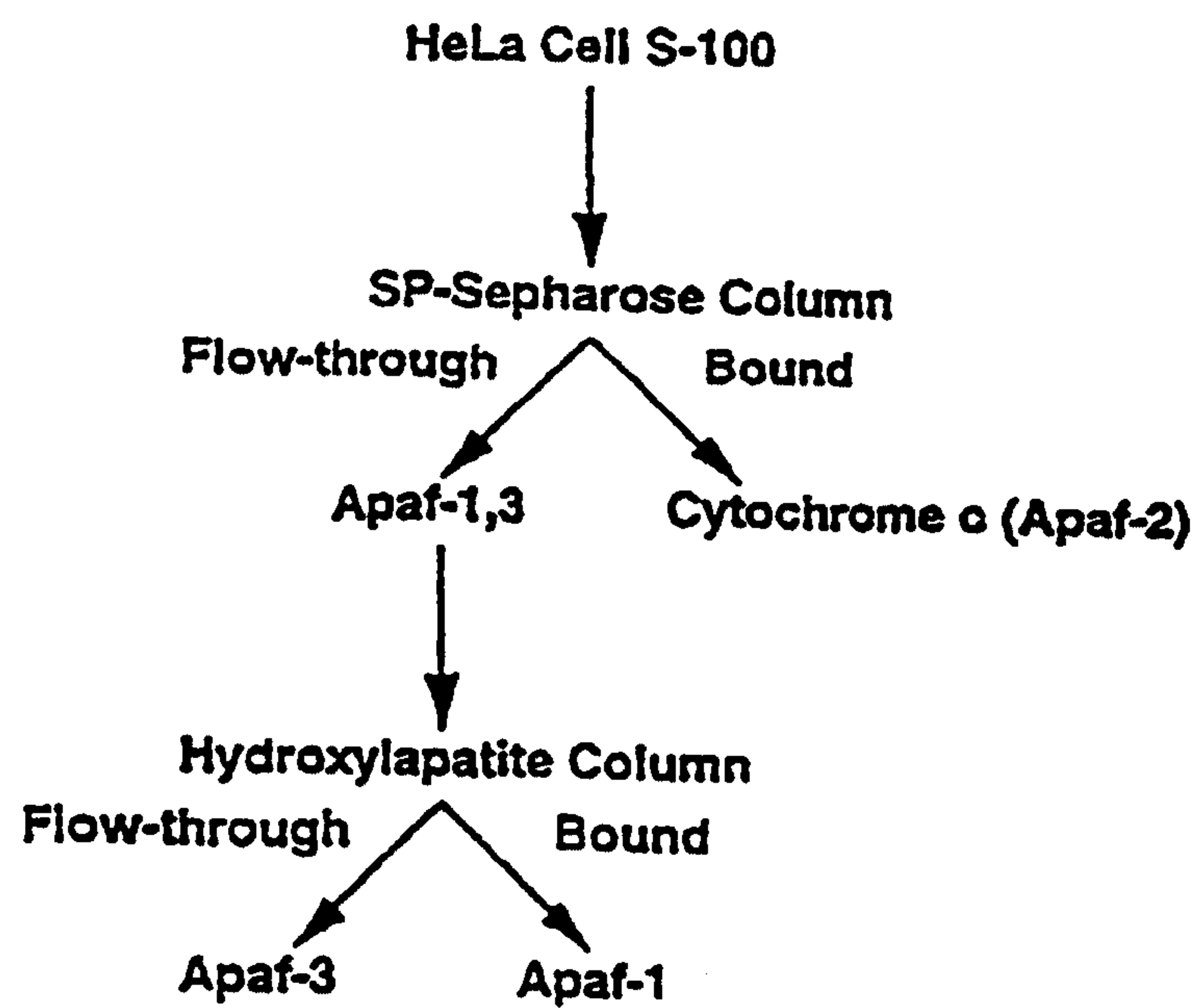
FIGURE 1

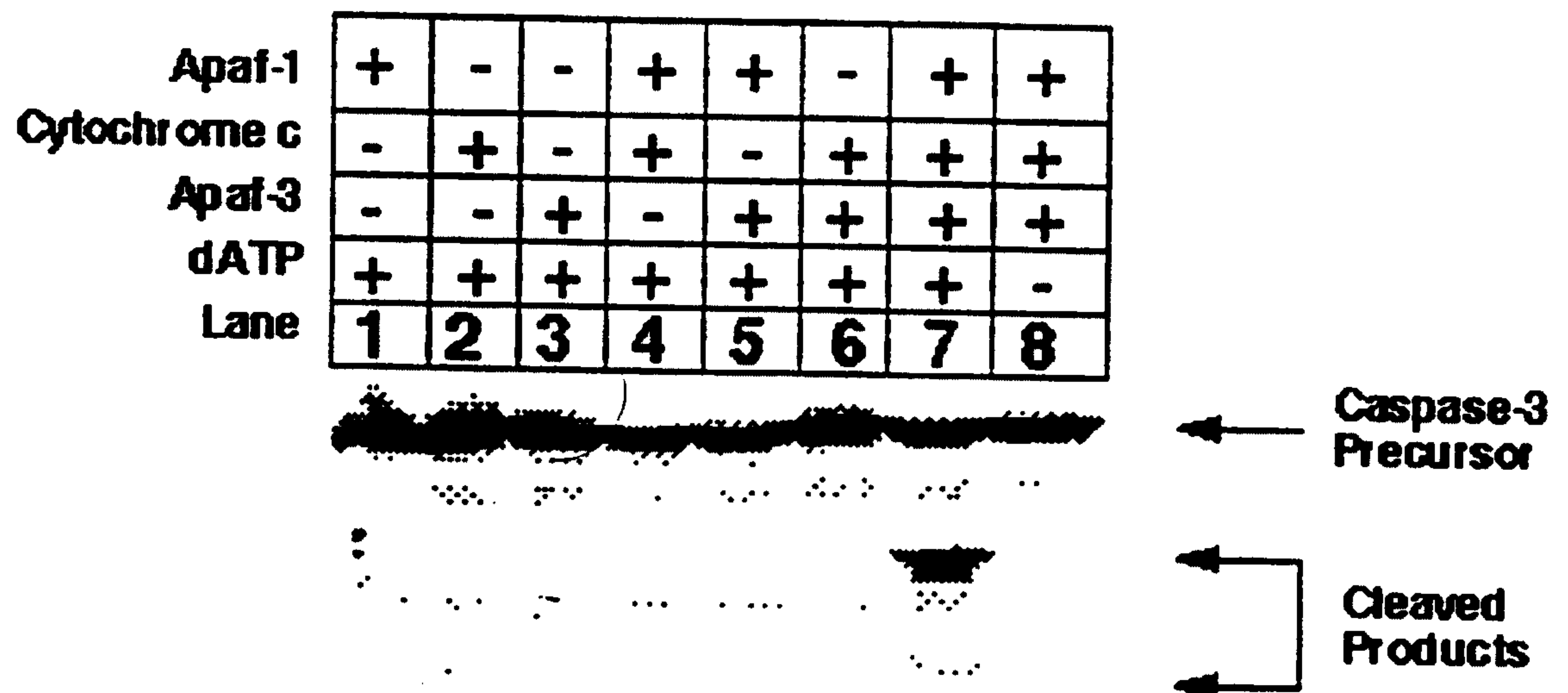
FIGURE 2

FIGURE 3

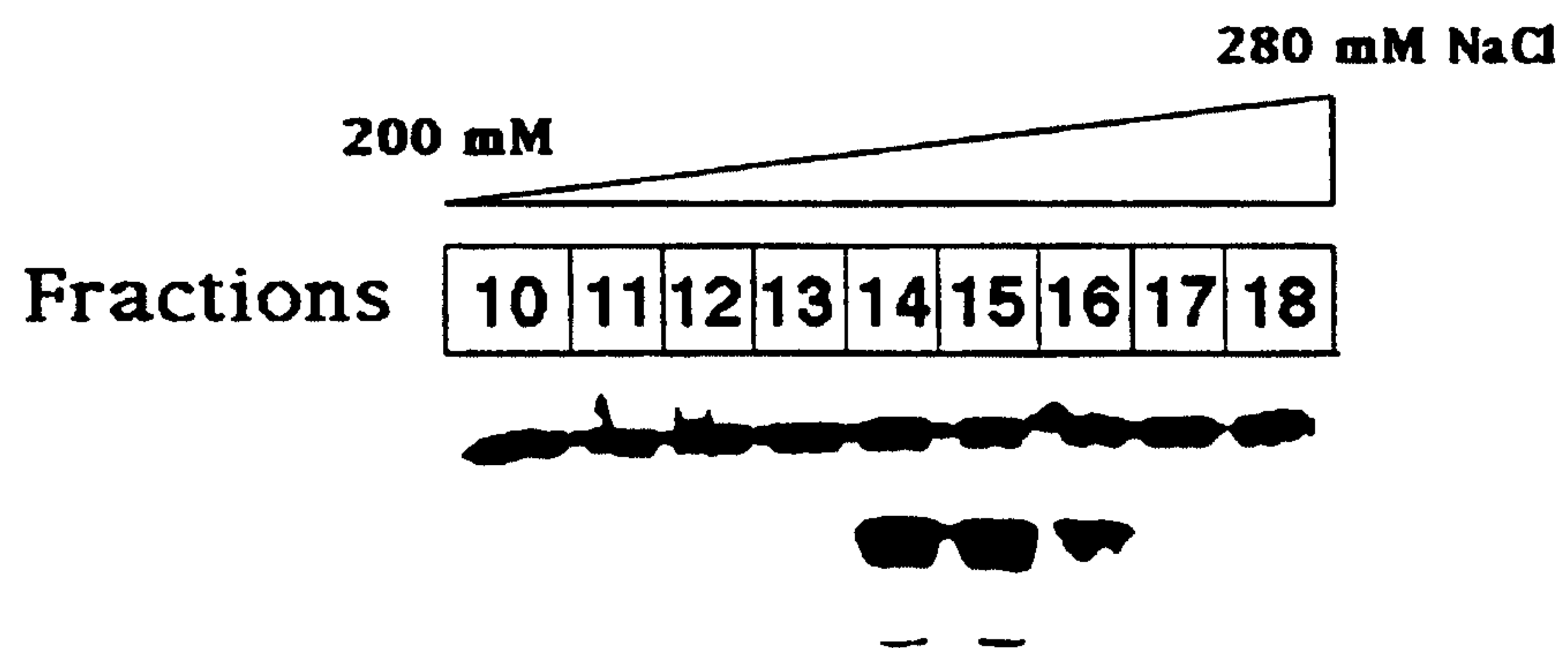
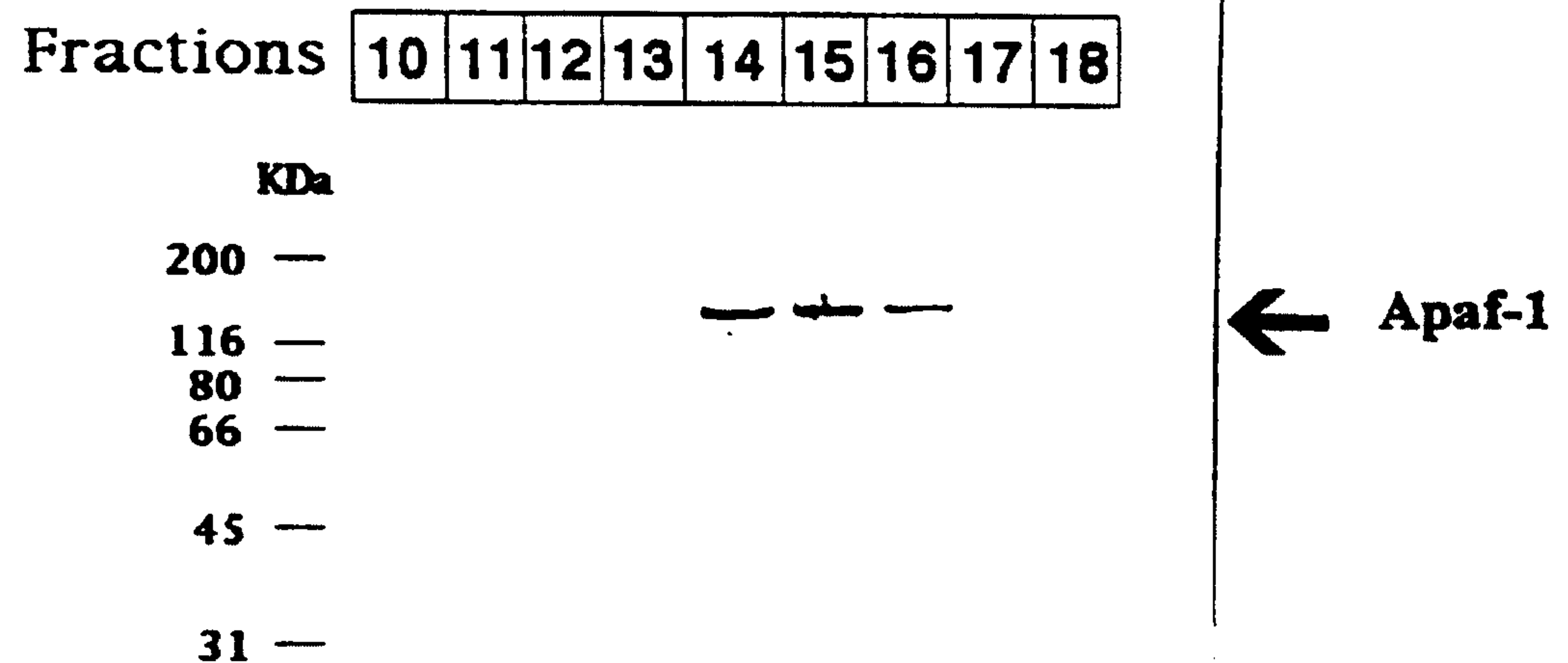


FIGURE 4

Sequence Range: 1 to 7042

FIGURE 5A

```

      10      20      30      40
      *      *      *      *
A AGA AGA GGT AGC GAG TGG ACG TGA CTG CTC TAT CCC GGG CAA AAG GGA
T TCT TCT CCA TCG CTC ACC TGC ACT GAC GAG ATA AGG CUC GTT TTC CCT
  R  G  S  E  W  T  *  L  L  Y  P  G  Q  K  G>

50      60      70      80      90
      *      *      *      *
TAG AAC CAG AGG TGG GGA GTC TGG GCA GTC GGC GAC CCG CGA AGA CTT
ATC TTG GTC TCC ACC CTT CAG ACC CGT CAG CCG CTC GGC GCT TCT GAA
  N  Q  R  W  G  V  W  A  V  G  D  P  R  R  L>

100     110     120     130     140
      *      *      *      *
GAG GTG CCG CAG CGG CAT CCG GAG TAG CGC CGG GCT CCC TCC GGG GTG
CTC CAC GGC GTC GCC GTA GGC CTC ATC GCG GCG CGA GGG AGG CCC CAC
  E  V  P  Q  R  H  P  E  *  R  R  A  P  S  G  V>

150     160     170     180     190
      *      *      *      *
CAG CCG CCG TCG GGG GAA GGG CGC CAC AGG CCG GGA AGA CCT CCT CCC
GTC GGC GGC AGC CUC CTT CUC GCG GTG TCC GGC CCT TCT GGA GGA GGG
  Q  P  P  S  G  E  G  R  H  R  P  G  R  P  P  P>

200     210     220     230     240
      *      *      *      *
TTT GTG TCC AGT AGT GGG GTC CAC CGG AGG GCG GCC CGT GGG CCG GGC
AAA CAC AGG TCA TCA CCC CAG GTG GCG TCC CGC CGG GCA CCC GGC CCG
  F  V  S  S  S  G  V  H  R  R  A  A  R  G  P  G>

250     260     270     280
      *      *      *      *
CTC ACC GCG GCG CTC CGG GAC TGT GGG GTC AGG CTG CGT TGG GTG GAC
GAG TGG CGC CGC GAG GGC CTG ACA CCG CAG TCC GAC GCA ACC CAC CTG
  L  T  A  A  L  R  D  C  G  V  R  L  R  W  V  D>

290     300     310     320     330
      *      *      *      *
GCC CAC CTC GCC AAC CTT CGG AGG TCC CTG GGG GTC TTC GTG CGC CCC
CGG GTG GAG CGG TTG GAA GCC TCC AGG GAC CCC CAG AAG CAC GCG GGG
  A  H  L  A  N  L  R  R  S  L  G  V  F  V  R  P>

340     350     360     370     380
      *      *      *      *
GGG GCT GCA GAG ATC CAG GGG AGG CCC CTG TGA CCC CCC GAC CTC CCC
CCC CGA CGT CTC TAG GTC CCC TCC GCG GAC ACT CCG GGC CTG GAC GGG
  G  A  A  E  I  Q  G  R  R  L  *  G  P  D  L  P>

390     400     410     420     430
      *      *      *      *
CCC CCC CCA CCC TAT CTC CCC ACA CAC ACC CCT CCA CCC CTA ATT CCC
GCC CCG CTT CCC ATA CAC CGC TCT GTC TCG GGA CGT GGG GAT TAA GCG
  R  G  E  G  Y  V  A  R  Q  S  P  A  P  L  I  P>

440     450     460     470     480
      *      *      *      *
GGT GGA AAA CTC CTG TTG CCG TTT CCC TCC ACC GGC CTG GAG TCT CCC
CCA CCT TTT GAG GAC AAC GGC AAA GGG AGG TGG CCG GAC CTC AGA GGG
  G  G  K  L  L  L  P  F  P  S  T  G  L  E  S  P>

```

FIGURE 5B

Translation
Start
548

490	500	510	520
AGT CTT GTC CCG GCA GTG CCG CCC TCC CCA CTA AGA CCT AGG CGC AAA			
TCA GAA CAG GGC CGT CAC GGC GGG AGG GGT GAT TCT GGA TCC GCG TTT			
S L V P A V P P S P L P P R R K>			
530	540	550	560
GGC TTG GCT CAT GGT TGA CAG CTC AGA GAG AGA AAG ATC TGA GGG AAG			
CCG AAC CGA GTA CCA ACT GTC GAG TCT CTC TCT TTC TAG ACT CCC TTC			
G L A H G * Q L R E R K I * G K>			
580	590	600	610
ATG GAT GCA AAA GCT CGA AAT TGT TTG CTT CAA CAT AGA GAA GCT CTG			
TAC CTA CGT TTT CGA GCT TTA ACA AAC GAA GTT GTA TCT CTT CGA GAC			
M D A K A R N C L L Q H R E A L>			
630	640	650	660
GAA AAG GAC ATC AAG ACA TCC TAC ATC ATG GAT CAC ATG ATT AGT GAT			
CTT TTC CTG TAG TTC TGT AGG ATG TAG TAC CTA GTG TAC TAA TCA CTA			
E K D I K T S Y I M D H M I S D>			
680	690	700	710
GGA TTT TTA ACA ATA TCA GAA GAG GAA AAA GTA AGA AAT GAG CCC ACT			
CCT AAA AAT TGT TAT AGT CTT CTC CTT TTT CAT TCT TTA CTC GGG TGA			
G F L T I S E E E K V F N E P T>			
730	740	750	760
CAA CAG CAA AGA GCA GCT ATG CTG ATT AAA ATG ATA CTT AAA AAA GAT			
GTT GTC GTT TCT CGT CGA TAC GAC TAA TTT TAC TAT GAA TTT TTT CTA			
Q Q Q R A A M L I K M I L K K D>			
770	780	790	800
AAT GAT TCC TAC GTA TCA TTC TAC AAT GCT CTA CTA CAT GAA GGA TAT			
TTA CTA AGG ATG CAT AGT AAG ATG TTA CGA GAT GAT GTA CTT CCT ATA			
N D S Y V S F Y N A L L H E G Y>			
820	830	840	850
AAA GAT CTT GCT GCC CTT CTC CAT GAT GGC ATT CCT GTT GTC TCT TCT			
TTT CTA GAA CGA CGG GAA GAG GTA CTA CCG TAA GGA CAA CAG AGA AGA			
K D L A A L L H D G I F V V S S>			
870	880	890	900
TCC AGT GTA AGG ACA GTC CTG TGT GAA GGT GGA GTA CCA CAG AGG CCA			
AGG TCA CAT TCC TGT CAG GAC ACA CTT CCA CCT CAT GGT GTC TCC GGT			
S S V R T V L C E G G V P Q R P>			
920	930	940	950
GTT GTT TTT GTC ACA AGG AAG AAG CTG GTG AAT GCA ATT CAG CAG AAG			
CAA CAA AAA CAG TGT TCC TTC TTC GAC CAC TTA CGT TAA GTC GTC TTC			
V V F V T R K K L V N A I Q Q K>			
970	980	990	1000

7/39

FIGURE 5C

```

CTC TCC AAA TTG AAA GGT GAA CCA GGA TGG GTC ACC ATA CAT GGA ATG
GAG AGG TTT AAC TTT CCA CTT GGT CCT ACC CAG TGG TAT GTA CCT TAC
  L  S  K  L  K  G  E  P  G  W  V  T  I  H  G  M>

1010      1020      1030      1040      1050
*          *          *          *          *
GCA GGC TGT GGG AAG TCT GTA TTA GCT GCA GAA GCT GTT AGA GAT CAT
CGT CCG ACA CCC TTC AGA CAT AAT CGA CGT CTT CGA CAA TCT CTA GTA
  A  G  C  G  K  S  V  L  A  A  E  A  V  R  D  H>

1060      1070      1080      1090      1100
*          *          *          *          *
TCC CTT TTA GAA GGT TGT TTC CCA GGG GGA GTG CAT TGG GTT TCA GTT
AGG GAA AAT CTT CCA ACA AAG GGT CCC CCT CAC GTA ACC CAA AGT CAA
  S  L  L  E  G  C  F  P  G  G  V  H  W  V  S  V>

1110      1120      1130      1140      1150
*          *          *          *          *
GGG AAA CAA GAC AAA TCT GGG CTT CTG ATG AAA CTG CAG AAT CTT TGC
CCC TTT GTT CTG TTT AGA CCC GAA GAC TAC TTT GAC GTC TTA GAA ACG
  G  K  Q  D  K  S  G  L  L  M  K  L  Q  N  L  C>

1160      1170      1180      1190      1200
*          *          *          *          *
ACA CGG TTG GAT CAG GAT GAG AGT TTT TCC CAG AGG CTT CCA CTT AAT
TGT GCC AAC CTA GTC CTA CTC TCA AAA AGG GTC TCC GAA GGT GAA TTA
  T  R  L  D  Q  D  E  S  F  S  Q  R  L  P  L  N>

1210      1220      1230      1240
*          *          *          *
ATT GAA GAG GCT AAA GAC CGT CTC CGC ATT CTG ATG CTT CGC AAA CAC
TAA CTT CTC CGA TTT CTG GCA GAG GCG TAA GAC TAC GAA GCG TTT GTG
  I  E  E  A  K  D  R  L  R  I  L  M  L  R  K  H>

1250      1260      1270      1280      1290
*          *          *          *          *
CCA AGG TCT CTC TTG ATC TTG GAT GAT GTT TGG GAC TCT TGG GTG TTG
GGT TCC AGA GAG AAC TAG AAC CTA CTA CAA ACC CTG AGA ACC CAC AAC
  P  R  S  L  L  I  L  D  D  V  W  D  S  W  V  L>

1300      1310      1320      1330      1340
*          *          *          *          *
AAA GCT TTT GAC AGT CAG TGT CAG ATT CTT CTT ACA ACC AGA GAC AAG
TTT CGA AAA CTG TCA GTC ACA GTC TAA GAA GAA TGT TGG TCT CTG TTC
  K  A  F  D  S  Q  C  Q  I  L  L  T  T  R  D  K>

1350      1360      1370      1380      1390
*          *          *          *          *
AGT GTT ACA GAT TCA GTA ATG GGT CCT AAA TAT GTA GTC CCT GTG GAG
TCA CAA TGT CTA AGT CAT TAC CCA GGA TTT ATA CAT CAG GGA CAC CTC
  S  V  T  D  S  V  M  G  P  K  Y  V  V  P  V  E>

1400      1410      1420      1430      1440
*          *          *          *          *
AGT TCC TTA GGA AAG GAA AAA GGA CTT GAA ATT TTA TCC CTT TTT GTT
TCA AGG AAT CCT TTC CTT TTT CCT GAA CTT TAA AAT AGG GAA AAA CAA
  S  S  L  G  K  E  K  G  L  E  I  L  S  L  F  V>

1450      1460      1470      1480
*          *          *          *
AAT ATG AAG AAG GCA GAT TTG CCA GAA CAA GCT CAT AGT ATT ATA AAA
TTA TAC TTC TTC CGT CTA AAC GGT CTT GTT CGA GTA TCA TAA TAT TTT

```

FIGURE 5D

N M K K A D L P E Q A H S I I K>
 1490 1500 1510 1520 1530
 * * * * *
 GAA TGT AAA GGC TCT CCC CTT GTA GTA TCT TTA ATT GGT GCA CTT TTA
 CTT ACA TTT CCG AGA GGG GAA CAT CAT AGA AAT TAA CCA CGT GAA AAT
 E C K G S P L V V S L I G A L L>
 1540 1550 1560 1570 1580
 * * * * *
 CGT GAT TTT CCC AAT CGC TGG GAG TAC TAC CTC AAA CAG CTT CAG AAT
 GCA CTA AAA GGG TTA GCG ACC CTC ATG ATG GAG TTT GTC GAA GTC TTA
 R D F P N R W E Y Y L K Q L Q N>
 1590 1600 1610 1620 1630
 * * * * *
 AAG CAG TTT AAG AGA ATA AGG AAA TCT TCG TCT TAT GAT TAT GAG GCT
 TTC GTC AAA TTC TCT TAT TCC TTT AGA AGC AGA ATA CTA ATA CTC CGA
 K Q F K R I R K S S S Y D Y E A>
 1640 1650 1660 1670 1680
 * * * * *
 CTA GAT GAA GCC ATG TCT ATA AGT GTT GAA ATG CTC AGA GAA GAC ATC
 GAT CTA CTT CGG TAC AGA TAT TCA CAA CTT TAC GAG TCT CTT CTG TAG
 L D E A M S I S V E M L R E D L>
 1690 1700 1710 1720
 * * * * *
 AAA GAT TAT TAC ACA GAT CTT TCC ATC CTT CAG AAG GAC GTT AAG GTG
 TTT CTA ATA ATG TGT CTA GAA AGG TAG GAA GTC TTC CTG CAA TTC CAC
 K D Y Y T D L S I L Q F D V K V>
 1730 1740 1750 1760 1770
 * * * * *
 CCT ACA AAG GTG TTA TGT ATT CTC TGG GAC ATG GAA ACT GAA GAA GTT
 GGA TGT TTC CAC AAT ACA TAA GAG ACC CTG TAC CTT TGA CTT CTT CAA
 P T K V L C I L W D M E T E E V>
 1780 1790 1800 1810 1820
 * * * * *
 GAA GAC ATA CTG CAG GAG TTT GTA AAT AAG TCT CTT TTA TTC TGT GAT
 CTT CTG TAT GAC GTC CTC AAA CAT TTA TTC AGA GAA AAT AAG ACA CTA
 E D I L Q E F V N K S L L F C D>
 1830 1840 1850 1860 1870
 * * * * *
 CGG AAT GGA AAG TCG TTT CGT TAT TAT TTA CAT GAT CTT CAA GTA GAT
 GCC TTA CCT TTC AGC AAA GCA ATA ATA AAT GTA CTA GAA GTT CAT CTA
 R N G K S F R Y Y L H D L Q V D>
 1880 1890 1900 1910 1920
 * * * * *
 TTT CTT ACA GAG AAG AAT TGC AGC CAG CTT CAG GAT CTA CAT AAG AAG
 AAA GAA TGT CTC TTC TTA ACG TCG GTC GAA GTC CTA GAT GTA TTC TTC
 F L T E K N C S Q L Q D L H K K>
 1930 1940 1950 1960
 * * * * *
 ATA ATC ACT CAG TTT CAG AGA TAT CAC CAG CCG CAT ACT CTT TCA CCA
 TAT TAG TGA GTC AAA GTC TCT ATA GTG GTC GGC GTA TGA GAA AGT GGT
 I I T Q F Q R Y H Q P H T L S P>

FIGURE 5E

1970	1980	1990	2000	2010
GAT CAG GAA GAC TGT ATG TAT TGG TAC AAC TTT CTG GCC TAT CAC ATG				
CTA GTC CTT CTG ACA TAC ATA ACC ATG TTG AAA GAC CGG ATA GTG TAC				
D Q E D C M Y W Y N F L A Y H M>				
2020	2030	2040	2050	2060
GCC AGT GCC AAG ATG CAC AAG GAA CTT TGT GCT TTA ATG TTT TCC CTG				
CGG TCA CGG TTC TAC GTG TTC CTT GAA ACA CGA AAT TAC AAA AGG GAC				
A S A K M H K E L C A L M F S L>				
2070	2080	2090	2100	2110
GAT TGG ATT AAA GCA AAA ACA GAA CTT GTA GGC CCT GCT CAT CTG ATT				
CTA ACC TAA TTT CGT TTT TGT CTT GAA CAT CCG GGA CGA GTA GAC TAA				
D W I K A K T E L V G F A H L I>				
2120	2130	2140	2150	2160
CAT GAA TTT GTG GAA TAC AGA CAT ATA CTA GAT GAA AAG GAT TGT GCA				
GTA CTT AAA CAC CTT ATG TCT GTA TAT GAT CTA CTT TTC CTA ACA CGT				
H E F V E Y R H I L D E K D C A>				
2170	2180	2190	2200	
GTC AGT GAG AAT TTT CAG GAG TTT TTA TCT TTA AAT GGA CAC CTT CTT				
CAG TCA CTC TTA AAA GTC CTC AAA AAT AGA AAT TTA CCT GTG GAA GAA				
V S E N F Q E F L S L N G H L L>				
2210	2220	2230	2240	2250
GGA CGA CAG CCA TTT CCT AAT ATT GTA CAA CTG GGT CTC TGT GAG CCG				
CCT GCT GTC GGT AAA GGA TTA TAA CAT GTT GAC CCA GAG ACA CTC GGC				
G R Q P F P N I V Q L G L C E P>				
2260	2270	2280	2290	2300
GAA ACT TCA GAA GTT TAT CAG CAA GCT AAG CTG CAG GCC AAG CAG GAG				
CTT TGA AGT CTT CAA ATA GTC GTT CGA TTC GAC GTC CGG TTC GTC CTC				
E T S E V Y Q Q A K L Q A K Q E>				
2310	2320	2330	2340	2350
GTC GAT AAT GGA ATG CTT TAC CTG GAA TGG ATA AAC AAA AAA AAC ATC				
CAG CTA TTA CCT TAC GAA ATG GAC CTT ACC TAT TTG TTT TTT TTG TAG				
V D N G M L Y L E W I N K K N I>				
2360	2370	2380	2390	2400
ACG AAT CTT TCC CGC TTA GTT GTC CGC CCC CAC ACA GAT GCT GTT TAC				
TGC TTA GAA AGG GCG AAT CAA CAG GCG GGG GTG TGT CTA CGA CAA ATG				
T N L S R L V V R P H T D A V Y>				
2410	2420	2430	2440	
CAT GCC TGC TTT TCT GAG GAT GGT CAG AGA ATA GCT TCT TGT GGA GCT				
GTA CGG ACG AAA AGA CTC CTA CCA GTC TCT TAT CGA AGA ACA CCT CGA				
H A C F S E D G Q R I A S C G A>				
2450	2460	2470	2480	2490

FIGURE 5F

```

GAT AAA ACC TTA CAG GTG TTC AAA GCT GAA ACA GGA GAG AAA CTT CTA
CTA TTT TGG AAT GTC CAC AAG TTT CGA CTT TGT CCT CTC TTT GAA GAT
D K T L Q V F K A E T G E K L L>

2500      2510      2520      2530      2540
*          *          *          *          *
GAA ATC AAG GCT CAT GAG GAT GAA GTG CTT TGT TGT GCA TTC TCT ACA
CTT TAG TTC CGA GTA CTC CTA CTT CAC GAA ACA ACA CGT AAG AGA TGT
E I K A H E D E V L C C A F S T>

2550      2560      2570      2580      2590
*          *          *          *          *
GAT GAC AGA TTT ATA GCA ACC TGC TCA GTG GAT AAA AAA GTG AAG ATT
CTA CTG TCT AAA TAT CGT TGG ACG AGT CAC CTA TTT TTT CAC TTC TAA
D D R F I A T C S V D K K V K L>

2600      2610      2620      2630      2640
*          *          *          *          *
TGG AAT TCT ATG ACT GGG GAA CTA GTA CAC ACC TAT GAT GAG CAC TCA
ACC TTA AGA TAC TGA CCC CTT GAT CAT GTG TGG ATA CTA CTC GTG AGT
W N S M T G E L V H T Y D E H S>

2650      2660      2670      2680
*          *          *          *
GAG CAA GTC AAT TGC TGC CAT TTC ACC AAC AGT AGT CAT CAT CTT CTC
CTC GTT CAG TTA ACG ACG GTA AAG TGG TTG TCA TCA GTA GTA GAA GAG
E Q V N C C H F T N S S H H L L>

2690      2700      2710      2720      2730
*          *          *          *          *
TTA GCC ACT GGG TCA AGT GAC TGC TTC CTC AAA CTT TGG GAT TTG AAT
AAT CGG TGA CCC AGT TCA CTG ACG AAG GAG TTT GAA ACC CTA AAC TTA
L A T G S S D C F L K L W D L N>

2740      2750      2760      2770      2780
*          *          *          *          *
CAA AAA GAA TGT CGA AAT ACC ATG TTT GGT CAT ACA AAT TCA GTC AAT
GTT TTT CTT ACA GCT TTA TGG TAC AAA CCA GTA TGT TTA AGT CAG TTA
Q K E C R N T M F G H T N S V N>

2790      2800      2810      2820      2830
*          *          *          *          *
CAC TGC AGA TTT TCA CCA GAT GAT AAG CTT TTG GCT AGT TGT TCA GCT
GTG ACG TCT AAA AGT GGT CTA CTA TTC GAA AAC CGA TCA ACA AGT CGA
H C R F S P D D K L L A S C S A>

2840      2850      2860      2870      2880
*          *          *          *          *
GAT GGA ACC TTA AAG CTT TGG GAT GCG ACA TCA GCA AAT GAG AGG AAA
CTA CCT TGG AAT TTC GAA ACC CTA CGC TGT AGT CGT TTA CTC TCC TTT
D G T L K L W D A T S A N E R K>

2890      2900      2910      2920
*          *          *          *
AGC ATT AAT GTG AAA CAG TTC TTC CTA AAT TTG GAG GAC CCT CAA GAG
TCG TAA TTA CAC TTT GTC AAG AAG GAT TTA AAC CTC CTG GGA GTT CTC
S I N V K Q F F L N L E D P Q E>

2930      2940      2950      2960      2970
*          *          *          *          *
GAT ATG GAA GTG ATA GTG AAG TGT TGT TCG TGG TCT GCT GAT GGT GCA
CTA TAC CTT CAC TAT CAC TTC ACA ACA AGC ACC AGA CGA CTA CCA CGT

```

FIGURE 5G

D M E V I V K C C S W S A D G A>
 2980 2990 3000 3010 3020
 * * * * *
 AGG ATA ATG GTG GCA GCA AAA AAT AAA ATC TTT TTG TGG AAT ACA GAC
 TCC TAT TAC CAC CGT CGT TTT TTA TTT TAG AAA AAC ACC TTA TGT CTG
 R E M V A A K N K I F L W N T D>
 3030 3040 3050 3060 3070
 * * * * *
 TCA CGT TCA AAG GTG GCT GAT TGC AGA GGA CAT TTA AGT TGG GTT CAT
 AGT GCA AGT TTC CAC CGA CTA ACG TCT CCT GTA AAT TCA ACC CAA GTA
 S R S K V A D C R G H L S W V H>
 3080 3090 3100 3110 3120
 * * * * *
 GGT GTG ATG TTT TCT CCT GAT GGA TCA TCA TTT TTG ACA TCT TCT GAT
 CCA CAC TAC AAA AGA GGA CTA CCT AGT AGT AAA AAC TGT AGA AGA CTA
 G V M F S P D G S S F L T S S D>
 3130 3140 3150 3160
 * * * *
 GAC CAG ACA ATC AGG CTC TGG GAG ACA AAG AAA GTA TGT AAG AAC TCT
 CTG GTC TGT TAG TCC GAG ACC CTC TGT TTC TTT CAT ACA TTC TTG AGA
 D Q T I R L W E T K K V C K N S>
 3170 3180 3190 3200 3210
 * * * * *
 GCT GTA ATG TTA AAG CAA GAA GTA GAT GTT GTG TTT CAA GAA AAT GAA
 CGA CAT TAC AAT TTC GTT CTT CAT CTA CAA CAC AAA GTT CTT TTA CTT
 A V M L K Q E V D V V F Q E N E>
 3220 3230 3240 3250 3260
 * * * * *
 GTG ATG GTC CTT GCA GTT GAC CAT ATA AGA CGT CTG CAA CTC ATT AAT
 CAC TAC CAG GAA CGT CAA CTG GTA TAT TCT GCA GAC GTT GAG TAA TTA
 V M V L A V D H I R R L Q L I N>
 3270 3280 3290 3300 3310
 * * * * *
 GGA AGA ACA GGT CAG ATT GAT TAT CTG ACT GAA GCT CAA GTT AGC TGC
 CCT TCT TGT CCA GTC TAA CTA ATA GAC TGA CTT CGA GTT CAA TCG ACG
 G R T G Q I D Y L T E A Q V S C>
 3320 3330 3340 3350 3360
 * * * * *
 TGT TGC TTA AGT CCA CAT CTT CAG TAC ATT GCA TTT GGA GAT GAA AAT
 ACA ACG AAT TCA GGT GTA GAA GTC ATG TAA CGT AAA CCT CTA CTT TTA
 C C L S P H L Q Y I A F G D E N>
 3370 3380 3390 3400
 * * * *
 GGA GCC ATT GAG ATT TTA GAA CTT GTA AAC AAT AGA ATC TTC CAG TCC
 CCT CGG TAA CTC TAA AAT CTT GAA CAT TTG TTA TCT TAG AAG GTC AGG
 G A I E I L E L V N N R I F Q S>
 3410 3420 3430 3440 3450
 * * * * *
 AGG TTT CAG CAC AAG AAA ACT GTA TGG CAC ATC CAG TTC ACA GCC GAT
 TCC AAA GTC GTG TTC TTT TGA CAT ACC GTG TAG GTC AAG TGT CGG CTA
 R F Q H K K T V W H I Q F T A D>

WO 98/55615

12/39

PCT/US98/11773

FIGURE 5H

```

3460      3470      3480      3490      3500
*         *         *         *         *
GAG AAG ACT CTT ATT TCA AGT TCT GAT GAT GCT GAA ATT CAG GTA TGG
CTC TTC TGA GAA TAA AGT TCA AGA CTA CTA CGA CTT TAA GTC CAT ACC
E   K   T   L   I   S   S   S   D   D   A   E   I   Q   V   W>

3510      3520      3530      3540      3550
*         *         *         *         *
AAT TGG CAA TTG GAC AAA TGT ATC TTT CTA CGA GCC CAT CAG GAA ACA
TTA ACC GTT AAC CTG TTT ACA TAG AAA GAT GCT CCG GTA GTC CTT TGT
N   W   Q   L   D   K   C   I   F   L   R   G   H   Q   E   T>

3560      3570      3580      3590      3600
*         *         *         *         *
GTG AAA GAC TTT AGA CTC TTG AAA AAT TCA AGA CTG CTT TCT TGG TCA
CAC TTT CTG AAA TCT GAG AAC TTT TTA AGT TCT GAC GAA AGA ACC AGT
V   K   D   F   R   L   L   K   N   S   R   L   L   S   W   S>

3610      3620      3630      3640
*         *         *         *
TTT GAT GGA ACA GTG AAG GTA TGG AAT ATT ATT ACT GGA AAT AAA GAA
AAA CTA CCT TGT CAC TTC CAT ACC TTA TAA TAA TGA CCT TTA TTT CTT
F   D   G   T   V   K   V   W   N   I   I   T   G   N   K   E>

3650      3660      3670      3680      3690
*         *         *         *         *
AAA GAC TTT GTC TGT CAC CAG GGT ACA GTA CTT TCT TGT GAC ATT TCT
TTT CTG AAA CAG ACA GTG GTC CCA TGT CAT GAA AGA ACA CTG TAA AGA
K   D   F   V   C   H   Q   G   T   V   L   S   C   D   I   S>

3700      3710      3720      3730      3740
*         *         *         *         *
CAC GAT GCT ACC AAG TTT TCA TCT ACC TCT GCT GAC AAG ACT GCA AAG
GTG CTA CGA TGG TTC AAA AGT AGA TGG AGA CGA CTG TTC TGA CGT TTC
H   D   A   T   K   F   S   S   T   S   A   D   K   T   A   K>

3750      3760      3770      3780      3790
*         *         *         *         *
ATC TGG AGT TTT GAT CTC CTT TTG CCA CTT CAT GAA TTG AGG GGC CAC
TAG ACC TCA AAA CTA GAG GAA AAC GGT GAA GTA CTT AAC TCC CCG GTG
I   W   S   F   D   L   L   L   P   L   H   E   L   R   G   H>

3800      3810      3820      3830      3840
*         *         *         *         *
AAC GGC TGT GTG CGC TGC TCT GCC TTC TCT GTG GAC AGT ACC CTG CTG
TTG CCG ACA CAC GCG ACG AGA CGG AAG AGA CAC CTG TCA TGG GAC GAC
N   G   C   V   R   C   S   A   F   S   V   D   S   T   L   L>

3850      3860      3870      3880
*         *         *         *
GCA ACG GGA GAT GAC AAT GGA GAA ATC AGG ATA TGG AAT GTC TCA AAC
CGT TGC CCT CTA CTG TTA CCT CTT TAG TCC TAT ACC TTA CAG AGT TTG
A   T   G   D   D   N   G   E   I   R   I   W   N   V   S   N>

3890      3900      3910      3920      3930
*         *         *         *         *
GGT GAG CTT CTT CAT TTG TGT GCT CCG CTT TCA GAA GAA GGA GCT GCT
CCA CTC GAA GAA GTA AAC ACA CGA CGC GAA AGT CTT CTT CCT CGA CGA
G   E   L   L   H   L   C   A   P   L   S   E   E   G   A   A>

3940      3950      3960      3970      3980
*         *         *         *         *

```

WO 98/55615

13/39

PCT/US98/11773

FIGURE 51

ACC CAT GGA GGC TGG GTG ACT GAC CTT TGC TTT TCT CCA GAT GGC AAA
 TGG GTA CCT CCG ACC CAC TGA CTG GAA ACG AAA AGA GGT CTA CCG TTT
 T H G G W V T D L C F S P D G K>

3990 4000 4010 4020 4030
 ATG CTT ATC TCT GCT GGA GGA TAT ATT AAG TGG TGG AAC GTT GTC ACT
 TAC GAA TAG AGA CGA CCT CCT ATA TAA TTC ACC ACC TTG CAA CAG TGA
 M L I S A G G Y I K W W N V V T>

4040 4050 4060 4070 4080
 GGG GAA TCC TCA CAG ACC TTC TAC ACA AAT GGA ACC AAT CTT AAG AAA
 CCC CTT AGG AGT GTC TGG AAG ATG TGT TTA CCT TGG TTA GAA TTC TTT
 G E S S Q T F Y T N G T N L K K>

4090 4100 4110 4120
 ATA CAC GTG TCC CCT GAC TTC AAA ACA TAT GTG ACT GTG GAT AAT CTT
 TAT GTG CAC AGG GGA CTG AAG TTT TGT ATA CAC TGA CAC CTA TTA GAA
 I H V S P D F K T Y V T V D N L>

4130 4140 4150 4160 4170
 GGT ATT TTA TAT ATT TTA CAG ACT TTA GAA TAA AAT AGT TAA GCA TTA
 CCA TAA AAT ATA TAA AAT GTC TGA AAT CTT ATT TTA TCA ATT CGT AAT
 G I L Y I L Q T L E N S A L>

4180 4190 4200 4210 4220
 ATG TAG TTG AAC TTT TTA AAT TTT TGA ATT GGA AAA AAA TTC TAA TGA
 TAC ATC AAC TTG AAA AAT TTA AAA ACT TAA CCT TTT TTT AAG ATT ACT
 M * L N F L N F * I G K K F * >

4230 4240 4250 4260 4270
 AAC CCT GAT ATC AAC TTT TTA TAA AGC TCT TAA TTG TTG TGC AGT ATT
 TTG GGA CTA TAG TTG AAA AAT ATT TCG AGA ATT AAC AAC ACG TCA TAA
 N P D I N F L * S S * L L C S I>

4280 4290 4300 4310 4320
 GCA TTC ATT ACA AAA GTG TTT GTG GTT GGA TGA ATA ATA TTA ATG TAG
 CGT AAG TAA TGT TTT CAC AAA CAC CAA CCT ACT TAT TAT AAT TAC ATC
 A F I T K V F V V G * I I L M >

4330 4340 4350 4360
 CTT TTT CCC AAA TGA ACA TAC CTT TAA TCT TGT TTT TCA TGA TCA TCA
 GAA AAA GGG TTT ACT TGT ATG GAA ATT AGA ACA AAA AGT ACT AGT AGT
 L F P K * T Y L * S C F S * S S>

4370 4380 4390 4400 4410
 TTA ACA GTT TGT CCT TAG GAT GCA AAT GAA AAT GTG AAT ACA TAC CTT
 AAT TGT CAA ACA GGA ATC CTA CGT TTA CTT TTA CAC TTA TGT ATG GAA
 L T V C P * D A N E N V N T Y L>

4420 4430 4440 4450 4460
 GTT GTA CTG TTG GTA AAA TTC TGT CTT GAT GCA TTC AAA ATG GTT GAC
 CAA CAT GAC AAC CAT TTT AAG ACA GAA CTA CGT AAG TTT TAC CAA CTG

Translation
 4159

FIGURE 5J

V V L L V K F C L D A F K M V D>

4470 4480 4490 4500 4510

ATA ATT AAT GAG AAG AAT TTG GAA GAA ATT GGT ATT TTA ATA CTG TCT
TAT TAA TTA CTC TTC TTA AAC CTT CTT TAA CCA TAA AAT TAT GAC AGA
I I N E K N L E E I G I L I L S>

4520 4530 4540 4550 4560

GTA TTT ATT ACT GTT ATG CAG GCT GTG CCT CAG GGT AGC AGT GGC CTG
CAT AAA TAA TGA CAA TAC GTC CGA CAC GGA GTC CCA TCG TCA CCG GAC
V F I T V M Q A V P Q G S S G L>

4570 4580 4590 4600

CTT TTT GAA CCA CAC TTA CCC CAA GGG GGT TTT GTT CTC CTA AAT ACA
GAA AAA CTT GGT GTG AAT GGG GTT CCC CCA AAA CAA GAG GAT TTA TGT
L F E P H L P Q G G F V L L N T>

4610 4620 4630 4640 4650

ATC TTA GAG GTT TTT TGC ACT CTT TAA ATT TGC TTT AAA AAT ATT GTG
TAG AAT CTC CAA AAA ACG TGA GAA ATT TAA ACG AAA TTT TTA TAA CAC
I L E V F C T L * I C F K N I V>

4660 4670 4680 4690 4700

TCT GTG TGC ATA GTC TGC AGC ATT TCC TTT AAT TGA CTC AAT AAG TGA
AGA CAC ACG TAT CAG ACG TCG TAA AGG AAA TTA ACT GAG TTA TTC ACT
S V C I V C S I S F N * L N K *>

4710 4720 4730 4740 4750

GTC TCG GAT TTA GCA GGC CCC CCC ACC TTT TTT TTT TGT TTT TGG AGA
CAG AAC CTA AAT CGT CCG GGG GGG TGG AAA AAA AAA ACA AAA ACC TCT
V L D L A G P P T F F F C F W R>

4760 4770 4780 4790 4800

CAG AGT CTT GCT TTG TTG CCA GGC TGG AGT GCA GTG GCG CGA TCT CGG
GTC TCA GAA CGA AAC AAC GGT CCG ACC TCA CGT CAC CGC GCT AGA GCC
Q S L A L L P G W S A V A R S R>

4810 4820 4830 4840

CTC ACC ACA ATC GCT GCC TCC TGG GTT CAA GCA ATT CTC CTG CCT CAG
GAG TGG TGT TAG CGA CCG AGG ACC CAA GTT CGT TAA GAG GAC GGA GTC
L T T E A A S W V Q A I L L P Q>

4850 4860 4870 4880 4890

CCT CCC GAG TAG CTG GGA CTA CAG GTG TGC GCA CAT GCC AGG CTA ATT
GGA GGG CTC ATC GAC CCT GAT GTC CAC ACG CGT GTA CGG TCC GAT TAA
P P E * L G L Q V C A H A R L I>

4900 4910 4920 4930 4940

TTT GTA TTT TTA GTA GAG ACG GGG TTT CAC CAT GTT GGC CGG GAT GGT
AAA CAT AAA AAT CAT CTC TGC CCC AAA GTG GTA CAA CCG GCC CTA CCA
F V F L V E T G F H H V G R D G>

FIGURE 5K

```

4950      4960      4970      4980      4990
*         *         *         *         *
CTC GAT CTC TTG ACC TCA TGA TCT ACC CGC CTT GGC CTC CCA AAG TGC
GAG CTA GAG AAC TGG AGT ACT AGA TGG GCG GAA CCG GAG GGT TTC ACG
L   D   L   L   T   S   *   S   T   R   L   G   L   P   K   C>

5000      5010      5020      5030      5040
*         *         *         *         *
TGA GAT TAC AGG CGT GAG CCA CCG TGC CTG GCC AGG CCC CTT CTC TTT
ACT CTA ATG TCC GCA CTC GGT GGC ACG GAC CGG TCC GGG GAA GAG AAA
*   D   Y   R   R   E   P   P   C   L   A   R   P   L   L   F>

5050      5060      5070      5080
*         *         *         *
TAA TGG AGA CAG GGT CTT GCA CTA TCA CCC AGG CTG GAG TGC AGT GGC
ATT ACC TCT GTC CCA GAA CGT GAT AGT GGG TCC GAC CTC ACG TCA CCG
*   W   R   Q   G   L   A   L   S   P   R   L   E   C   S   G>

5090      5100      5110      5120      5130
*         *         *         *         *
ATA ATC ATA CCT CAT TGC AGC CTC AGA CTC CTG GGT TCA AGC AAT CCT
TAT TAG TAT GGA GTA ACG TCG GAG TCT GAG GAC CCA AGT TCG TTA GGA
I   I   I   P   H   C   S   L   R   L   L   G   S   S   N   P>

5140      5150      5160      5170      5180
*         *         *         *         *
CCT GCC TCA GCC TCC CAA GTA GCT GAG ACT GCA GGC ACG AGC CAC CAC
GGA CGG AGT CGG AGG GTT CAT CGA CTC TGA CGT CCG TGC TCG GTG GTG
P   A   S   A   S   Q   V   A   E   T   A   G   T   S   H   H>

5190      5200      5210      5220      5230
*         *         *         *         *
ACC CAG CTA ATT TTT AAG TTT TCT TGT AGA GAC AGG GTC TCA CTA TGT
TGG GTC GAT TAA AAA TTC AAA AGA ACA TCT CTG TCC CAG AGT GAT ACA
T   Q   L   I   F   K   F   S   C   R   D   R   V   S   L   C>

5240      5250      5260      5270      5280
*         *         *         *         *
TGT CTA GGC TGG TCT TGA ACT CTT GGC CTC AAG TAA TCC TCC TGC CTC
ACA GAT CCG ACC AGA ACT TGA GAA CCG GAG TTC ATT AGG AGG ACG GAG
C   L   G   W   S   *   T   L   G   L   K   *   S   S   C   L>

5290      5300      5310      5320
*         *         *         *
AGC CTC CCA AAG TGT TGG GAT TGC AGA TAT GAG CCA CTG GCC TGG CCT
TCG GAG GGT TTC ACA ACC CTA ACG TCT ATA CTC GGT GAC CGG ACC GGA
S   L   P   K   C   W   D   C   R   Y   E   P   L   A   W   P>

5330      5340      5350      5360      5370
*         *         *         *         *
TCA GCA GTT CTT TTT GTG AAG TAA AAC TTG TAT GTT GGA AAG AGT AGA
AGT CGT CAA GAA AAA CAC TTC ATT TTG AAC ATA CAA CCT TTC TCA TCT
S   A   V   L   F   V   K   *   N   L   Y   V   G   X   S   R>

5380      5390      5400      5410      5420
*         *         *         *         *
TTT TAT TCG TCT ACC CTT TTC TCA CTG TAG CTG CTG GCA GCC CTG TGC
AAA ATA ACC AGA TGG GAA AAG AGT GAC ATC GAC GAC CGT CGG GAC ACG
F   Y   W   S   T   L   F   S   L   *   L   L   A   A   L   C>

5430      5440      5450      5460      5470
*         *         *         *         *

```

FIGURE 5L

```

CAT ATC TGG ACT CTA GTT GTC AGT ATC TGA GTT GGA CAC TAT TCC TGC
GTA TAG ACC TGA GAT CAA CAG TCA TAG ACT CAA CCT GTG ATA AGG ACG
H I W T L V V S I * V G H Y S C>

5480      5490      5500      5510      5520
*         *         *         *         *
TCC CTC TTG TTT CTT ACA TAT CAG ACT TCT TAC TTG AAT GAA ACC TGA
AGG GAG AAC AAA GAA TGT ATA GTC TGA AGA ATG AAC TTA CTT TGG ACT
S L L F L T Y Q T S Y L N E T *>

5530      5540      5550      5560
*         *         *         *
TCT TTC CTA ATC CTC ACT TTT TTC TTT TTT AAA AAG CAG TTT CTC CAC
AGA AAG GAT TAG GAG TGA AAA AAG AAA AAA TTT TTC GTC AAA GAG GTG
S F L I L T F F F F K K Q F L H>

5570      5580      5590      5600      5610
*         *         *         *         *
TGC TAA ATG TTA GTC ATT GAG GTG GGG CCA ATT TTA ATC ATA AGC CTT
ACG ATT TAC AAT CAG TAA CTC CAC CCC GGT TAA AAT TAG TAT TCG GAA
C * M L V I E V G P I L I I S L>

5620      5630      5640      5650      5660
*         *         *         *         *
AAT AAG ATT TTT CTA AGA AAT GTG AAA TAG AAC AAT TTT CAT CTA ATT
TTA TTC TAA AAA GAT TCT TTA CAC TTT ATC TTG TTA AAA GTA GAT TAA
N K I F L R N V K * N N F H L I>

5670      5680      5690      5700      5710
*         *         *         *         *
CCA TTT ACT TTT AGA TGA ATG GCA TTG TGA ATG CCA TTC TTT TAA TGA
GGT AAA TGA AAA TCT ACT TAC CGT AAC ACT TAC GGT AAG AAA ATT ACT
P F T F R * M A L * M P F F * *>

5720      5730      5740      5750      5760
*         *         *         *         *
ATT TCA AGA GAA TTC TCT GGT TTT CTG TGT AAT TCC AGA TGA GTC ACT
TAA AGT TCT CTT AAG AGA CCA AAA GAC ACA TTA AGG TCT ACT CAG TGA
I S R E F S G F L C N S R * V T>

5770      5780      5790      5800
*         *         *         *
GTA ACT CTA GAA GAT TAA CCT TCC AGC CAA CCT ATT TTC CTT TCC CTT
CAT TGA GAT CTT CTA ATT GGA AGG TCG GTT GGA TTA AAG GAA AGG GAA
V T L E D * P S S Q P I F L S L>

5810      5820      5830      5840      5850
*         *         *         *         *
GTC TCT CTC ATC CTC TTT TCC TTC CTT CTT TCC TTT CTC TTC TTT TAT
CAG AGA GAG TAG GAG AAA AGG AAG GAA GAA AGG AAA GAG AAG AAA ATA
V S L I L F S F L L S F L F F Y>

5860      5870      5880      5890      5900
*         *         *         *         *
CTC CAA GGT TAA TCA GGA AAA ATA GCT TTT GAC AGG GGA AAA AAC TCA
GAG GTT CCA ATT AGT CCT TTT TAT CGA AAA CTG TCC CCT TTT TTG AGT
L Q G * S G K I A F D F G K N S>

5910      5920      5930      5940      5950
*         *         *         *         *
ATA ACT AGC TAT TTT TGA CCT CCT GAT CAG GAA CTT TAG TTG AAG CGT
TAT TGA TCG ATA AAA ACT GGA GGA CTA GTC CTT GAA ATC AAC TTC GCA

```

WO 98/55615

PCT/US98/11773

17/39

FIGURE 5M

I T S V F * P P D Q E L * L K R>

5960 5970 5980 5990 6000

AAA TCT AAA GAA ACA TTT TCT CTG AAA TAT ATT ATT AAG GGC AAT GGA
 TTT AGA TTT CTT TGT AAA AGA GAC TTT ATA TAA TAA TTC CCG TTA CCT
 K S K E T F S L K Y I I K G N G>

6010 6020 6030 6040

GAT AAA TTA ATA GTA GAT GTG GTT CCC AGA AAA TAT AAT CAA AAT TCA
 CTA TTT AAT TAT CAT CTA CAC CAA GGG TCT TTT ATA TTA GTT TTA AGT
 D K L I V D V V P R K Y N Q N S>

6050 6060 6070 6080 6090

AAG ATT TTT TTT GTT TCT GTA ACT GGA ACT AAA TCA AAT GAT TAC TAG
 TTC TAA AAA AAA CAA AGA CAT TGA CCT TGA TTT AGT TTA CTA ATG ATC
 K I F F V S V T G T K S N D Y *>

6100 6110 6120 6130 6140

TGT TAA TAG TAG ATA ACT TGT TTT TAT TGT TGG TGC ATA TTA GTA TAA
 ACA ATT ATC ATC TAT TGA ACA AAA ATA ACA ACC ACC TAT AAT CAT ATT
 C * * I T C F Y C W C I L V *>

6150 6160 6170 6180 6190

CTG TGG GGT AGG TCG GGG AGA GGG TAA GGG AAT AGA TCA CTC AGA TGT
 GAC ACC CCA TCC AGC CCC TCT CCC ATT CCC TTA TCT AGT GAG TCT ACA
 L W G R S G R G * G N R S L R C>

6200 6210 6220 6230 6240

ATT TTA GAT AAG CTA TTT AGC CTT TGA TGG AAT CAT AAA TAC AGT GAA
 TAA AAT CTA TTC GAT AAA TCG GAA ACT ACC TTA GTA TTT ATG TCA CTT
 I L D K L F S L * W N H K Y S =>

6250 6260 6270 6280

TAC AAT CCT TTG CAT TGT TAA GGA GGT TTT TTG TTT TTA AAT GGT GGG
 ATG TTA GGA AAC GTA ACA ATT CCT CCA AAA AAC AAA AAT TTA CCA CCC
 Y N P L H C * G G F L F L N G G>

6290 6300 6310 6320 6330

TCA AGG AGC TAG TTT ACA GGC TTA CTG TGA TTT AAG CAA ATG TGA AAA
 AGT TCC TCG ATC AAA TGT CCG AAT GAC ACT AAA TTT GTT TAC ACT TTT
 S R S * F T G L L * F K Q M * K>

6340 6350 6360 6370 6380

GTG AAA CCT TAA TTT TAT CAA AAG AAA TTT CTG TAA ATG GTA TGT CTC
 CAC TTT GGA ATT AAA ATA GTT TTC TTT AAA GAC ATT TAC CAT ACA GAG
 V K P * F Y Q K K F L * M V C L>

6390 6400 6410 6420 6430

CTT AGA ATA CCC AAA TCA TAA TTT TAT TTG TAC ACA CTG TTA GCG GCT
 GAA TCT TAT GGG TTT AGT ATT AAA ATA AAC ATG TGT GAC AAT CCC CGA
 L R I P K S * F Y L Y T L L G A>

FIGURE 5N

```

      6440      6450      6460      6470      6480
      *         *         *         *         *
CAT CTC ATG TAG GCA GAG TAT AAA GTA TTA CCT TTT GGA ATT AAA AGC
GTA GAG TAC ATC CGT CTC ATA TTT CAT AAT GGA AAA CCT TAA TTT TCG
H   L   M   *   A   E   Y   K   V   L   P   F   G   I   K   S>

      6490      6500      6510      6520
      *         *         *         *
CAC TGA CTG TTA TAA AGT ATA ACA ACA CAC ATC AGG TTT TAA AAA GCC
GTG ACT GAC AAT ATT TCA TAT TGT TGT GTG TAG TCC AAA ATT TTT CGG
H   *   L   L   *   S   I   T   T   H   I   R   F   *   K   A>

6530      6540      6550      6560      6570
      *         *         *         *         *
TTG AAT GGC CCT TGT CTT AAA AAG AAA TTA GGA GCC AGG TGC GGT GGC
AAC TTA CCG GGA ACA GAA TTT TTC TTT AAT CCT CGG TCC ACG CCA COG
L   N   G   P   C   L   K   K   K   L   G   A   R   C   G   G>

6580      6590      6600      6610      6620
      *         *         *         *         *
ACG TGC CTG TAG TCC CAG CTC CTT GGG AGG CTG AGA CAG GAG GAT TCC
TGC ACG GAC ATC AGG GTC GAG GAA CCC TCC GAC TCT GTC CTC CTA AGG
T   C   L   *   S   Q   L   L   G   R   L   F   Q   E   D   S>

      6630      6640      6650      6660      6670
      *         *         *         *         *
TTG AGC CCT GGA GTT TGA GTC CAG CCT GGG TGA CAT AGC AAG ACC CTG
AAC TCG GGA CCT CAA ACT CAG GTC GGA CCC ACT GTA TCG TTC TGG GAC
L   S   P   G   V   *   V   Q   P   G   *   H   S   K   T   L>

      6680      6690      6700      6710      6720
      *         *         *         *         *
TCT TAA AAG AAA AAT GGG AAG AAA GAC AAG GTA ACA TGA AGA AAG AAG
AGA ATT TTC TTT TTA CCC TTC TTT CTG TTC CAT TGT ACT TCT TTC TTC
S   *   K   K   N   G   K   K   D   K   V   T   *   R   K   K>

      6730      6740      6750      6760
      *         *         *         *
AGA TAC CTA GTA TGA TGG AGC TGC AAA TTT CAT GGC AGT TCA TGC AGT
TCT ATG GAT CAT ACT ACC TCG ACG TTT AAA GTA CCG TCA AGT ACG TCA
R   Y   L   V   *   N   S   C   K   F   H   G   S   S   C   S>

6770      6780      6790      6800      6810
      *         *         *         *         *
CGG TCA AGA GGA GGA TTT TGT TTT GTA GTT TGC AGA TGA GCA TTT CTA
GCC AGT TCT CCT CCT AAA ACA AAA CAT CAA ACG TCT ACT CGT AAA GAT
R   S   R   G   G   F   C   F   V   V   C   R   *   A   F   L>

6820      6830      6840      6850      6860
      *         *         *         *         *
AAG CAT TTT CCC TTG CTG TAT TTT TTT GTA TTA TAA ATT ACA TTG GAC
TTC GTA AAA GCG AAC GAC ATA AAA AAA CAT AAT ATT TAA TGT AAC CTG
K   H   F   P   L   L   Y   F   F   V   L   *   I   T   L   D>

      6870      6880      6890      6900      6910
      *         *         *         *         *
TTC ATA TAT ATA ATT TTT TTT TAC ATT ATA TGT CTC TTG TAT GTT TTG
AAG TAT ATA TAT TAA AAA AAA ATG TAA TAT ACA GAG AAC ATA CAA AAC
F   I   Y   I   I   F   F   Y   I   I   C   L   L   Y   V   L>

      6920      6930      6940      6950      6960
      *         *         *         *         *

```

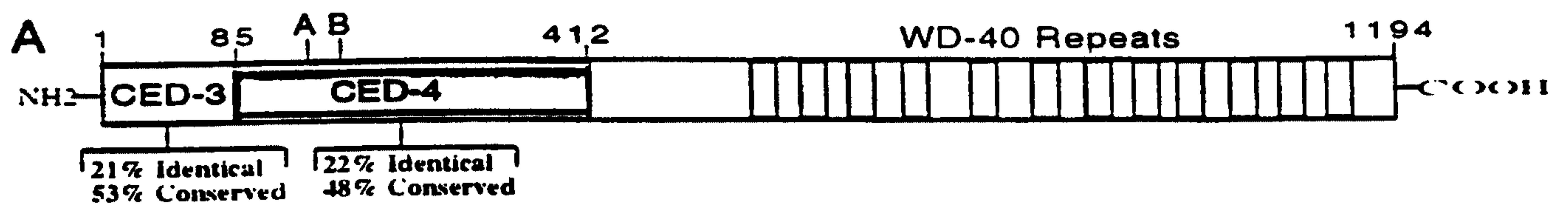
FIGURE 50

AAA	CTC	TTG	TAT	TTA	TGA	TAT	AGC	TTA	TAT	GAT	TTT	TTT	GCC	TTG	GTA
TTT	GAG	AAC	ATA	AAT	ACT	ATA	TCG	AAT	ATA	CTA	AAA	AAA	CGG	AAC	CAT
K	L	L	Y	L	*	Y	S	L	Y	D	F	F	A	L	V>
		6970				6980				6990					7000
		*				*				*					*
TAC	ATT	TTA	AAA	TAT	GAA	TTT	AAA	AAA	TTT	TTG	TAA	AAA	TAA	AAT	TCA
ATG	TAA	AAT	TTT	ATA	CTT	AAA	TTT	TTT	AAA	AAC	ATT	TTT	ATT	TTA	AGT
Y	I	L	K	Y	E	F	K	K	F	L	*	K	*	N	S>
7010			7020				7030								7040
*			*			*	*			*					*
CRA	AAT	TGT	TTT	GAA	AAA	CAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA	AAA
GTT	TTA	ACA	AAA	CTT	TTT	GTT	TTT	TTT	TTT	TTT	TTT	TTT	TTT	TTT	TTT
Q	N	C	F	E	K	Q	K	K	K	K	K	K	K	K	K>

New 7810

FIGURE 6

MDAKARNCLLOHREALEKDIKTSYIMDHMISDGFLTISEE	40
EKVRNEPTQQORAAMLIKMIKKDND SYVSFYNALLHEGY	80
KDLAALLHDGIPVVSSSSSVRTVLC EGGVPQRPVVFVTRKK	120
LVNAIOOKLSKLGEPGWVTIHGMAGCGKSVLAAEAVRDH	160
SLLEGCFPGGVHWSVGKQDKSGLLMKLQNLCTRLDODES	200
FSORLP LNIEEAKDRLRILMLRKHPRSLLILDDVWDSWVL	240
KAFDSQCCQILLTTRDKSVTDSVMGPKYVVPVVESSLGKEKG	280
LEILSLFVNMKKADLPEQAH SIIKECKGSPLVVSIGALL	320
RDFPNRWEYYLKQLQNKQFKRIRKSSSYDYEALDEAMSIS	360
VEMLREDIKDYTDLSILQKDVKVPTKVL CILWDMETEEV	400
EDILQEFVNKSLLFCDRNGKSF RYYLHDLQVDFLTEKNCS	440
QLQDLHKKIITQFQRYHQPHTLS PDQEDCMYWYNFLAYHM	480
ASAKMHKELCALMFSLDWIKAKTEL VGPAHLIHEFVEYRH	520
ILDEKDCAVSENFQEFSLNGHLLGRQPFPNIVQLGLCEP	560
ETSEVYQQA KLOAKQEV DNGMLYLEWINKKNITNLSRLVV	600
RPHTDVYHACESE DGORIASCGADKTLOVFKAETGEKLL	640
EIKAH EDEVLCCEAESTDDRFIATCSVDKVKIWN SMTGEL	680
VHTYDEHSEOVNCCHFTNSSHLLLATGSSDCFLKLWDLN	720
QKECRNTMFGHTNSVNHCRFSPDDKLLASCSADGTLKLWD	760
ATSANERKSINVKOFFLNLEDPOEDMEVIVKCCSWSADGA	800
RIMVAAKNKIFLWN TDSRSKVADCRGHLSWVHGVMFSPDG	840
SSFLTSSDDOTIRLWETKKVCKNSAVMLKQEV DVVFOENE	880
VMVLAVDHIRRLQLINGRTGQIDYLTEAQVSCCCLS PHLO	920
VTARGDENGATELLELVNNRIFQSRF DHKKT VWHIOFTAD	960
EKTLISSSDAEIOVWNWQLDKCIFLRGHOETVKDFRLLK	1000
NSRLLSNSFDGTVKVWN IITGNKEKDFVCHOGTVLSCDIS	1040
HDA TKFSSTSA DKTAKIWSFDLLLPLHELRGHNGCVRC SA	1080
FSVDSTLLATGDDNGEIRIWNVSN GELLHLCAPLSEEGAA	1120
THGGWVTDLCESPDGKMLISAGGYIKWNNVVTGESSQTFY	1160
TNGTNLKKIHVSPDFKTYVTVDNLGILYILQ TLE	1194

FIGURE 7

WO 98/55615

22/39

PCT/US98/11773

FIGURE 8

Apaf-1 1 MDAKARNCLLOHREAL-EKDIKTSYIMDHMISDGFLTISEE
CED-3 1 MMRODRRSLLERNIMMFSSHLKVDEILEVLIKQVLNSDNG

Apaf-1 40 EKVRNEPTQQORAAMLIKMILKKDND SYVSFYNALLHEGYK
CED-3 41 DMINSCGTVREKRREIVKAVORRGDVAEDAEYDALRSTGHE

Apaf-1 80 DLAALL Apaf-1 85
CED-3 81 GLAEVL CED-3 86

FIGURE 9

Apaf-1 92 PVVSSSSV R T V L C E G G V P Q R P V V F V I R K K L V N A I Q Q K L S K
 CED-4 108 P Q F S R Q M L D R K I L L G N V P K O M T C Y I - R E Y H V D R V I K K L D E

Apaf-1 132 L - K G E P G W V I I H G M A G C G K S V L A A E A V R D H S L L E G C F P G G
 CED-4 147 M C D L D S F F L F L H G R A G S G K S V I A S O A L S K S D Q L I G I N Y D S

Apaf-1 171 V H W V - S V G K Q D K S G L L M K L Q N L C T R L D Q D E S F S - Q R L P L N
 CED-4 188 I V W L K D S G T A P K S I E D L F T D I L L M L K S E D D L L N F P S V E H V

Apaf-1 209 I E E A K D R L R I L M L R K H P R S L L I L D D V W D S W V L K - A F D S Q C
 CED-4 227 T S V V L K R M I C N A L I D R P N I L F V F D D V V O E E T I R W A Q E L R L

Apaf-1 248 Q I L L T T R D K S V T D S V M G P K Y V V P V E S S I G K E K G L E I L S L F
 CED-4 265 R C L V T T R D V E L S N A A S O T C E F I E V - T S L E I D E C Y D F L E A Y

Apaf-1 288 - V N M K K A D L P E Q A - H S I I K E C K G S P - - L V V S L I G A L L R D F
 CED-4 306 G M P M P V G E K E E D V L N K T I E L S S G N P A T L M M F E K S C E P K T F

Apaf-1 324 P N R W - E Y Y L K Q L Q N K Q F K R I R K S S S Y D Y E A L D E A M S I S V E
 CED-4 346 - E K M A Q L N N K - L E S R G L V G V E C I T P Y S Y K S L A M A L Q R C V E

Apaf-1 363 M L R E D I K D Y Y I D L S I L Q K D V K V P T K V L - C I L W - D M E T E E V
 CED-4 384 V L S D E D R S A L A F A V M P P G V D I P V K L W S C V I P V D I C S N E E

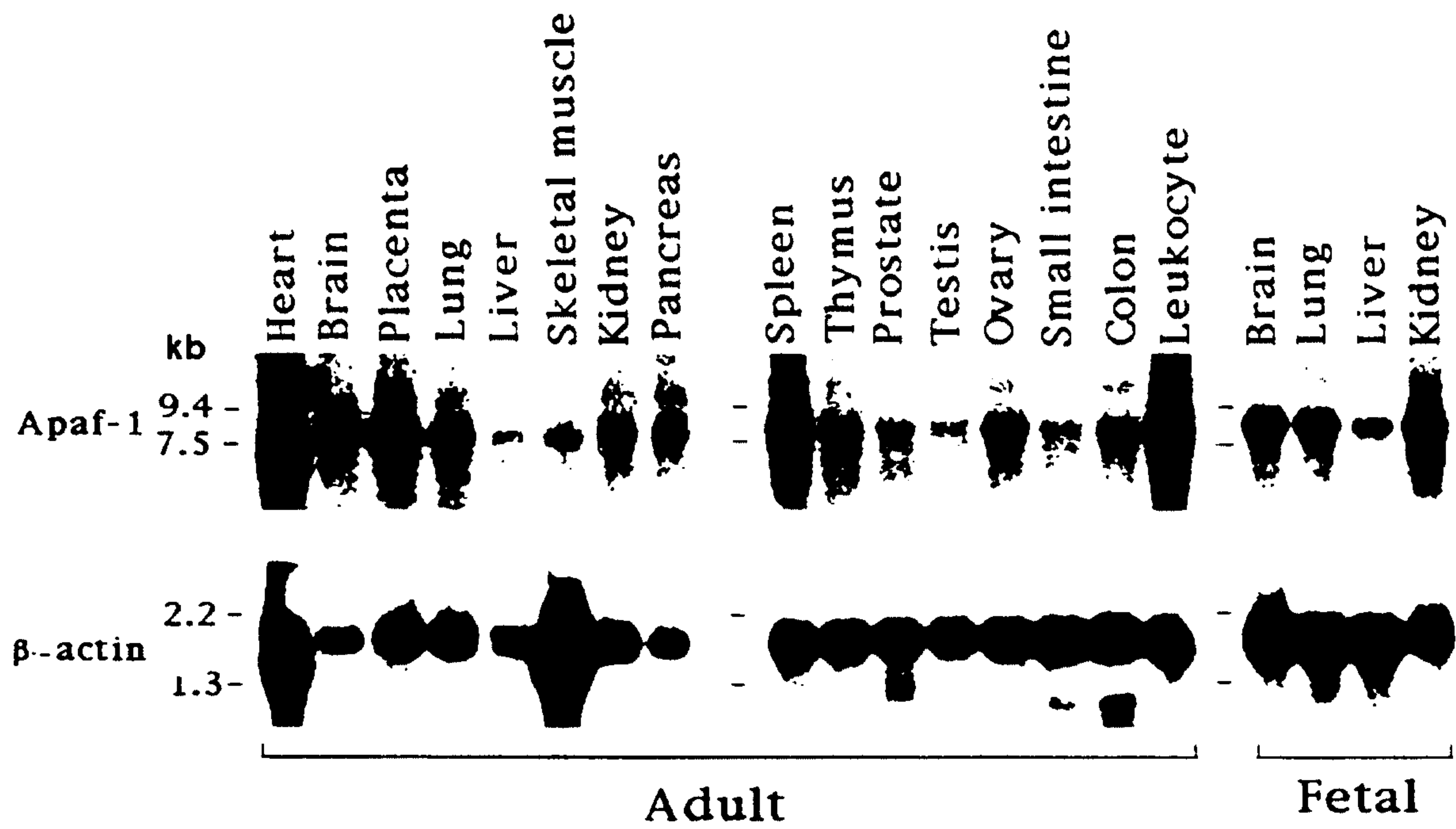
Apaf-1 401 E D I L Q E F V N K S L Apaf-1 412
 CED-4 424 E Q - L D D E V A D R L CED-4 435

WO 98/55615

PCT/US98/11773

24/39

FIGURE 10



WO 98/55615

25/39

PCT/US98/11773

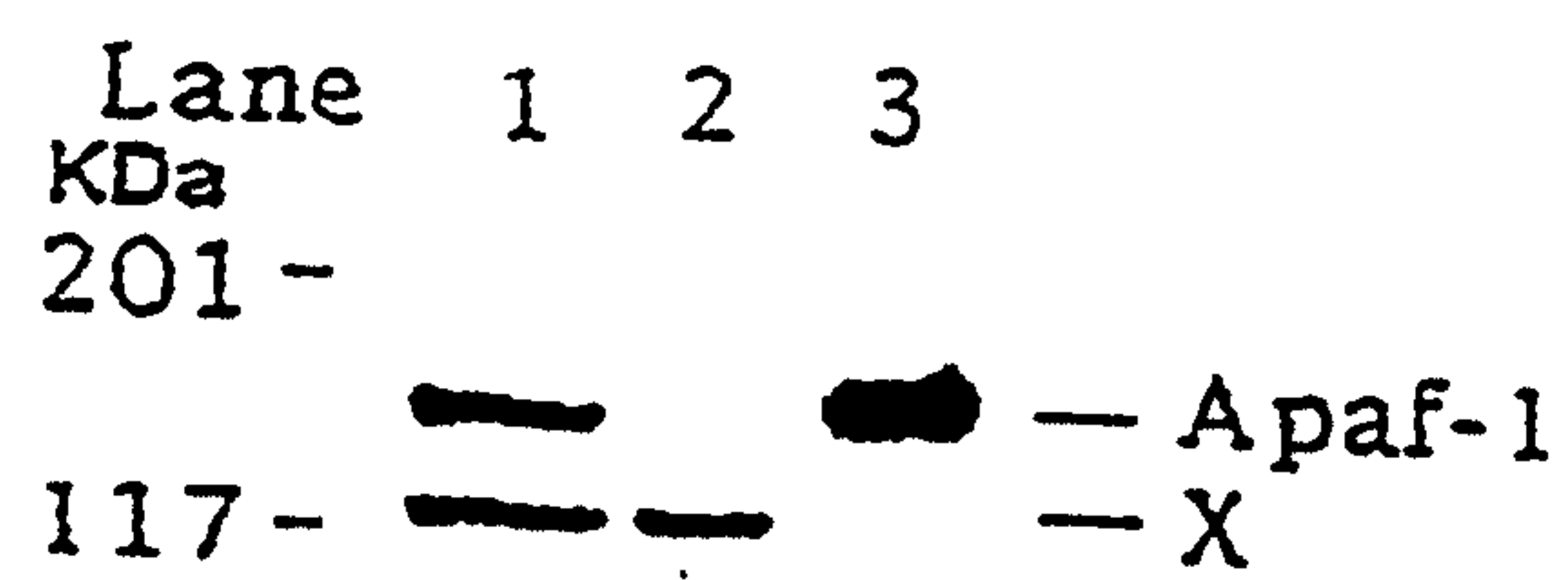
FIGURE 11

FIGURE 12

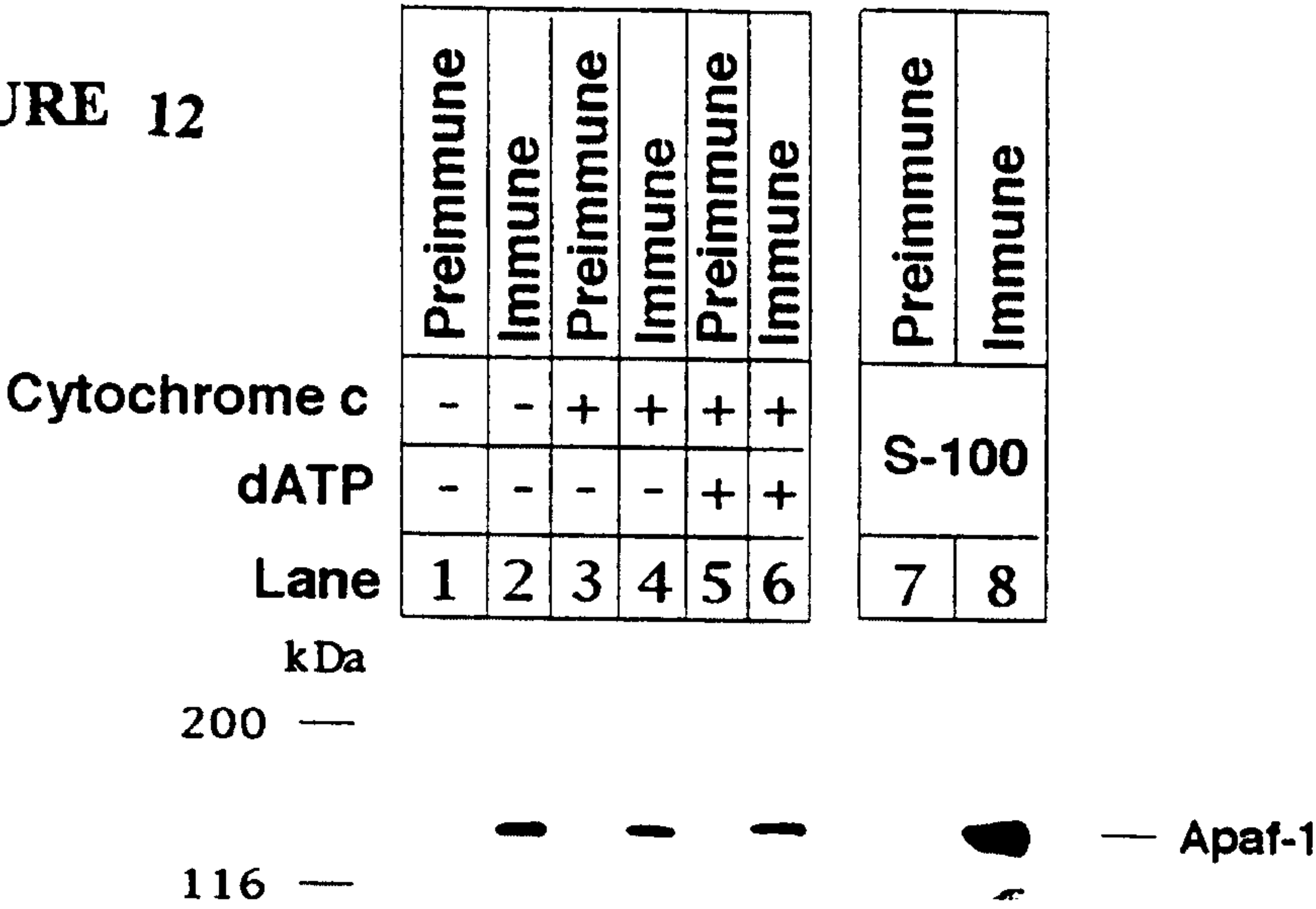


FIGURE 13

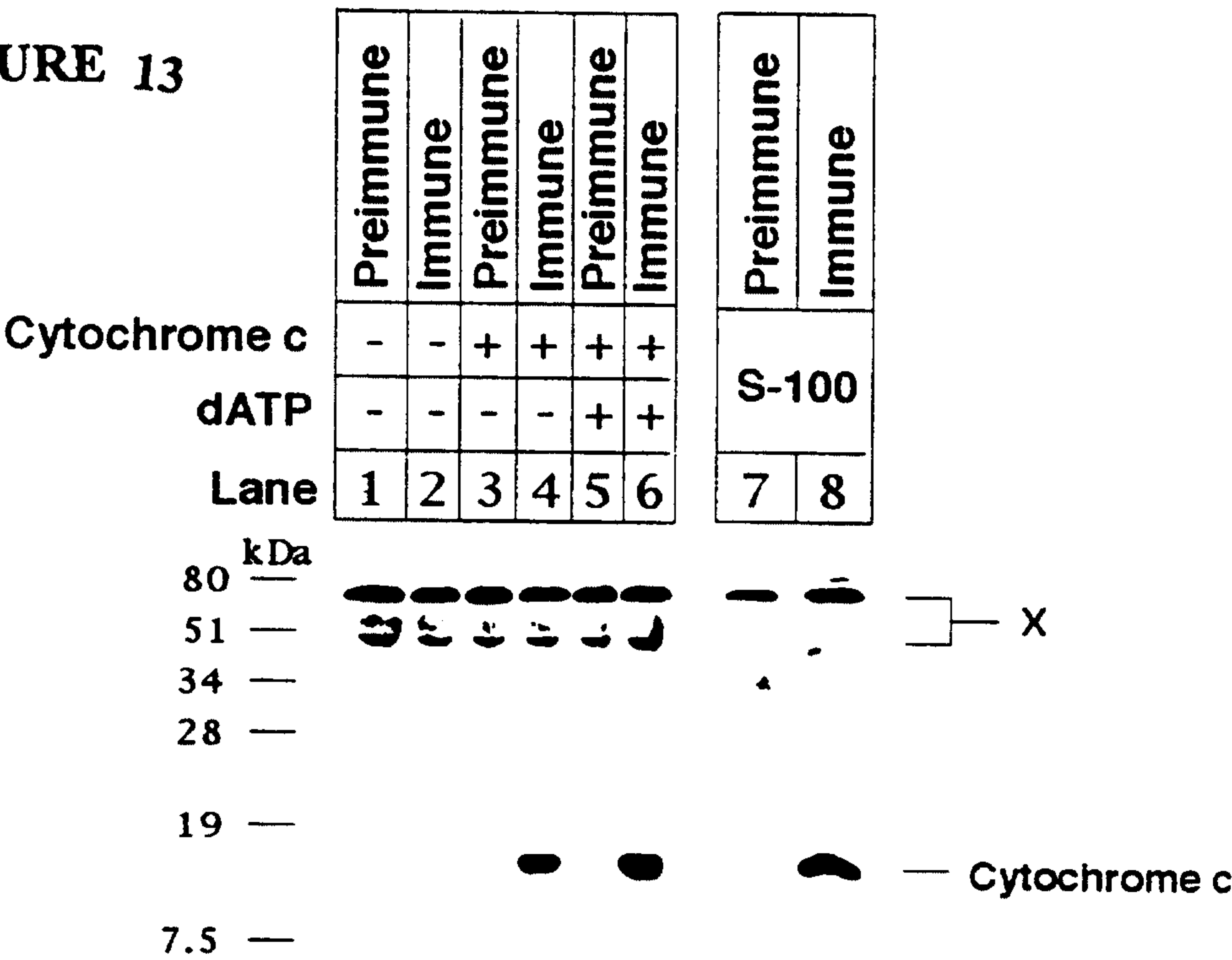


FIGURE 14

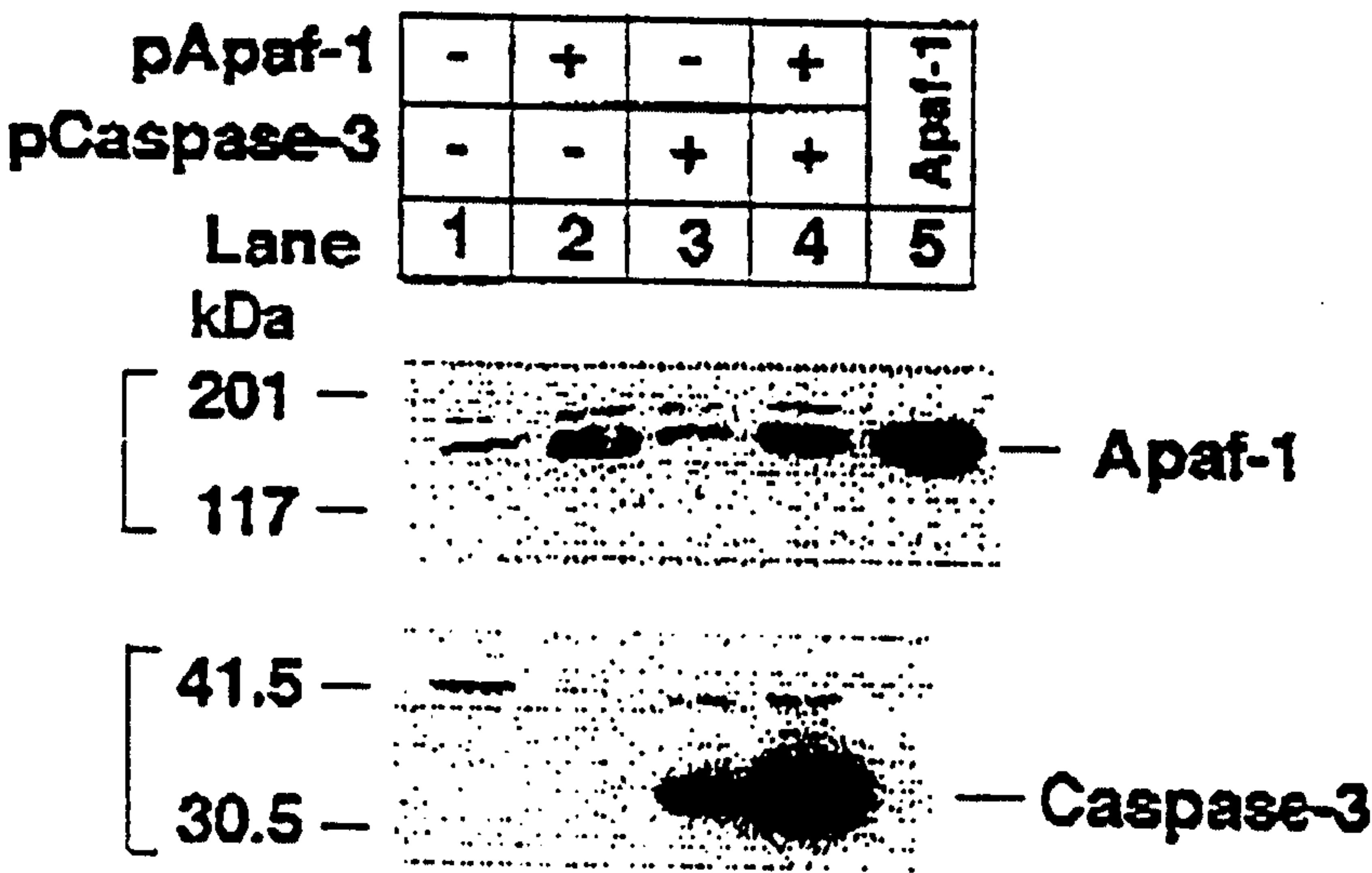


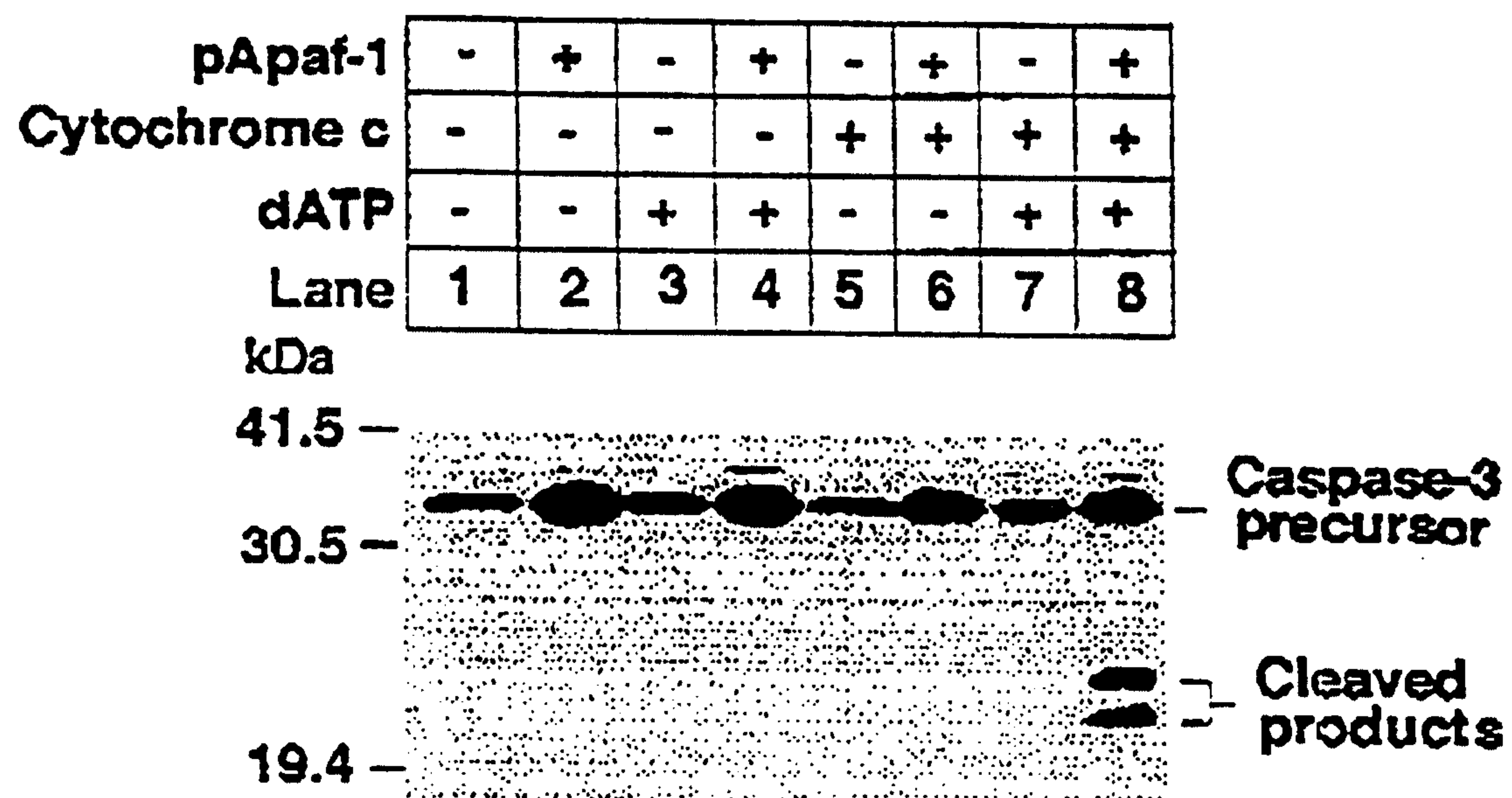
FIGURE 15

FIGURE 16A

Apaf-1L

Contig Length:	7810 bases
Average Length/Sequence:	660 bases
Total Sequence Length:	42284 bases
Top Strand:	34 sequences
Bottom Strand:	30 sequences
Total:	64 sequences

10 20 30 40
AAGAAGAGGTAGCGAGTGGACGTGACTGCTCTATCCCGGG 40
CAAAAGGGATAGAACCAGAGGTGGGGAGTCTGGGCAGTCG 80
GCGACCCGCGAAGACTTGAGGTGCCGCAGCGGCATCCGGA 120
GTAGCGCCGGGCTCCCTCCGGGGTGCAGCCGCGTCGGGG 160
GAAGGGCGCCACAGGCCGGGAAGACCTCCTCCCTTTGTGT 200
210 220 230 240
CCAGTAGTGGGGTCCACCGGAGGGCGGCCCGTGGGCCGGG 240
CCTCACCGCGGCGCTCCGGGACTGTGGGGTCAGGCTGCGT 280
TGGGTGGACGCCCACCTCGCCAACCTTCGGAGGTCCCTGG 320
GGGTCTTCGTGCGCCCCGGGGCTGCAGAGATCCAGGGGAG 360
GCGCCTGTGAGGCCCGGACCTGCCCGGGGCGAAGGGTAT 400
410 420 430 440
GTGGCGAGACAGAGCCCTGCACCCCTAATTCCCGGTGGAA 440
AACTCCTGTTGCCGTTTCCCTCCACCGGCCTGGAGTCTCC 480
CAGTCTTGTCGCGGAGTGCCGCCCTCCCCACTAAGACCT 520
AGGCGCAAAGGCTTGGCTCATGGTTGACAGCTCAGAGAGA 560
GAAAGATCTGAGGGAAGATGGATGCAAAAGCTCGAAATTG 600
610 620 630 640
TTTGCTTCAACATAGAGAAGCTCTGGAAAAGGACATCAAG 640
ACATCCTACATCATGGATCACATGATTAGTGATGGATTTT 680
TAACAATATCAGAAGAGGAAAAAGTAAGAAATGAGCCAC 720
TCAACAGCAAAGAGCAGCTATGCTGATTAAAATGATACTT 760
AAAAAAGATAATGATTCCTACGTATCATTCTACAATGCTC 800

Start
548

FIGURE 16B

810 820 830 840
TACTACATGAAGGATATAAAGATCTTGCTGCCCTTCTCCA 840
TGATGGCATTCTCTGTTGTCTCTTCTTCCAGTGGTAAAGAT 880
TCAGTTAGTGGAATAACTTCGTATGTAAGGACAGTCCTGT 920
GTGAAGGTGGAGTACCACAGAGGCCAGTTGTTTTTGTAC 960
AAGGAAGAAGCTGGTGAATGCAATTCAGCAGAAGCTCTCC 1000
1010 1020 1030 1040
AAATTGAAAGGTGAACCAGGATGGGTCACCATACATGGAA 1040
TGGCAGGCTGTGGGAAGTCTGTATTAGCTGCAGAAGCTGT 1080
TAGAGATCATTCCCTTTTAGAAGGTTGTTTCCCAGGGGGA 1120
GTGCATTGGGTTTCAGTTGGGAAACAAGACAAATCTGGGC 1160
TTCTGATGAAACTGCAGAATCTTTGCACACGGTTGGATCA 1200
1210 1220 1230 1240
GGATGAGAGTTTTTCCCAGAGGCTTCCACTTAATATTGAA 1240
GAGGCTAAAGACCGTCTCCGCATTCTGATGCTTCGCAAAC 1280
ACCCAAGGTCTCTCTTGATCTTGGATGATGTTTGGGACTC 1320
TTGGGTGTTGAAAGCTTTTGACAGTCAGTGTCAGATTCTT 1360
CTTACAACCAGAGACAAGAGTGTTACAGATTCAGTAATGG 1400
1410 1420 1430 1440
GTCCTAAATATGTAGTCCCTGTGGAGAGTTCCTTAGGAAA 1440
GGAAAAAGGACTTGAAATTTTATECCTTTTGTAAATATG 1480
AAGAAGGCAGATTTGCCAGAACAAGCTCATAGTATTATAA 1520
AAGAATGTAAAGGCTCTCCCCTTGTAGTATCTTTAATTGG 1560
TGCACTTTTACGTGATTTTCCCAATCGCTGGGAGTACTAC 1600
1610 1620 1630 1640
CTCAAACAGCTTCAGAATAAGCAGTTTAAGAGAATAAGGA 1640
AATCTTCGTCTTATGATTATGAGGCTCTAGATGAAGCCAT 1680
GTCTATAAGTGTTGAAATGCTCAGAGAAGACATCAAAGAT 1720
TATTACACAGATCTTTCCATCCTTCAGAAGGACGTTAAGG 1760
TGCCTACAAAGGTGTTATGTATTCTCTGGGACATGGAAAC 1800

WO 98/55615

PCT/US98/11773

31/39

FIGURE 16C

1810 1820 1830 1840
TGAAGAAGTTGAAGACATACTGCAGGAGTTTGTAATAAG 1840
TCTCTTTTATTCTGTGATCGGAATGGAAAGTCGTTTCGTT 1880
ATTATTTACATGATCTTCAAGTAGATTTTCTTACAGAGAA 1920
GAATTGCAGCCAGCTTCAGGATCTACATAAGAAGATAATC 1960
ACTCAGTTTCAGAGATATCACCAGCCGCATACTCTTTCAC 2000

2010 2020 2030 2040
CAGATCAGGAAGACTGTATGTATTGGTACAACCTTTCTGGC 2040
CTATCACATGGCCAGTGCCAAGATGCACAAGGAACTTTGT 2080
GCTTTAATGTTTTCCCTGGATTGGATTAAAGCAAAAACAG 2120
AACTTGTAGGCCCTGCTCATCTGATTCATGAATTTGTGGA 2160
ATACAGACATATACTAGATGAAAAGGATTGTGCAGTCAGT 2200

2210 2220 2230 2240
GAGAATTTTCAGGAGTTTTTATCTTTAAATGGACACCTTC 2240
TTGGACGACAGCCATTTCTTAATATTGTACAACCTGGGTCT 2280
CTGTGAGCCGGAACTTCAGAAGTTTATCAGCAAGCTAAG 2320
CTGCAGGCCAAGCAGGAGGTCGATAATGGAATGCTTTACC 2360
TGGAATGGATAAACAAAAAAACATCACGAATCTTTCCCG 2400

2410 2420 2430 2440
CTTAGTTGTCCGCCCCCACACAGATGCTGTTTACCATGCC 2440
TGCTTTTCTGAGGATGGTCAGAGAATAGCTTCTTGTGGAG 2480
CTGATAAAACCTTACAGGTGTTCAAAGCTGAAACAGGAGA 2520
GAAACTTCTAGAAATCAAGGCTCATGAGGATGAAGTGCTT 2560
TGTTGTGCATTCTCTACAGATGACAGATTTATAGCAACCT 2600

2610 2620 2630 2640
GCTCAGTGGATAAAAAAGTGAAGATTTGGAATTCTATGAC 2640
TGGGGAAGTAGTACACACCTATGATGAGCACTCAGAGCAA 2680
GTCAATTGCTGCCATTTACCAACAGTAGTCATCATCTTC 2720
TCTTAGCCACTGGGTCAAGTGACTGCTTCCTCAAACCTTG 2760
GGATTTGAATCAAAAAGAATGTCGAAATACCATGTTTGGT 2800

FIGURE 16D

2810 2820 2830 2840
CATACAAATTCAGTCAATCACTGCAGATTTTCACCAGATG 2840
ATAAGCTTTTGGCTAGTTGTTTCAGCTGATGGAACCTTAAA 2880
GCTTTGGGATGCGACATCAGCAAATGAGAGGAAAAGCATT 2920
AATGTGAAACAGTTCTTCCTAAATTTGGAGGACCCTCAAG 2960
AGGATATGGAAGTGATAGTGAAGTGTTGTTTCGTGGTCTGC 3000
3010 3020 3030 3040
TGATGGTGCAAGGATAATGGTGGCAGCAAAAAATAAAATC 3040
TTTTTGTGGAATACAGACTCACGTTCAAAGGTGGCTGATT 3080
GCAGAGGACaTTTAAGTTGGGTTTCATGGTGTGATGTTTTTC 3120
TCCTGATGGATCATCATTTTTTGACATCTTCTGATGACCAG 3160
ACAATCAGGCTCTGGGAGACAAAGAAAGTATGTAAGAAGT 3200
3210 3220 3230 3240
CTGCTGTAATGTTAAAGCAAGAAGTAGATGTTGTGTTTCA 3240
AGAAAATGAAGTGATGGTCCTTGCAGTTGACCATATAAGA 3280
CGTCTGCAACTCATTAAATGGAAGAACAGGTCAGATTGATT 3320
ATCTGACTGAAGCTCAAGTTAGCTGCTGTTGCTTAAGTCC 3360
ACATCTTCAGTACATTGCATTTGGAGATGAAAATGGAGCC 3400
3410 3420 3430 3440
ATTGAGATTTTAGAAGCTTGTAACAATAGAATCTTCCAGT 3440
CCAGGTTTCAGCACAAGAAAAGTGTATGGCACATCCAGTT 3480
CACAGCCGATGAGAAGACTCTTATTTCAGTTCTGATGAT 3520
GCTGAAATTCAGGTATGGAATTGGCAATTGGACAAATGTA 3560
TCTTTCTACGAGGCCATCAGGAAACAGTGAAAGAcTTTAG 3600
3610 3620 3630 3640
ACTCTTGAAAAATTCAAGACTGCTTTCTTGGTCATTTGAT 3640
GGAACAGTGAAGGTATGGAATATTACTGGAAATAAAG 3680
AAAAAGACTTTGTCTGTCACCAGGGTACAGTACTTTCTTG 3720
TGACATTTCTCACGATGCTACCAAGTTTTTCATCTACCTCT 3760
GCTGACAAGACTGCAAAGATCTGGAGTTTTGATCTCCTTT 3800

FIGURE 16E

3810 3820 3830 3840
TGCCACTTCATGAATTGAGGGGCCACAACGGCTGTGTGCG 3840
CTGCTCTGCCTTCTCTGTGGACAGTACCCTGCTGGCAACG 3880
GGAGATGACAATGGAGAAATCAGGATATGGAATGTCTCAA 3920
ACGGTGAGCTTCTTCATTTGTGTGCTCCGCTTTCAGAAGA 3960
AGGAGCTGCTACCCATGGAGGCTGGGTGACTGACCTTTGC 4000
4010 4020 4030 4040
TTTTCTCCAGATGGCAAATGCTTATCTCTGCTGGAGGAT 4040
ATATTAAGTGGTGGAAACGTTGTCACTGGGGAATCCTCACA 4080
GACCTTCTACACAAATGGAACCAATCTTAAGAAAATACAC 4120
GTGTCCCCTGACTTCAAAACATATGTGACTGTGGATAATC 4160
TTGGTATTTTATATATTTTACAGACTTTAGAATAAAATAG 4200
4210 4220 4230 STOP 4240
TTAAGCATTAAATGTAGTTGAACTTTTTAAATTTTTGAATT 4240
GGAAAAAAATTCTAATGAAACCCTGATATCAACTTTTTAT 4280
AAAGCTCTTAATTGTTGTGCAGTATTGCATTCATTACAAA 4320
AGTGTTTGTGGTTGGATGAATAATATTAATGTAGCTTTTT 4360
CCCAAATGAACATACCTTTAATCTTGTTTTTCATGATCAT 4400
4410 4420 4430 4440
CATTACAGTTTGTCTTAGGATGCAAATGAAAATGTGAA 4440
TACATACCTTGTTGTA CTGTTGGTAAAATTCTGTCTTGAT 4480
GCATTCAAAATGGTTGACATAATTAATGAGAAGAATTTGG 4520
AAGAAATTGGTATTTTAACTGTCTGTATTTATTACTGT 4560
TATGCAGGCTGTGCCTCAGGGTAGCAGTGGCCTGCTTTTT 4600
4610 4620 4630 4640
GAACCACACTTACCCCAAGGGGGTTTTGTTCTCCTAAATA 4640
CAATCTTAGAGGTTTTTTGCACTCTTTAAATTTGCTTTAA 4680
AAATATTGTGTCTGTGTGCATAGTCTGCAGCATTTTCCTTT 4720
AATTGACTCAATAAGTGAGTCTTGGATTTAGCAGGCCCCC 4760
CCACCTTTTTTTTTTTGTTTTTGGAGACAGAGTCTTGCTTT 4800

4192 = end

FIGURE 16F

4810 4820 4830 4840
GTTGCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACC 4840
ACAATCGCTGCCTCCTGGGTTCAAGCAATTCTCCTGCCTC 4880
AGCCTCCCGAGTAGCTGGGACTACAGGTGTGCGCACATGC 4920
CAGGCTAATTTTTGTATTTTAGTAGAGACGGGGTTTCAC 4960
CATGTTGGCCGGGATGGTCTCGATCTCTTGACCTCATGAT 5000
5010 5020 5030 5040
CTACCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGCG 5040
TGAGCCACCGTGCCTGGCCAGGCCCTTCTCTTTTAATGG 5080
AGACAGGGTCTTGCACTATCACCCAGGCTGGAGTGCAGTG 5120
GCATAATCATACCTCATTGCAGCCTCAGACTCCTGGGTTC 5160
AAGCAATCCTCCTGCCTCAGCCTCCCAAGTAGCTGAGACT 5200
5210 5220 5230 5240
GCAGGCACGAGCCACCACACCCAGCTAATTTTAAAGTTTT 5240
CTTGTAGAGACAGGGTCTCACTATGTTGTCTAGGCTGGTC 5280
TTGAACTCTTGGCCTCAAGTAATCCTCCTGCCTCAGCCTC 5320
CCAAAGTGTTGGGATTGCAGATATGAGCCACTGGCCTGGC 5360
CTTCAGCAGTTCTTTTTGTGAAGTAAACTTGTATGTTGG 5400
5410 5420 5430 5440
AAAGAGTAGATTTTATTGGTCTACCCTTTTCTCACTGTAG 5440
CTGCTGGCAGCCCTGTGCCATATCTGGACTCTAGTTGTCA 5480
GTATCTGAGTTGGACACTATTCCTGCTCCCTCTTGTTTCT 5520
TACATATCAGACTTCTTACTTGAATGAAACCTGATCTTTC 5560
CTAATCCTCACTTTTTTCTTTTTTAAAAAGCAGTTTCTCC 5600
5610 5620 5630 5640
ACTGCTAAATGTTAGTCATTGAGGTGGGGCCAATTTTAAT 5640
CATAAGCCTTAATAAGA+TTTTCTAAGAAATG+GAAATAG 5680
AACAAATTTTCATCTAATTCATTTACTTTTAGATGAATGG 5720
CATTGTGAATGCCATTCTTTTAATGAATTTCAAGAGAATT 5760
CTCTGGTTTTCTGTGTAATTCAGATGAGTCACTGTA ACT 5800

FIGURE 16G

5810 5820 5830 5840
CTAGAAGATTAACCTTCCAGCCAACCTATTTTCCTTTCCC 5840
TTGTCTCTCTCATCCTCTTTTCCTTCCTTCTTTCTTTCT 5880
CTTCTTTTATCTCCAAGGTTAATCAGGAAAAATAGCTTTT 5920
GACAGGGGAAAAAACTCAATAACTAGCTATTTTTGACCTC 5960
CTGATCAGGAACTTTAGTTGAAGCGTAAATCTAAAGAAAC 6000
6010 6020 6030 6040
ATTTTCTCTGAAATATATTATTAAGGGCAATGGAGATAAA 6040
TTAATAGTAGATGTGGTTCCTCAGAAAATATAATCAAAATT 6080
CAAAGATTTTTTTTTGTTTCTGTAACCTGGAATAAATCAA 6120
TGATTACTAGTGTTAATAGTGATAACTTGTTTTTATTGT 6160
TGGTGCATATTAGTATAACTGTGGGGTAGGTCGGGGAGAG 6200
6210 6220 6230 6240
GGTAAGGGAATAGATCACTCAGATGTATTTTAGATAAGCT 6240
ATTTAGCCTTTGATGGAATCATAAATACAGTGAATACAAT 6280
CCTTTGCATTGTTAAGGAGGTTTTTTGTTTTTAAATGGTG 6320
GGTCAAGGAGCTAGTTTACAGGCTTACTGTGATTTAAGCA 6360
AATGTGAAAAGTGAAACCTTAATTTTATCAAAGAAATTT 6400
6410 6420 6430 6440
CTGTAAATGGTATGTCTCCTTAGAATACCCAAATCATAAT 6440
TTTATTTGTACACACTGTTAGGGGCTCATCTCATGTAGGC 6480
AGAGTATAAAGTATTACCTTTTGGAATTAAAAGCCACTGA 6520
CTGTTATAAAGTATAACAACACACATCAGGTTTTAAAAAG 6560
CCTTGAATGGCCCTTGTCTTAAAAAGAAATTAGGAGCCAG 6600
6610 6620 6630 6640
GTGCGGTGGCACGTGCCTGTAGTCCCAGCTCCTTGGGAGG 6640
CTGAGACAGGAGGATTCTTTGAGCCCTGGAGTTTGAGTCC 6680
AGCCTGGGTGACATAGCAAGACCCTGTCTTAAAAGAAAAA 6720
TGGGAAGAAAGACAAGGTAACATGAAGAAAGAAGAGATAC 6760
CTAGTATGATGGAGCTGCAAATTTTCATGGCAGTTCATGCA 6800

FIGURE 16H

6810 6820 6830 6840
GTCGGTCAAGAGGAGGATTTTGTGTTTGTAGTTTGCAGATG 6840
AGCATTTCTAAAGCATTTTCCCTTGCTGTATTTTTTTGTA 6880
TTATAAATTACATTGGACTTCATATATATAA+TTTTTTTTT 6920
ACATTATATGTCTCTTGATGTTTTGAAACTCTTGTATTT 6960
ATGATATAGCTTATATGATTTTTTTGCCTTGGTATACATT 7000
7010 7020 7030 7040
TTAAAATATGAATTTAAAAAATTTTTGTAAAAATAAAATT 7040
CACAAAATTGTTTTGAAAAACAAAAAATAAAAAA 7075

FIGURE 17

MDAKARN C LLQHR EALEKDI KT SYI MDHM I SDG FLTIS EE 40
 E KVRN E PT QQQRAAML IK M ILKKDND SYVSFY NALLHEGY 80
 K DLAALLH D GI PV VSS SSGKDS VSG ITS YVRT VLCEGGVP 120
 QR PVV FVTRKKLVNA IQQK LS KLKGEPGWVTI HGMAGCGK 160
 SVLAAE AVRDSHLL E GCF PGGVHW VS VGKQDKSGLLMKLQ 200
 N LCT RLDQDESFS QRLPLN I E EAKDR LR IL MLRK HP RS LL 240
 I LDDVW DS W VLKAFDS QCQ ILLTTRDKSVTDSVMGPKYVV 280
 P VESS LGKE KGLE ILS LF VN MK KA DLP EQAHS I KECKGS 320
 P LVVS LI G ALLR DFPN RW EYYLKQLQNKQF KRI RKS SSYD 360
 YE ALDEAMS IS VEMLR ED I KDYYT DLS I LQKDVK VPTKVL 400
 C I LW DM ETE EVED I LQEF VNKSLLF CDR NGKSFYRYLHDL 440
 QVD FL TE KN C SQ LQ DLH K K ITQFQRYHQPH TLSPDQEDC 480
 MYWYNFLAYH MASAKMHKELCALMFS LDWI KAKTEL VGPA 520
 H LI HE FVEYR H I LDEKDCAVS E NFQEFLS LN GHLLGR QP F 560
 P NI VQ LG L CE P ET SEVYQQAKLQAKQEVDNGM LYLEWINK 600
 KNI TNLSRLVVRP HTDAVYH AC FS ED GQR IA S C GADKTLQ 640
 VFK AETG EK LLE I K AH EDEVLC CAF STDDRF IATCS VD KK 680
 VKIW NS MT G ELVHTYDEHS EQVNCCH FTNS S HHLLLATGS 720
 SDCFLKLWDL NQ KEC RNTMF GHTN SV NHC RFSPDDKLLAS 760
 CS ADG T LKLWDATS ANE RKS INVK QFF LNLEDP QEDMEVI 800
 VKCCS WSADGAR I MVAANKI FLWNTDS RS KVADCRGHLS 840
 WVHGVMFS PDG SS F LTSSDDQT I RLWET KKVCKNSAVMLK 880
 QEVDVVF QENEVMVL AVDH IRR LQL IN GRTG QI DYLT EAQ 920
 VS CCCLS PH LQYI AFG DEN GAI EI LELVN NR IFQS R FQH K 960
 KTVW H I QFTADE KTLI S SSDDAE IQVWN WQLDKCI FLRGH 1000
 QETVKDF R LLK NS RLLS WS FDGTVKVWN I ITGN KEKDFVC 1040
 H QGTVLS CDIS H DAT KF SS TS ADKTAK IWS FELL LP LH EL 1080
 RG HN GCV RCS AFS VDS TLLATGDDNGE IRI WNVS NG ELLH 1120
 LC AP LS E EGAAT HGGWVTDLCFS PDGKML ISAGGYIKWWN 1160
 VVT GE SS QT F YT NGTN LKKI HVS PD FKTYVTVD NLGI LYI 1200
 LQ TLE 1205

Figure 18

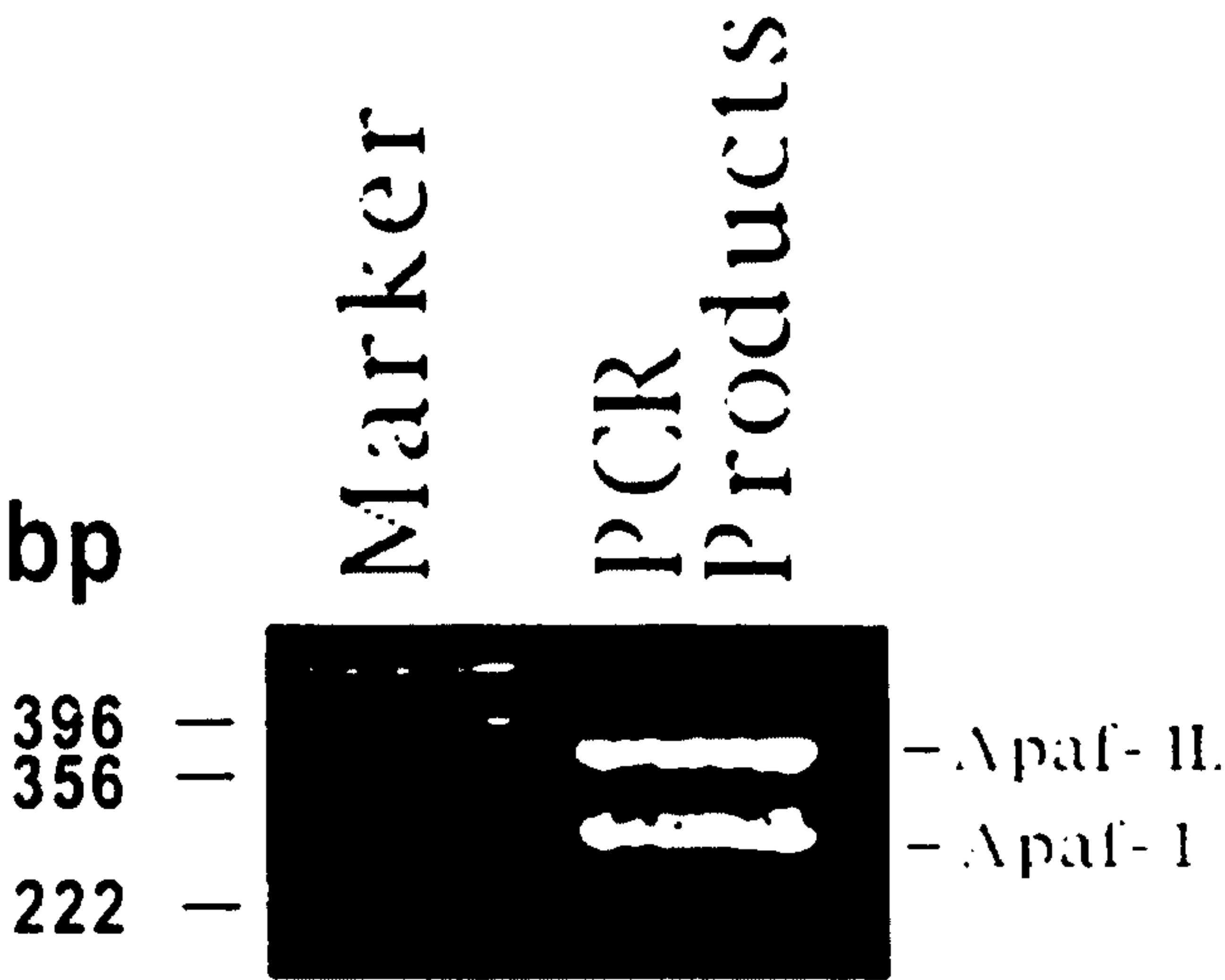
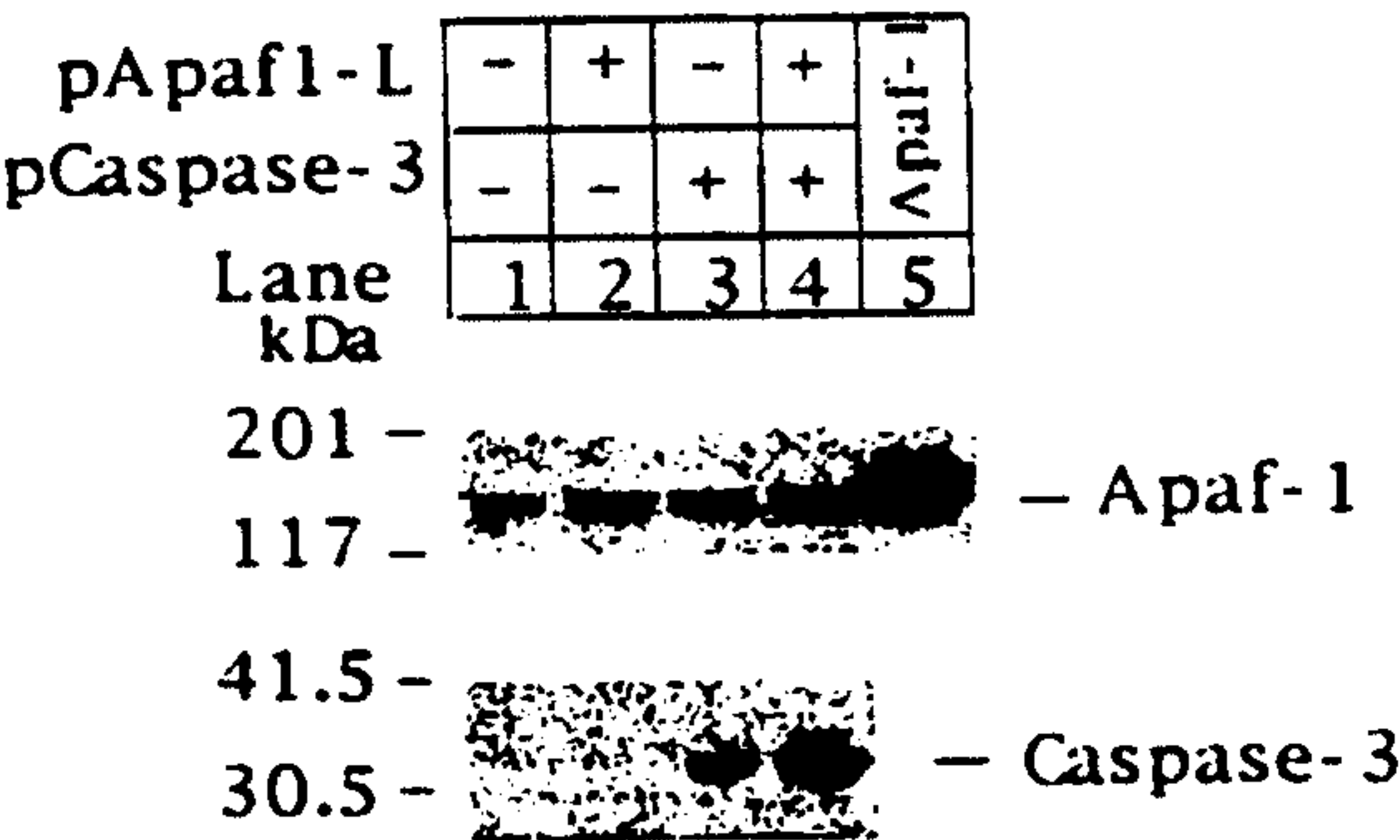


Figure 19

A



B

