(12) STANDARD PATENT (11) Application No. AU 2001239832 B2
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Compositions and methods for the early diagnosis of ovarian cancer

(51)[7] International Patent Classification(s)

A61K 038/17	**G01N 033/53**
C07H 021/04	**C12Q 001/68**
A61K 048/00	**C07K 014/435**

(21) Application No: **2001239832** (22) Date of Filing: **2001.02.20**

(87) WIPO No: **WO01/62271**

(30) Priority Data

(31) Number	(32) Date	(33) Country
09/510,738	**2000.02.22**	**US**

(43) Publication Date: **2001.09.03**
(43) Publication Journal Date: **2001.11.15**
(44) Accepted Journal Date: **2004.09.30**

(71) Applicant(s)
The Board of Trustees of the University of Arkansas

(72) Inventor(s)
O'Brien, Timothy J.

(74) Agent / Attorney
Callinan Lawrie, 711 High Street, Kew, VIC, 3101

(56) Related Art
Tanimoto et al, "Cancer Research", July 1997, vol 57, no 14, p 2884-2887
WO 98/41656

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 August 2001 (30.08.2001)

PCT

(10) International Publication Number
WO 01/62271 A1

(51) International Patent Classification[7]: A61K 38/17, 48/00, C07H 21/04, C07K 14/435, C12Q 1/68, G01N 33/53

(21) International Application Number: PCT/US01/05703

(22) International Filing Date: 20 February 2001 (20.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/510,738 22 February 2000 (22.02.2000) US

(71) Applicant: THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS [US/US]; 2404 North University Avenue, Little Rock, AR 72207-3608 (US).

(72) Inventor: O'BRIEN, Timothy, J.; 2610 North Pierce, Little Rock, AR 72207 (US).

(74) Agent: ADLER, Benjamin, A.; Adler & Associates, 8011 Candle Lane, Houston, TX 77071 (US).

(81) Designated States (national): AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, 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, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/62271 A1

(54) **Title:** COMPOSITIONS AND METHODS FOR THE EARLY DIAGNOSIS OF OVARIAN CANCER

(57) **Abstract:** The disclosed nucleic acid primer sets, used in combination with quantitative amplification (PCR) of tissue cDNA, can indicate the presence of specific proteases in a tissue sample. The detected proteases are themselves specifically overexpressed in certain cancers, and the presence of their genetic precursors may serve for early detection of associated ovarian and other malignancies, and for the design of interactive therapies for cancer treatment.

COMPOSITIONS AND METHODS FOR THE EARLY DIAGNOSIS OF OVARIAN CANCER

BACKGROUND OF THE INVENTION

Field of the Invention

Generally, the present invention relates to the fields of molecular biology and medicine. More specifically, the present invention is in the field of cancer, especially ovarian cancer diagnosis.

Background of the Invention

To date, ovarian cancer remains the number one killer of women with gynecologic malignant hyperplasia. Approximately 75% of women diagnosed with such cancers are already at an advanced stage (III and IV) of the disease at their initial diagnosis. During the past 20 years, neither diagnosis nor five year survival rates have greatly improved for these patients. This is substantially due to the high percentage of high-stage initial detections of the disease. Therefore, the challenge remains to develop new markers that improve early diagnosis and thereby reduce the percentage of high-stage initial diagnoses.

Extracellular proteases have already been implicated in the growth, spread and metastatic progression of many cancers, due to the ability of malignant cells not only to grow *in situ*, but to dissociate from the primary tumor and to invade new surfaces. The ability to disengage from one tissue and re-engage the surface of another tissue is what provides for the morbidity and mortality associated with this disease. Therefore, extracellular proteases may be good candidates for markers of neoplastic development.

In order for malignant cells to grow, spread or metastasize, they must have the capacity to invade local host tissue, dissociate or shed from the primary tumor, and for metastasis to occur, enter and survive in the bloodstream, implant by invasion into the surface of the target organ and establish an environment conducive for new colony growth (including the induction of angiogenic and growth factors). During this progression, natural tissue barriers have to be degraded, including basement membranes and connective tissue. These barriers include collagen, laminin, proteoglycans and extracellular matrix glycoproteins, including fibronectin. Degradation of these natural barriers, both those surrounding the primary tumor and at the sites of metastatic invasion, is believed to be brought about by the action of a matrix of extracellular proteases.

Proteases have been classified into four families: serine proteases, metallo-proteases, aspartic proteases and cysteine proteases. Many proteases have been shown to be involved in the human disease process and these enzymes are targets for the development of inhibitors as new therapeutic agents. Additionally, certain individual proteases have been shown to be induced and overexpressed in a diverse group of cancers, and as such, are potential candidates for markers of early diagnosis and possible therapeutic intervention. A group of examples are shown in Table 1.

TABLE 1

Known proteases expressed in various cancers

	Gastric	Brain	Breast	Ovarian
Serine Proteases:	uPA	uPA	NES-1	NES-1
	PAI-1 PAI-1	uPA	uPA	
		tPA		PAI-2
Cysteine Proteases:	Cathepsin B	Cathepsin L	Cathepsin B	Cathepsin B
	Cathepsin L		Cathepsin L	Cathepsin L
Metallo-proteases:	Matrilysin*	Matrilysin	Stromelysin-3	MMP-2
	Collagenase*	Stromelysin	MMP-8	
	Stromelysin-1*	Gelatinase B	MMP-9	
			Gelatinase A	

uPA, Urokinase-type plasminogen activator; tPA, Tissue-type plasminogen activator; PAI-I, Plasminogen activator 0 inhibitors; PAI-2, Plasminogen activator inhibitors; NES-1, Normal epithelial cell-specific-1; MMP, Matrix P metallo-protease. *Overexpressed in gastrointestinal ulcers.

Significantly, there is a good body of evidence supporting the downregulation or inhibition of individual proteases and the reduction in invasive capacity or malignancy. In work by Clark *et al.*, inhibition of *in vitro* growth of human small cell lung cancer was demonstrated using a general serine protease inhibitor. More recently, Torres-Rosedo *et al.*, [*Proc.Natl.Acad.Sci.USA,* 90, 7181-7185 (1993)] demonstrated an inhibition of hepatoma tumor cell growth using specific antisense inhibitors for the serine protease hepsin gene. Metastatic potential of melanoma cells has also been shown to be reduced in a mouse model using a synthetic inhibitor (batimastat) of metallo-proteases. Powell *et al.* [*Cancer Research*, 53, 417-422 (1993)] presented evidence to confirm that the expression of extracellular proteases in relatively non-invasive tumor cells enhances their malignant progression using a tumorgenic, but non-metastatic, prostate cell line. Specifically, enhanced metastasis was demonstrated after introducing and expressing the PUMP-1 metallo-protease gene. There is also a body of data to support the notion that expression of cell surface proteases on relatively non-metastatic cell types increases the invasive potential of such cells.

Thus, the prior art is deficient in a tumor marker useful as an indicator of early disease, particularly for ovarian cancers. The present invention fulfills this long-standing need and desire in the art.

SUMMARY OF THE INVENTION

This invention allows for the detection of cancer, especially ovarian cancer, by screening for hepsin mRNA in tissue, which is indicative of the hepsin protease, which is shown herein to be specifically associated with the surface of 80 percent of ovarian and other tumors. Proteases are considered to be an integral part of tumor growth and metastasis, and therefore, markers indicative of their presence or absence are useful for the diagnosis of cancer. Furthermore, the present invention is useful for treatment (*i.e.*, by inhibiting hepsin or expression of hepsin), for targeted therapy, for vaccination, etc.

In one embodiment of the present invention, there is provided a method of diagnosing cancer in an individual, comprising the steps of obtaining a biological sample from an individual and detecting hepsin in the sample. The presence of hepsin in the sample is indicative of the presence of carcinoma in the individual, wherein the absence of hepsin in the sample is indicative of the absence of carcinoma in the individual.

In another embodiment of the present invention, there is provided a method for detecting malignant hyperplasia in a biological sample, comprising the steps of isolating mRNA from the sample; and detecting hepsin mRNA in the sample. The presence of the hepsin mRNA in the sample is indicative of the presence of malignant hyperplasia, and the absense of the hepsin mRNA in the sample is indicative of the absence of malignant hyperplasia.

In yet another embodiment of the present invention, there is provided a method for detecting malignant hyperplasia in a biological sample, comprising the steps of isolating protein from the sample; and detecting hepsin protein in the sample. The presence of the hepsin protein in the sample is indicative of the presence of malignant hyperplasia, wherein the absense of the hepsin protein in the sample is indicative of the absence of malignant hyperplasia. This method may further comprise the step of comparing the hepsin protein to reference information, wherein the comparison provides a diagnosis of the malignant hyperplasia, or alternatively, determines a treatment of the malignant hyperplasia.

In still yet another embodiment of the present invention, there is provided a method of inhibiting expression of hepsin in a cell, comprising the step of introducing a vector into a cell, wherein the vector comprises a hepsin gene in opposite orientation operably linked to elements necessary for expression. Expression of the vector produces hepsin antisense mRNA in the cell, which hybridizes to endogenous hepsin mRNA and thereby inhibits expression of hepsin in the cell.

In yet another embodiment of the present invention, there is provided a method of inhibiting a hepsin protein in a cell, comprising the step of introducing an antibody specific for a hepsin protein or a fragment thereof into a cell. Binding of the antibody inhibits the hepsin protein.

In another embodiment of the present invention, there is provided a method of targeted therapy to an individual, comprising the step of administering a compound to an individual, wherein the compound has a targeting moiety and a therapeutic moiety, wherein the targeting moiety is specific for hepsin.

In yet another embodiment of the present invention, there is provided a method of vaccinating an individual against hepsin, comprising the steps of inoculating an individual with a hepsin protein or fragment thereof, wherein the hepsin protein or fragment thereof lack hepsin protease activity. Inoculation with the hepsin protein or fragment thereof elicits an immune response in the individual, thereby vaccinating the individual against hepsin.

In still another embodiment of the present invention, there is provided an oligonucleotide having a sequence complementary to SEQ ID No.188. Also embodied is a composition comprising the above-described oligonucleotide and a physiologically acceptable carrier therefore. Additionally embodied is a method of treating a neoplastic state in an individual in need of such treatment, comprising the step of administering to the individual an effective dose of the above-described oligonucleotide.

In another embodiment of the present invention, there is provided a method of screening for compounds that inhibit hepsin activity, comprising the steps of contacting a sample with a compound, wherein the sample comprises hepsin protein; and assaying for hepsin protease activity. A decrease in the hepsin protease activity in the presence of the compound relative to hepsin protease activity in the absence of the compound is indicative of a compound that inhibits hepsin activity.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention. These embodiments are given for the purpose of disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The appended drawings have been included herein so that the above-recited features, advantages and objects of the invention will become clear and can be understood in detail. These drawings form a part of the specification. It is to be noted, however, that the appended drawings illustrate preferred embodiments of the invention and should not be considered to limit the scope of the invention.

Figure 1 shows agarose gel comparison of PCR products derived from normal and carcinoma cDNA.

Figure 2 shows Northern blot analysis of ovarian tumors using hepsin, SCCE, PUMP-1, TADG-14 and β-tubulin probes.

Figure 3 shows amplification with serine protease redundant primers: histidine sense (S1) with aspartic acid antisense (AS1), using normal cDNA (Lane 1) and tumor cDNA (Lane 2); and histidine sense (S1) with serine antisense (AS2), using normal cDNA (Lane 3) and tumor cDNA (Lane 4).

Figure 4 shows amplification with cysteine protease redundant primers. Normal (Lane 1), low malignant potential (Lane 2), serious carcinoma (Lane 3), mucinous carcinoma (Lane 4), and clear cell carcinoma (Lane 5).

Figure 5 shows amplification with metallo-protease redundant primers. Normal (Lane 1), low malignant potential (Lane 2), serious carcinoma (Lane 3), mucinous carcinoma (Lane 4), and clear cell carcinoma (Lane 5).

Figure 6 shows amplification with specific primers directed towards the serine protease, hepsin. Expression in normal (Lanes 1-3), low malignant potential tumors (Lanes 4-8), and ovarian carcinomas (Lanes 9-12).

Figure 7 shows hepsin expression levels in normal, low malignant potential tumors, and ovarian carcinomas. S=serious, M=mucinous, LMP=low malignant potential.

Figure 8 shows serine protease stratum corneum chymotrypsin enzyme (SCCE) expression in normal, low malignant potential tumors, and ovarian carcinomas.

Figure 9 shows metallo-protease PUMP-1 (MMP-7) gene expression in normal (lanes 1-2) and ovarian carcinomas tissue (Lanes 3-10).

Figure 10A shows Northern blot analysis of hepsin expression in normal ovary and ovarian carcinomas. *Lane 1*, normal ovary (case 10); *lane 2*, serous carcinoma (case 35); *lane 3*, mucinous carcinoma (case 48); *lane 4*, endometrioid carcinoma (case 51); and *lane 5*, clear cell carcinoma (case 54). In cases 35, 51 and 54, more than a 10-fold increase in the hepsin 1.8 kb transcript abundance was observed. **Figure 10B** shows Northern blot analysis of hepsin in normal human fetal. **Figure 10C** shows Northern blot analysis of hepsin in adult tissues. Significant overexpression of the hepsin transcript is noted in both fetal liver and fetal kidney. Notably, hepsin overexpression is not observed in normal adult tissue. Slight expression above the background level is observed in the adult prostate.

Figure 11A shows hepsin expression in normal (N), mucinous (M) and serous (S) low malignant potential (LMP) tumors and carcinomas (CA). β-tubulin was used as an internal control. **Figure 11B** shows the ratio of hepsin:β-tubulin expression in normal ovary, LMP tumor, and ovarian carcinoma. Hepsin mRNA expression levels were significantly elevated in LMP tumors, ($p < 0.005$) and carcinomas ($p < 0.0001$) compared to levels in normal ovary. All 10 cases of normal ovaries showed a relatively low level of hepsin mRNA expression.

Figure 12A shows northern blot analysis of mRNA expression of the SCCE gene in fetal tissue. **Figure 12B** shows northern blot analysis of mRNA expression of the SCCE gene in ovarian tissue.

Figure 13A shows a comparison of quantitative PCR of SCCE cDNA from normal ovary and ovarian carcinomas. **Figure 13B** shows a bar graph comparing the ratio of SCCE to β-tubulin in 10 normal and 44 ovarian carcinoma tissues.

Figure 14 shows a comparison by quantitative PCR of normal and ovarian carcinoma expression of mRNA for protease M.

Figure 15 shows the TADG-12 catalytic domain including an insert near the His 5'-end.

Figure 16A shows northern blot analysis comparing TADG-14 expression in normal and ovarian carcinoma tissues. **Figure 16B** shows preliminary quantitative PCR amplification of normal and carcinoma cDNAs using specific primers for TADG-14.

Figure 17A shows northern blot analysis of the PUMP-1 gene in human fetal tissue. **Figure 17B** shows northern blot analysis of the PUMP-1 gene in normal ovary and ovarian carcinomas.

Figure 18A shows a comparison of PUMP-1 expression in normal and carcinoma tissues using quantitative PCR with an internal β-tubulin control. **Figure 18B** shows the ratio of mRNA expression of PUMP-1 compared to the internal control β-tubulin in 10 normal and 44 ovarian carcinomas.

Figure 19 shows a comparison of PCR amplified products for the hepsin, SCCE, protease M, PUMP-1 and Cathepsin L genes.

DETAILED DESCRIPTION OF THE INVENTION

This invention identifies a hepsin protease on ovarian and other tumor cells which is characteristic of this type of cancer, and in various combinations with other proteases, is characteristic of individual tumor types. Such information can provide the basis for diagnostic tests (assays or immunohistochemistry), prognostic evaluation (depending on the display pattern) and therapeutic intervention utilizing either antibodies directed at the protease, antisense vehicles for downregulation or protease inhibitors both from established inhibition data and/or for the design of new drugs. Long-term treatment of tumor growth, invasion and metastasis has not succeeded with existing chemotherapeutic agents - most tumors become resistant to drugs after multiple cycles of chemotherapy.

A primary object of the present invention is a method for detecting the presence of malignant hyperplasia in a tissue sample. It is an advantage of the present invention that it has as a particular object the detection of cancer in ovarian tissue. The cancer is detected by analyzing a biological sample for the presence of markers to proteases that are specific indicators of certain types of cancer cells. This object may be accomplished by isolating mRNA from a sample or by detection of proteins by polyclonal or preferably monoclonal antibodies. When using mRNA detection, the method may be carried out by combining the isolated mRNA with reagents to convert to cDNA according to standard methods; treating the converted cDNA with amplification reaction reagents (such as cDNA PCR reaction reagents) in a container along with an appropriate mixture of nucleic acid primers selected from the list in

Table 2 or as detailed above; reacting the contents of the container to produce amplification products; and analyzing the amplification products to detect the presence of malignant hyperplasia markers in the sample. For mRNA, the analyzing step may be accomplished using Northern Blot analysis to detect the presence of malignant hyperplasia markers in the amplification product. Northern Blot analysis is known in the art. The analysis step may be further accomplished by quantitatively detecting the presence of malignant hyperplasia marker in the amplification produce, and comparing the quantity of marker detected against a panel of expected values for known presence or absence in normal and malignant tissue derived using similar primers.

Another embodiment of the present invention are various nucleic acid sequences that are useful in the methods disclosed herein. These nucleic acid sequences are listed in Table 2. It is anticipated that these nucleic acid sequences be used in mixtures to accomplish the utility of this invention. Features of such mixtures include: SEQ ID No. 1 with SEQ ID No. 2; SEQ ID No. 1 with SEQ ID No. 3; SEQ ID No. 4 with SEQ ID No. 5; SEQ ID No. 6 with SEQ ID No. 7; SEQ ID No. 8 with SEQ ID No. 9; and SEQ ID No. 10 with SEQ ID No. 11. The skilled artisan may be able to develop other nucleic acid sequences and mixtures thereof to accomplish the benefit of this invention, but it is advantageous to have the sequences listed in Table 2 available without undue experimentation.

The present invention is directed toward a method of diagnosing cancer in an individual, comprising the steps of obtaining a biological sample from an individual; and detecting hepsin in the sample. The presence of hepsin in the sample is indicative of the presence of cancer in the individual, wherein the absence of hepsin in the sample is indicative of the absence of cancer in the individual. Generally, detection of the hepsin is by means such as Northern blot, Western blot, PCR, dot blot, ELISA sandwich assay, radioimmunoassay, DNA array chips and flow cytometry. An example of a typical cancer diagnosed by this method is ovarian cancer.

The present invention is also directed toward a method for detecting malignant hyperplasia in a biological sample, comprising the steps of isolating mRNA from the sample; and detecting hepsin mRNA in the sample. The presence of the hepsin mRNA in the sample is indicative of the presence of malignant hyperplasia, wherein the absense of the hepsin mRNA in the sample is indicative of the absence of malignant hyperplasia. This method may further comprise the step of comparing the hepsin mRNA to reference information, wherein the comparison provides a diagnosis and/or determines a treatment of the malignant hyperplasia. A typical means of detection of hepsin mRNA is by PCR amplification, which, preferably, uses primers shown in SEQ ID No. 8 and SEQ ID No. 9. Representative biological samples include a tissue and a bodily fluid, wherein the bodily fluid is preferably blood.

The present invention is additionally directed toward a method for detecting malignant hyperplasia in a biological sample, comprising the steps of isolating protein from the sample; and detecting hepsin protein in the sample. The presence of the hepsin protein in the sample is indicative of the presence of malignant hyperplasia, wherein the absense of the hepsin protein in the sample is indicative of the absence of malignant hyperplasia. This method also may comprise the step of comparing the hepsin protein to reference information, wherein the comparison provides a diagnosis or determines a treatment of the malignant hyperplasia. Preferably, the detection of the hepsin protein is by immunoaffinity to an antibody which is specific for hepsin. Representative biological samples are a tissue and a bodily fluid, and it is preferable that the bodily fluid is blood.

The present invention is further directed toward a method of inhibiting expression of hepsin in a cell, comprising the step of introducing a vector into a cell, wherein the vector comprises a hepsin gene in opposite orientation operably linked to elements necessary for expression, wherein expression of the vector produces hepsin antisense mRNA in the cell. The hepsin antisense mRNA hybridizes to endogenous hepsin mRNA, thereby inhibiting expression of hepsin in the cell.

The present invention is still further directed toward a method of inhibiting a hepsin protein in a cell, comprising the step of introducing an antibody into a cell, wherein the antibody is specific for a hepsin protein or a fragment thereof. Binding of the antibody to hepsin inhibits the hepsin protein. Preferably, the hepsin fragment is a 9-residue fragment up to a 20-residue fragment, and more preferably, the 9-residue fragment is SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 and 154.

The present invention is also directed toward a method of targeted therapy to an individual, comprising the step of administering a compound to an individual, wherein the compound has a targeting moiety and a therapeutic moiety, and wherein the targeting moiety is specific for hepsin. Preferably, the targeting moiety is an antibody specific for hepsin or a ligand or ligand binding domain that binds hepsin. Likewise, the therapeutic moiety is preferably a radioisotope, a toxin, a chemotherapeutic agent, an immune stimulant or cytotoxic agent. Generally, the individual suffers from a disease such as ovarian cancer, lung cancer, prostate cancer, colon cancer or another cancer in which hepsin is overexpressed.

The present invention is additionally directed toward a method of vaccinating an individual against hepsin, comprising the steps of inoculating an individual with a hepsin protein or fragment thereof, wherein the hepsin protein or fragment thereof lack hepsin protease activity. Inoculation with the hepsin protein, or fragment thereof, elicits an immune response in the individual, thereby vaccinating the individual against hepsin. Generally, this method is applicable when the individual has cancer, is suspected of having cancer or is at risk of getting cancer. Sequences of preferred hepsin proteins or fragment thereof are shown in SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 and 154.

The present invention is yet directed toward a method of producing immune-activated cells directed toward hepsin, comprising the steps of exposing dendritic cells to hepsin protein or fragment thereof, which lacks hepsin protease activity. Typically, exposure to hepsin protein or fragment thereof activates the dendritic cells, thereby producing immune-activated cells directed toward hepsin. Generally, the immune-activated cells are B-cells, T-cells and/or dendrites. Preferably, the hepsin fragment is a 9-residue fragment up to a 20-residue fragment, and more preferably, the 9-residue fragment is SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 or 154. Oftentimes, the dendritic cells are isolated from an individual prior to exposure and then reintroduced into the individual subsequent to the exposure. Typically, the individual has cancer, is suspected of having cancer or is at risk of getting cancer.

The present invention is further directed toward an immunogenic composition, comprising an immunogenic fragment of hepsin protein and an appropriate adjuvant. Preferably, the fragment is a 9-residue fragment up to a 20-residue fragment, and more preferably, the 9-residue fragment is SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 or 154.

The present invention is further directed toward an oligonucleotide having a sequence complementary to SEQ ID No.188 or a frgament thereof. The present invention further provides a composition comprising the above-described oligonucleotide and a physiologically acceptable carrier therefore, and a method of treating a neoplastic state in an individual in need of such treatment, comprising the step of administering to the individual an effective dose of the above-described oligonucleotide. Typically, the neoplastic state may be ovarian cancer, breast cancer, lung cancer, colon cancer, prostate cancer or another cancer in which hepsin is overexpressed.

The present invention is still further directed toward a method of screening for compounds that inhibit hepsin activity, comprising the steps of contacting a sample with a compound, wherein the sample comprises hepsin protein; and assaying for hepsin protease activity. A decrease in the hepsin protease activity in the presence of the compound relative to hepsin protease activity in the absence of the compound is indicative of a compound that inhibits hepsin activity.

The present invention is yet additionally directed toward a method for detecting ovarian malignant hyperplasia in a biological sample, comprising the steps of isolating the proteases or protease mRNA present in the biological sample; and detecting specific proteases or protease mRNA present in the biological sample. The proteases are selected from the group consisting of hepsin, protease M, complement factor B, SCCE, cathepsin L and PUMP-1. This method may further comprise the step of comparing the specific proteases or protease mRNA detected to reference information, wherein the comparison provides a diagnoses or determines a treatment of the malignant hyperplasia. Typically, the protease mRNA is detected by amplification of total mRNA, and the protease is detected with an antibody. Representative biological samples are blood, urine, saliva, tears, interstitial fluid, ascites fluid, tumor tissue biopsy and circulating tumor cells.

It will be apparent to one skilled in the art that various substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D.N. Glover ed. 1985); "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" (B.D. Hames & S.J. Higgins eds. 1985); "Transcription and Translation" (B.D. Hames & S.J. Higgins eds. 1984); "Animal Cell Culture" (R.I. Freshney, ed. 1986); "Immobilized Cells And Enzymes" (IRL Press, 1986); B. Perbal, "A Practical Guide To Molecular Cloning" (1984). Therefore, if appearing herein, the following terms shall have the definitions set out below.

As used herein, the term "cDNA" shall refer to the DNA copy of the mRNA transcript of a gene.

As used herein, the term "derived amino acid sequence" shall mean the amino acid sequence determined by reading the triplet sequence of nucleotide bases in the cDNA.

As used herein the term "screening a library" shall refer to the process of using a labeled probe to check whether, under the appropriate conditions, there is a sequence complementary to the probe present in a particular DNA library. In addition, "screening a library" could be performed by PCR.

As used herein, the term "PCR" refers to the polymerase chain reaction that is the subject of U.S. Patent Nos. 4,683,195 and 4,683,202 to Mullis, as well as other improvements now known in the art.

The amino acid described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form can be substituted for any L-amino acid residue, as long as the desired functional property of immunoglobulin-binding is retained by the polypeptide. NH_2 refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, *J Biol. Chem.*, 243:3552-59 (1969), abbreviations for amino acid residues may be used.

It should be noted that all amino-acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-terminus to carboxy-terminus. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino-acid residues.

A "replicon" is any genetic element (*e.g.*, plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; *i.e.*, capable of replication under its own control.

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment. A "vector" may further be defined as a replicable nucleic acid construct, *e.g.*, a plasmid or viral nucleic acid.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in its either single-stranded form or as a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (*e.g.*, restriction fragments), viruses, plasmids, and chromosomes. The structure is discussed herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA).

An expression vector is a replicable construct in which a nucleic acid sequence encoding a polypeptide is operably linked to suitable control sequences capable of effecting expression of the polypeptide in a cell. The need for such control sequences will vary depending upon the cell selected and the transformation method chosen. Generally, control sequences include a transcriptional promoter and/or enhancer, suitable mRNA ribosomal binding sites and sequences which control the termination of transcription and translation. Methods which are well known to those skilled in the art can be used to construct expression vectors containing appropriate transcriptional and translational control signals. See, for example, techniques described in Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual* (2nd Ed.), Cold Spring Harbor Press, N.Y. A gene and its transcription control sequences are defined as being "operably linked" if the transcription control sequences effectively control transcription of the gene. Vectors of the invention include, but are not limited to, plasmid vectors and viral vectors. Preferred viral vectors of the invention are those derived from retroviruses, adenovirus, adeno-associated virus, SV40 virus, or herpes viruses. In general, expression vectors contain promoter sequences which facilitate the efficient transcription of the inserted DNA fragment and are used in connection with a specific host. The expression vector typically contains an origin of replication, promoter(s), terminator(s), as well as specific genes which are capable of providing phenotypic selection in transformed cells. The transformed hosts can be fermented and cultured according to means known in the art to achieve optimal cell growth.

An "origin of replication" refers to those DNA sequences that participate in DNA synthesis.

A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are typically determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (*e.g.*, mammalian) DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters typically contain Shine-Dalgarno ribosome-binding sequences in addition to the -10 and -35 consensus sequences.

An "expression control sequence" is a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then translated into the protein encoded by the coding sequence.

A "signal sequence" can be included near the coding sequence. This sequence encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell to direct the polypeptide to the cell surface or secrete the polypeptide into the media, and this signal peptide is clipped off by the host cell before the protein leaves the cell. Signal sequences can be found associated with a variety of proteins native to prokaryotes and eukaryotes.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

A cell has been "transformed" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

Two DNA sequences are "substantially homologous" when at least about 75% (preferably at least about 80%, and most preferably at least about 90% or 95%) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, *e.g.*, Maniatis et al., *supra*; DNA Cloning, Vols. I & II, *supra*; Nucleic Acid Hybridization, *supra*.

A "heterologous" region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. Another example is a construct where the coding sequence itself is not found in nature (*e.g.*, a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals which fluoresce when exposed to ultraviolet light, and others. A number of fluorescent materials are known and can be utilized as labels. These include, for example, fluorescein, rhodamine, auramine, Texas Red, AMCA blue and Lucifer Yellow. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate. Proteins can also be labeled with a radioactive element or with an enzyme. The radioactive label can be detected by any of the currently available counting procedures. The preferred isotope may be selected from ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re. Enzyme labels are likewise useful, and can be detected by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Many enzymes which can be used in these procedures are known and can be utilized. The preferred are peroxidase, β-glucuronidase, β-D-glucosidase, β-D-galactosidase, urease, glucose oxidase plus peroxidase and alkaline phosphatase. U.S. Patent Nos. 3,654,090, 3,850,752, and 4,016,043 are referred to by way of example for their disclosure of alternate labeling material and methods.

A particular assay system developed and utilized in the art is known as a receptor assay. In a receptor assay, the material to be assayed is appropriately labeled and then certain cellular test colonies are inoculated with a quantitiy of both the label after which binding studies are conducted to determine the extent to which the labeled material binds to the cell receptors. In this way, differences in affinity between materials can be ascertained.

An assay useful in the art is known as a "cis/trans" assay. Briefly, this assay employs two genetic constructs, one of which is typically a plasmid that continually expresses a particular receptor of interest when transfected into an appropriate cell line, and the second of which is a plasmid that expresses a reporter such as luciferase, under the control of a receptor/ligand complex. Thus, for example, if it is desired to evaluate a compound as a ligand for a particular receptor, one of the plasmids would be a construct that results in expression of the receptor in the chosen cell line, while the second plasmid would possess a promoter linked to the luciferase gene in which the response element to the particular receptor is inserted. If the compound under test is an agonist for the receptor, the ligand will complex with the receptor, and the resulting complex will bind the response element and initiate transcription of the luciferase gene. The resulting chemiluminescence is then measured photometrically, and dose response curves are obtained and compared to those of known ligands. The foregoing protocol is described in detail in U.S. Patent No. 4,981,784.

As used herein, the term "host" is meant to include not only prokaryotes but also eukaryotes such as yeast, plant and animal cells. A recombinant DNA molecule or gene which encodes a human hepsin protein of the present invention can be used to transform a host using any of the techniques commonly known to those of ordinary skill in the art. Especially preferred is the use of a vector containing coding sequences for the gene which encodes a human hepsin protein of the present invention for purposes of prokaryote transformation. Prokaryotic hosts may include *E. coli*, *S. tymphimurium*, *Serratia marcescens* and *Bacillus subtilis*. Eukaryotic hosts include yeasts such as *Pichia pastoris*, mammalian cells and insect cells.

As used herein, "substantially pure DNA" means DNA that is not part of a milieu in which the DNA naturally occurs, by virtue of separation (partial or total purification) of some or all of the molecules of that milieu, or by virtue of alteration of sequences that flank the claimed DNA. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote; or which exists as a separate molecule (*e.g.*, a cDNA or a genomic or cDNA fragment produced by polymerase chain reaction (PCR) or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence, *e.g.*, a fusion protein. Also included is a recombinant DNA which includes a portion of the nucleotides listed in SEQ ID No. 188 and which encodes an alternative splice variant of hepsin.

By a "substantially pure protein" is meant a protein which has been separated from at least some of those components which naturally accompany it. Typically, the protein is substantially pure when it is at least 60% (by weight) free from the proteins and other naturally-occurring organic molecules with which it is naturally associated in *vivo*. Preferably, the purity of the preparation (by weight) is at least 75%, more preferably at least 90%, and most preferably at least 99%. A substantially pure hepsin protein may be obtained, for example, by extraction from a natural source; by expression of a recombinant nucleic acid encoding a hepsin polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, *e.g.*, column chromatography, such as immunoaffinity chromatography using an antibody specific for hepsin, polyacrylamide gel electrophoresis, or HPLC analysis. A protein is substantially free of naturally associated components when it is separated from at least some of those contaminants which accompany it in its natural state. Thus, a protein which is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be, by definition, substantially free from its naturally associated components. Accordingly, substantially pure proteins include eukaryotic proteins synthesized in *E. coli*, other prokaryotes, or any other organism in which they do not naturally occur.

The term "oligonucleotide", as used herein, is defined as a molecule comprised of two or more ribonucleotides, preferably more than three. Its exact size will depend upon many factors, which, in turn, depend upon the ultimate function and use of the oligonucleotide. The term "primer", as used herein, refers to an oligonucleotide, whether occurring naturally (as in a purified restriction digest) or produced synthetically, and which is capable of initiating synthesis of a strand complementary to a nucleic acid when placed under appropriate conditions, *i.e.*, in the presence of nucleotides and an inducing agent, such as a DNA polymerase, and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, sequence and/or homology of primer and the method used. For example, in diagnostic applications, the oligonucleotide primer typically contains 15-25 or more nucleotides, depending upon the complexity of the target sequence, although it may contain fewer nucleotides.

The primers herein are selected to be "substantially" complementary to particular target DNA sequences. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment (*i.e.*, containing a restriction site) may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementary with the sequence to hybridize therewith and form the template for synthesis of the extension product.

The probe to which the DNA of the invention hybridizes preferably consists of a sequence of at least 20 consecutive nucleotides, more preferably 40 nucleotides, even more preferably

50 nucleotides, and most preferably 100 nucleotides or more (up to 100%) of the coding sequence of the nucleotides listed in SEQ ID No. 188 or the complement thereof. Such a probe is useful for detecting expression of hepsin in a cell by a method including the steps of (a) contacting mRNA obtained from the cell with a labeled hepsin hybridization probe; and (b) detecting hybridization of the probe with the mRNA.

By "high stringency" is meant DNA hybridization and wash conditions characterized by high temperature and low salt concentration, *e.g.*, wash conditions of 65°C at a salt concentration of approximately 0.1X SSC, or the functional equivalent thereof. For example, high stringency conditions may include hybridization at about 42°C in the presence of about 50% formamide; a first wash at about 65°C with about 2X SSC containing 1% SDS; followed by a second wash at about 65°C with about 0.1X SSC.

The DNA may have at least about 70% sequence identity to the coding sequence of the nucleotides listed in SEQ ID No. 188, preferably at least 75% (*e.g.*, at least 80%); and most preferably at least 90%. The identity between two sequences is a direct function of the number of matching or identical positions. When a position in both of the two sequences is occupied by the same monomeric subunit, *e.g.*, if a given position is occupied by an adenine in each of two DNA molecules, then they are identical at that position. For example, if 7 positions in a sequence 10 nucleotides in length are identical to the corresponding positions in a second 10-nucleotide sequence, then the two sequences have 70% sequence identity. The length of comparison sequences will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 100 nucleotides. Sequence identity is typically measured using sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705).

The present invention comprises a vector comprising a DNA sequence which encodes a hepsin protein, wherein said vector is capable of replication in a host, and comprises, in operable linkage: a) an origin of replication; b) a promoter; and c) a DNA sequence coding for said hepsin protein. Preferably, the vector of the present invention contains a portion of the DNA sequence shown in SEQ ID No. 188. Vectors may be used to amplify and/or express nucleic acid encoding a hepsin protein or fragment thereof.

In addition to substantially full-length proteins, the invention also includes fragments (*e.g.*, antigenic fragments) of the hepsin protein. As used herein, "fragment," as applied to a polypeptide, will ordinarily be at least 10 residues, more typically at least 20 residues, and preferably at least 30 (*e.g.*, 50) residues in length, but less than the entire, intact sequence. Fragments of the hepsin protein can be generated by methods known to those skilled in the art, *e.g.*, by enzymatic digestion of naturally occurring or recombinant hepsin protein, by recombinant DNA techniques using an expression vector that encodes a defined fragment of hepsin, or by chemical synthesis. The ability of a candidate fragment to exhibit a characteristic of hepsin (*e.g.*, binding to an antibody specific for hepsin) can be assessed by methods described herein. Purified hepsin or antigenic fragments of hepsin can be used to generate new antibodies or to test existing antibodies (*e.g.*, as positive controls in a diagnostic assay) by employing standard protocols known to those skilled in the art. Included in this invention is polyclonal antisera generated by using hepsin or a fragment of hepsin as the immunogen in, *e.g.*, rabbits. Standard protocols for monoclonal and polyclonal antibody production known to those skilled in this art are employed. The monoclonal antibodies generated by this procedure can be screened for the ability to identify recombinant hepsin cDNA clones, and to distinguish them from other cDNA clones.

Further included in this invention are hepsin proteins which are encoded, at least in part, by portions of SEQ ID No. 188, *e.g.*, products of alternative mRNA splicing or alternative protein processing events, or in which a section of hepsin sequence has been deleted. The fragment, or the intact hepsin polypeptide, may be covalently linked to another polypeptide, *e.g.*, one which acts as a label, a ligand or a means to increase antigenicity.

The invention also includes a polyclonal or monoclonal antibody which specifically binds to hepsin. The invention encompasses not only an intact monoclonal antibody, but also an immunologically-active antibody fragment, *e.g.*, a Fab or $(Fab)_2$ fragment; an engineered single chain Fv molecule; or a chimeric molecule, *e.g.*, an antibody which contains the binding specificity of one antibody, *e.g.*, of murine origin, and the remaining portions of another antibody, *e.g.*, of human origin.

In one embodiment, the antibody, or a fragment thereof, may be linked to a toxin or to a detectable label, *e.g.*, a radioactive label, non-radioactive isotopic label, fluorescent label, chemiluminescent label, paramagnetic label, enzyme label, or colorimetric label. Examples of suitable toxins include diphtheria toxin, *Pseudomonas* exotoxin A, ricin, and cholera toxin. Examples of suitable enzyme labels include malate hydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase, acetylcholinesterase, etc. Examples of suitable radioisotopic labels include ^{3}H, ^{125}I, ^{131}I, ^{32}P, ^{35}S, ^{14}C, etc.

Paramagnetic isotopes for purposes of *in vivo* diagnosis can also be used according to the methods of this invention. There are numerous examples of elements that are useful in magnetic resonance imaging. For discussions on *in vivo* nuclear magnetic resonance imaging, see, for example, Schaefer et al., (1989) *JACC* 14, 472-480; Shreve et al., (1986) *Magn. Reson. Med.* 3, 336-340; Wolf, G. L., (1984) *Physiol. Chem. Phys. Med. NMR* 16, 93-95; Wesbey et al., (1984) *Physiol. Chem. Phys. Med. NMR* 16, 145-155; Runge et al., (1984) *Invest. Radiol.* 19, 408-415. Examples of suitable fluorescent labels include a fluorescein label, an isothiocyalate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an ophthaldehyde label, a fluorescamine label, etc. Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, an aequorin label, etc.

Those of ordinary skill in the art will know of other suitable labels which may be employed in accordance with the present invention. The binding of these labels to antibodies or fragments thereof can be accomplished using standard techniques commonly known and used by those of ordinary skill in the art. Typical techniques are described by Kennedy et al., (1976) *Clin. Chim. Acta* 70, 1-31; and Schurs et al., (1977) *Clin. Chim. Acta* 81, 1-40. Coupling techniques mentioned in the latter are the glutaraldehyde method, the periodate method, the dimaleimide method, the m-maleimidobenzyl-N-hydroxy-succinimide ester method. All of these methods are incorporated by reference herein.

Also within the invention is a method of detecting hepsin protein in a biological sample, which includes the steps of contacting the sample with the labeled antibody, *e.g.*, radioactively tagged antibody specific for hepsin, and determining whether the antibody binds to a component of the sample. Antibodies to the hepsin protein can be used in an immunoassay to detect increased levels of hepsin protein expression in tissues suspected of neoplastic transformation. These same uses can be achieved with Northern blot assays and analyses.

As described herein, the invention provides a number of diagnostic advantages and uses. For example, the hepsin protein is useful in diagnosing cancer in different tissues since this protein is highly overexpressed in tumor cells. Antibodies (or antigen-binding fragments thereof) which bind to an epitope specific for hepsin are useful in a method of detecting hepsin protein in a biological sample for diagnosis of cancerous or neoplastic transformation. This method includes the steps of obtaining a biological sample (*e.g.*, cells, blood, plasma, tissue, etc.) from a patient suspected of having cancer, contacting the sample with a labeled antibody (*e.g.*, radioactively tagged antibody) specific for hepsin, and detecting the hepsin protein using standard immunoassay techniques such as an ELISA. Antibody binding to the biological sample indicates that the sample contains a component which specifically binds to an epitope within hepsin.

Likewise, a standard Northern blot assay can be used to ascertain the relative amounts of hepsin mRNA in a cell or tissue obtained from a patient suspected of having cancer, in accordance with conventional Northern hybridization techniques known to those of ordinary skill in the art. This Northern assay uses a hybridization probe, *e.g.*, radiolabelled hepsin cDNA, either containing the full-length, single stranded DNA having a sequence complementary to SEQ ID No. 188, or a fragment of that DNA sequence at least 20 (preferably at least 30, more preferably at least 50, and most preferably at least 100 consecutive nucleotides in length). The DNA hybridization probe can be labeled by any of the many different methods known to those skilled in this art.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion:

EXAMPLE 1

<u>Amplification of serine proteases using redundant and specific primers</u>

Only cDNA preparations deemed free of genomic DNA were used for gene expression analysis. Redundant primers were prepared for serine proteases, metallo-proteases and cysteine protease. The primers were synthesized to consensus sequences of amino acid surrounding the catalytic triad for serine proteases, *viz.* histidine ... aspartate ... and serine. The sequences of both sense (histidine & aspartate) and antisense (aspartate and serine) redundant primers are shown in Table 2.

TABLE 2

PCR Primers	5'→3'	SEQ ID No.
Redundant Primers:		
Serine Protease (histidine) = S1	tgggtigtiacigcigcica(ct)tg	1
Serine Protease (aspartic acid) = AS1	a(ag)ia(ag)igciatitcitticc	2
Serine Protease (serine) = AS11	a(ag)iggiccicci(cg)(ta)(ag)tcicc	3
Cysteine Protease – sense	ca(ag)ggica(ag)tg(ct)ggi(ta)(cg)itg(ct)tgg	4
Cysteine Protease - antisense	taiccicc(ag)tt(ag)caicc(ct)tc	5
Metallo Protease - sense	cci(ac)gitg(tc)ggi(ga)(ta)icciga	6
Metallo Protease - antisense	tt(ag)tgicciai(ct)tc(ag)tg	7
Specific Primers:		
Serine Protease (hepsin) = sense	tgtcccgatggcgagtgttt	8
Serine Protease (hepsin) = antisense	cctgttggccatagtactgc	9
Serine Protease (SCCE) = sense	agatgaatgagtacaccgtg	10
Serine Protease (SCCE) = antisense	ccagtaagtccttgtaaacc	11
Serine Protease (Comp B) = sense	aagggacacgagagctgtat	12
Serine Protease (Comp B) = antisense	aagtggtagttggaggaagc	13
Serine Protease (Protease M)= sense	ctgtgatccaccctgactat	20
Serine Protease (Protease M) = antisense	caggtggatgtatgcacact	21
Serine Protease (TADG12) = sense (Ser10-s)	gcgcactgtgtttatgagat	22
Serine Protease (TADG12) = antisense (Ser10-as)	ctctttggcttgtacttgct	23
Serine Protease (TADG13) = sense	tgagggacatcattatgcac	24
Serine Protease (TADG13) = antisense	caagttttccccataattgg	25
Serine Protease (TADG14) = sense	acagtacgcctgggagacca	26
Serine Protease (TADG14) = antisense	ctgagacggtgcaattctgg	27
Cysteine Protease (Cath-L) = sense	attggagagagaaaggctac	14
Cysteine Protease (Cath-L) = antisense	cttgggattgtacttacagg	15
Metallo Protease (PUMP1) = sense	cttccaaagtggtcacctac	16
Metallo Protease (PUMP1) = antisense	ctagactgctaccatccgtc	17

EXAMPLE 2

Carcinoma tissue

Several protease entities were identified and subcloned from PCR amplification of cDNA derived from serous cystadenocarcinomas. Therefore, the proteases described herein are reflective of surface activities for this type of carcinoma, the most common form of ovarian cancer. Applicant has also shown PCR amplification bands unique to the mucinous tumor type and the clear cell type of similar base pair size. About 20-25% of ovarian cancers are classified as either mucinous, clear cell, or endometrioid.

EXAMPLE 3

Ligation, transformation and sequencing

To determine the identity of the PCR products, all the appropriate bands were ligated into Promega T-vector plasmid and the ligation product was used to transform JM109 cells (Promega) grown on selective media. After selection and culturing of individual colonies, plasmid DNA was isolated by means of the WIZARD MINIPREP™ DNA purification system (Promega). Inserts were sequenced using a Prism Ready Reaction Dydeoxy Terminators cycle sequencing kit (Applied Biosystems). Residual dye terminators were removed from the completed sequencing reaction using a CENTRISEP SPIN™ column (Princeton Separation), and samples were loaded into an Applied Biosystems Model 373A DNA sequencing system. The results of subcloning and sequencing for the serine protease primers are summarized in Table 3.

TABLE 3

Serine protease candidates

Subclone	Primer Set	Gene Candidate
1	His-Ser	Hepsin
2	His-Ser	SCCE
3	His-Ser	Compliment B
4	His-Asp	Cofactor 1
5	His-Asp	TADG-12*
6	His-Ser	TADG-13*
7	His-Ser	TADG-14*
8	His-Ser	Protease M
9	His-Ser	TADG-15*

*indicates novel proteases

EXAMPLE 4

Cloning and characterization

Cloning and characterization of new gene candidates was undertaken to expand the panel representative of extracellular proteases specific for ovarian carcinoma subtypes. Sequencing of the PCR products derived from tumor cDNA confirms the potential candidacy of these genes. The three novel genes all have conserved residues within the catalytic triad sequence consistent with their membership in the serine protease family.

Applicant compared the PCR products amplified from normal and carcinoma cDNAs using sense-histidine and antisense-aspartate as well as sense-histidine and antisense-serine. The anticipated PCR products of approximately 200 bp and 500 bp for those pairs of primers were observed (aspartate is approximately 50-70 amino acids downstream from histidine, and serine is about 100-150 amino acids toward the carboxy end from histidine).

Figure 1 shows a comparison of PCR products derived from normal and carcinoma cDNA as shown by staining in an agarose gel. Two distinct bands in Lane 2 were present in the primer pair sense-His/antisense ASP (AS1) and multiple bands of about 500 bp are noted in the carcinoma lane for the sense-His/antisense-Ser (AS2) primer pairs in Lane 4.

EXAMPLE 5

Quantitative PCR

The mRNA overexpression of hepsin was detected and determined using quantitative PCR. Quantitative PCR was performed generally according to the method of Noonan et al. [*Proc.Natl.Acad.Sci.,USA*, 87:7160-7164 (1990)]. The following oligonucleotide primers were used:

hepsin:

forward 5'-TGTCCCGATGGCGAGTGTTT-3' (SEQ ID No. 8), and

reverse 5'-CCTGTTGGCCATAGTACTGC-3' (SEQ ID No. 9);

and β-tubulin:

forward 5'- TGCATTGACAACGAGGC -3' (SEQ ID No. 18), and

reverse 5'- CTGTCTTGA CATTGTTG -3' (SEQ ID No. 19).

β-tubulin was utilized as an internal control. The predicted sizes of the amplified genes were 282 bp for hepsin and 454 bp for β-tubulin. The primer sequences used in this study were designed according to the cDNA sequences described by Leytus *et al.* [*Biochemistry, 27,* 1067-1074 (1988)] for hepsin, and Hall *et al.* [*Mol. Cell. Biol., 3,* 854-862 (1983)] for β-tubulin. The PCR reaction mixture consisted of cDNA derived from 50 ng of mRNA converted by conventional techniques, 5 pmol of sense and antisense primers for both the hepsin gene and the β-tubulin gene, 200 μmol of dNTPs, 5 μCi of α-^{32}PdCTP and 0.25 units of Taq DNA polymerase with reaction buffer (Promega) in a final volume of 25 μl. The target sequences were amplified in parallel with the β-tubulin gene. Thirty cycles of PCR were carried out in a Thermal Cycler (Perkin-Elmer Cetus). Each cycle of PCR included 30 sec of denaturation at 95°C, 30 sec of annealing at 63°C and 30 sec of extension at 72°C. The PCR products were separated on 2% agarose gels and the radioactivity of each PCR product was determined by using a PhosphorImager™ (Molecular Dynamics). Student's *t* test was used for comparison of mean values.

Experiments comparing PCR amplification in normal ovary and ovarian carcinoma suggested overexpression and/or alteration in mRNA transcript in tumor tissues. Northern blot analysis of TADG-14 confirms a transcript size of 1.4 kb and data indicate overexpression in ovarian carcinoma (Figure 2). Isolation and purification using both PCR and a specific 250 bp PCR product to screen positive plaques yielded a 1.2 kb clone of TADG-14. Other proteases were amplified by the same method using the appropriate primers from Table 2.

EXAMPLE 6

Tissue bank

A tumor tissue bank of fresh frozen tissue of ovarian carcinomas as shown in Table 4 was used for evaluation. Approximately 100 normal ovaries removed for medical reasons other than malignancy were obtained from surgery and were available as controls.

TABLE 4

Ovarian cancer tissue bank

	Total	Stage I/11	Stage III/IV	No Stage
Serous				
Malignant	166	15	140	8
LMP	16	9	7	0
Benign	12	0	0	12
Mucinous				
Malignant	26	6	14	6
LMP	28	25	3	0
Benign	3	0	0	3
Endometrioid				
Malignant	38	17	21	0
LMP	2	2	0	0
Benign	0	0	0	0
Other*				
Malignant	61	23	29	9
LMP	0	0	0	0
Benign	5	0	0	5

*Other category includes the following tumor types: Brenner's tumor, thecoma, teratoma, fibrothecoma, fibroma, granulosa cell, clear cell, germ cell, mixed mullerian, stromal, undifferentiated, and dysgerminoma.

From the tumor bank, approximately 100 carcinomas were evaluated encompassing most histological sub-types of ovarian carcinoma, including borderline or low-malignant potential tumors and overt carcinomas. The approach included using mRNA prepared from fresh frozen tissue (both normal and malignant) to compare expression of genes in normal, low malignant potential tumors and overt carcinomas. The cDNA prepared from polyA$^+$ mRNA was deemed to be genomic DNA-free by checking all preparations with primers that encompassed a known intron-exon splice site using both β-tubulin and p53 primers.

EXAMPLE 7

Northern blots

Significant information can be obtained by examining the expression of these candidate genes by Northern blot. Analysis of normal adult multi-tissue blots offers the opportunity to identify normal tissues which may express the protease. Ultimately, if strategies for inhibition of proteases for therapeutic intervention are to be developed, it is essential to appreciate the expression of these genes in normal tissue if and when it occurs.

Significant information is expected from Northern blot analysis of fetal tissue. Genes overexpressed in carcinomas are often highly expressed in organogenesis. As indicated, the hepsin gene cloned from hepatoma cells and overexpressed in ovarian carcinoma is overtly expressed in fetal liver. Hepsin gene expression was also detected in fetal kidney, and therefore, could be a candidate for expression in renal carcinomas.

Northern panels for examining expression of genes in a multi-tissue normal adult as well as fetal tissue are commercially available (CLONTECH). Such evaluation tools are not only important to confirm the overexpression of individual transcripts in tumor versus normal tissues, but also provides the opportunity to confirm transcript size, and to determine if alternate splicing or other transcript alteration may occur in ovarian carcinoma.

EXAMPLE 8

Northern blot analysis

Northern blot analysis was performed as follows: 10 µg of mRNA was loaded onto a 1% formaldehyde-agarose gel, electrophoresed and blotted onto a HyBond-N+™ nylon membrane (Amersham). ^{32}P-labeled cDNA probes were made using Prime-a-Gene Labeling System™ (Promega). The PCR products amplified by specific primers were used as probes. Blots were prehybridized for 30 min and then hybridized for 60 min at 68°C with ^{32}P-labeled cDNA probe in ExpressHyb™ Hybridization Solution (CLONTECH). Control hybridization to determine relative gel loading was accomplished using the β-tubulin probe.

Normal human tissues including spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocyte, heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas and normal human fetal tissues; brain, lung, liver and kidney (Human Multiple Tissue Northern Blot; CLONTECH) were all examined using the same hybridization procedure.

EXAMPLE 9

PCR products corresponding to serine, cysteine and metallo-proteases

Based on their unique expression in either low malignant potential tumors or carcinomas, PCR-amplified cDNA products were cloned and sequenced and the appropriate gene identified based upon nucleotide and amino acid sequences stored in the GCG and EST databases. Figures 3, 4 & 5 show the PCR product displays comparing normal and carcinomatous tissues using redundant primers for serine proteases (Figure 3), for cysteine proteases (Figure 4) and for metallo-proteases (Figure 5). Note the differential expression in the carcinoma tissues versus the normal tissues. The proteases were identified using redundant cDNA primers (see Table 2) directed towards conserved sequences that are associated with intrinsic enzyme activity (for serine proteases, cysteine proteases and metallo-proteases) by comparing mRNA expression in normal, low malignant potential and overt ovarian carcinoma tissues according to Sakanari *et al.* [*Biochemistry* *86*, 4863-4867 (1989)].

EXAMPLE 10

Serine proteases

For the serine protease group, using the histidine domain primer sense, S1, in combination with antisense primer AS2, the following proteases were identified:

(a) Hepsin, a trypsin-like serine protease cloned from hepatoma cells shown to be a cell surface protease essential for the growth of hepatoma cells in culture and highly expressed in hepatoma tumor cells (Figure 3, Lane 4);

(b) Complement factor B protease (human factor IX), a protease involved in the coagulation cascade and associated with the production and accumulation of fibrin split products associated with tumor cells (Figure 3, Lane 4). Compliment factor B belongs in the family of coagulation factors X (Christmas factor). As part of the intrinsic pathway, compliment factor B catalyzes the proteolytic activation of coagulation factor X in the presence of Ca^{2+} phospholipid and factor VIIIa e5; and

(c) A stratum corneum chymotryptic enzyme (SCCE) serine protease involved in desquarnation of skin cells from the human stratum corneum (Figure 3, Lane 4). SCCE is expressed in keratinocytes of the epidermis and functions to degrade the cohesive structures in the cornified layer to allow continuous skin surface shedding.

EXAMPLE 11

Cysteine proteases

In the cysteine protease group, using redundant sense and anti-sense primers for cysteine proteases, one unique PCR product was identified by overexpression in ovarian carcinoma when compared to normal ovarian tissue (Figure 4, Lanes 3-5). Cloning and sequencing this PCR product identified a sequence of

Cathepsin L, which is a lysomal cysteine protease whose expression and secretion is induced by malignant transformation, growth factors and tumor promoters. Many human tumors (including ovarian) express high levels of Cathepsin L. Cathepsin L cysteine protease belongs in the stromolysin family and has potent elastase and collagenase activities. Published data indicates increased levels in the serum of patients with mucinous cystadenocarcinoma of the ovary. It has not heretofore been shown to be expressed in other ovarian tumors.

EXAMPLE 12

Metallo-proteases

Using redundant sense and anti-sense primers for the metallo-protease group, one unique PCR product was detected in the tumor tissue which was absent in normal ovarian tissue (Figure 5, Lanes 2-5). Subcloning and sequencing this product indicates it has complete homology in the appropriate region with the so-called PUMP-1 (MMP-7) gene. This zinc-binding metallo-protease is expressed as a proenzyme with a signal sequence and is active in gelatin and collagenase digestion. PUMP-1 has also been shown to be induced and overexpressed in 9 of 10 colorectal carcinomas compared to normal colon tissue, suggesting a role for this substrate in the progression of this disease.

EXAMPLE 13

Expression of hepsin

The expression of the serine protease hepsin gene in 8 normal, 11 low malignant potential tumors, and 14 carcinoma (both mucinous and serous type) by quantitative PCR using hepsin-specific primers (see Table 2) was determined (primers directed toward the β-tubulin message were used as an internal standard) (Table 5). These data confirm the overexpression of the hepsin surface protease gene in ovarian carcinoma, including both low malignant potential tumors and overt carcinoma. Expression of hepsin is increased over normal levels in low malignant potential tumors, and high stage tumors (Stage III) of this group have higher expression of hepsin when compared to low stage tumors (Stage 1) (Table 6). In overt carcinoma, serous tumors exhibit the highest levels of hepsin expression, while mucinous tumors express levels of hepsin comparable with the high stage low malignant potential group (Figures 6 & 7).

TABLE 5

Patient Characteristics and Expression of Hepsin Gene

Case	Histological type[a]	Stage/Grade	LN[b]	mRNA expression of hepsin[c]
1	normal ovary			n
2	normal ovary			n
3	normal ovary			n
4	normal ovary			n
5	normal ovary			n

6	normal ovary			n
7	normal ovary			n
8	normal ovary			n
9	normal ovary			n
10	normal ovary			n
11	S adenoma (LMP)	1 / 1	N	4+
12	S adenoma (LMP)	1 / 1	NE	4+
13	S adenoma (LMP)	1 / 1	NE	n
14	S adenoma (LMP)	1 / 1	N	2+
15	S adenoma (LMP)	3 / 1	P	4+
16	S adenoma (LMP)	3 / 1	P	4+
17	S adenoma (LMP)	3 / 1	P	4+
18	M adenoma (LMP)	1 / 1	NE	4+
19	M adenoma (LMP)	1 / 1	N	n
20	M adenoma (LMP)	1 / 1	N	n
21	M adenoma (LMP)	1 / 1	N	n
22	M adenoma (LMP)	1 / 1	NE	n
23	S carcinoma	1 / 2	N	4+
24	S carcinoma	1 / 3	N	4+
25	S carcinoma	3 / 1	NE	2+
26	S carcinoma	3 / 2	NE	4+
27	S carcinoma	3 / 2	P	4+
28	S carcinoma	3 / 2	NE	2+
29	S carcinoma	3 / 3	NE	2+
30	S carcinoma	3 / 3	NE	4+
31	S carcinoma	3 / 3	NE	4+
32	S carcinoma	3 / 3	NE	4+
33	S carcinoma	3 / 3	N	4+

34	S	carcinoma	3/3	NE	n
35	S	carcinoma	3/3	NE	4+
36	S	carcinoma	3/3	NE	4+
37	S	carcinoma	3/3	NE	4+
38	S	carcinoma	3/3	N	4+
39	S	carcinoma	3/2	NE	2+
40	S	carcinoma	3/3	NE	4+
41	S	carcinoma	3/2	NE	4+
42	M	carcinoma	1/2	N	n
43	M	carcinoma	2/2	NE	4+
44	M	carcinoma	2/2	N	4+
45	M	carcinoma	3/1	NE	n
46	M	carcinoma	3/2	NE	4+
47	M	carcinoma	3/2	NE	n
48	M	carcinoma	3/3	NE	n
49	E	carcinoma	2/3	N	4+
50	E	carcinoma	3/2	NE	4+
51	E	carcinoma	3/3	NE	4+
52	C	carcinoma	1/3	N	4+
53	C	carcinoma	1/1	N	4+
54	C	carcinoma	3/2	P	4+

[a]S, serous; M, mucinous; E, endometrioid; C, clear cell; [b]LN, lymph node metastasis; P, positive; N, negative; NE, not examined; [c]n, normal range = mean ±2SD; 2+, mean ±2SD to ±4SD; 4+, mean ±4SD or greater.

TABLE 6

Overexpression of hepsin in normal ovaries and ovarian tumors

Type	N	Hepsin Overexpression	Ratio of Hepsin to β-tubulin
Normal	10	0 (0%)	0.06 ± 0.05
LMP	12	7 (58.3%)	0.26 ± 0.19
Serous	7	6 (85.7%)	0.34 ± 0.20
Mucinous	5	1 (20.0%)	0.14 ± 0.12
Carcinomous	32	27 (84.4%)	0.46 ± 0.29
Serous	19	18 (94.7%)	0.56 ± 0.32
Mucinous	7	3 (42.9%)	0.26 ± 0.22
Endometrioid	3	3 (100%)	0.34 ± 0.01
Clear Cell	3	3 (100%)	0.45 ± 0.08

EXAMPLE 14

Expression of SCCE and PUMP-1

Studies using both SCCE-specific primers (Figure 8) and PUMP-specific primers (Figure 9) indicate overexpression of these proteases in ovarian carcinomas.

EXAMPLE 15

Summary of known proteases detected herein

Most of the proteases described herein were identified from the sense-His/antisense-Ser primer pair, yielding a 500 bp PCR product (Figure 1, Lane 4). Some of the enzymes are familiar, a short summary of each follows.

Hepsin

Hepsin is a trypsin-like serine protease cloned from hepatoma cells. Hepsin is an extracellular protease (the enzyme includes a secretion signal sequence) which is anchored in the plasma membrane by its amino terminal domain, thereby exposing its catalytic domain to the extracellular matrix. Hepsin has also been shown to be expressed in breast cancer cell lines and peripheral nerve cells. Hepsin has never before been associated with ovarian carcinoma. Specific primers for the hepsin gene were synthesized and the expression of hepsin examined using Northern blots of fetal tissue and ovarian tissue (both normal and ovarian carcinoma).

Figure 10A shows that hepsin was expressed in ovarian carcinomas of different histologic types, but not in normal ovary. Figure 10B shows that hepsin was expressed in fetal liver and fetal kidney as anticipated, but at very low levels or not at all in fetal brain and lung. Figure 10C shows that hepsin overexpression is not observed in normal adult tissue. Slight expression above the background level is observed in the adult prostate. The mRNA identified in both Northern blots was the appropriate size for the hepsin transcript. The expression of hepsin was examined in 10 normal ovaries and 44 ovarian tumors using specific primers to β-tubulin and hepsin in a quantitative PCR assay, and found it to be linear over 35 cycles. Expression is presented as the ratio of ^{32}P-hepsin band to the internal control, the ^{32}P-β-tubulin band.

Hepsin expression was investigated in normal (N), mucinous (M) and serous (S) low malignant potential (LMP) tumors and carcinomas (CA). Figure 11A shows quantitative PCR of hepsin and internal control β-tubulin. Figure 11B shows the ratio of hepsin:β-tubulin expression in normal ovary, LMP tumor, and ovarian carcinoma. It was observed that Hepsin mRNA expression levels were significantly elevated in LMP tumors, ($p < 0.005$) and carcinomas ($p < 0.0001$) compared to levels in normal ovary. All 10 cases of normal ovaries showed a relatively low level of hepsin mRNA expression.

Hepsin mRNA is highly overexpressed in most histopathologic types of ovarian carcinomas including some low malignant potential tumors (see Figures 11A & 11B). Most noticeably, hepsin is highly expressed in serous, endometrioid and clear cell tumors tested. It is highly expressed in some mucinous tumors, but it is not overexpressed in the majority of such tumors.

Stratum corneum chymotrypsin enzyme (SCCE)

The PCR product identified was the catalytic domain of the sense-His/antisense-Ser of the stratum corneum chymotrypsin enzyme. This extracellular protease was cloned, sequenced and shown to be expressed on the surface of keratinocytes in the epidermis. Stratum corneum chymotrypsin enzyme is a chymotrypsin-like serine protease whose function is suggested to be in the catalytic degradation of intercellular cohesive structures in the stratum corneum layer of the skin. This degradation allows continuous shedding (desquamation) of cells from the skin surface. The subcellular localization of stratum corneum chymotrypsin enzyme is in the upper granular layer in the stratum corneum of normal non-palmoplantar skin and in the cohesive parts of hypertrophic plantar stratum corneum. Stratum corneum chymotrypsin enzyme is exclusively associated with the stratum corneum and has not so far been shown to be expressed in any carcinomatous tissues.

Northern blots were probed with the PCR product to determine expression of stratum corneum chymotrypsin enzyme in fetal tissue and ovarian carcinoma (Figures 12A & 12B). Noticeably, detection of stratum corneum chymotrypsin enzyme messenger RNA on the fetal Northern was almost non-existent (a problem with the probe or the blot was excluded by performing the proper controls). A faint band appeared in fetal kidney. On the other hand, stratum corneum chymotrypsin enzyme mRNA is abundant in the ovarian carcinoma mRNA (Figure 12B). Two transcripts of the correct size are observed for stratum corneum chymotrypsin enzyme. The same panel of cDNA used for hepsin analysis was used for stratum corneum chymotrypsin enzyme expression.

No stratum corneum chymotrypsin enzyme expression was detected in the normal ovary lane of the Northern blot. A comparison of all candidate genes, including a loading marker (β-tubulin), was shown to confirm that this observation was not a result of a loading bias. Quantitative PCR using stratum corneum chymotrypsin enzyme primers, along with β-tubulin internal control primers, confirmed the overexpression of stratum corneum chymotrypsin enzyme mRNA in carcinoma of the ovary with no expression in normal ovarian tissue (Figure 13).

Figure 13A shows a comparison using quantitative PCR of stratum corneum chymotrypsin enzyme cDNA from normal ovary and ovarian carcinomas. Figure 13B shows the ratio of stratum corneum chymotrypsin enzyme to the β-tubulin internal standard in 10 normal and 44 ovarian carcinoma tissues. Again, it is observed that stratum corneum chymotrypsin enzyme is highly overexpressed in ovarian carcinoma cells. It is also noted that some mucinous tumors overexpress stratum corneum chymotrypsin enzyme, but the majority do not.

Protease M

Protease M was identified from subclones of the His--ser primer pair. This protease was first cloned by Anisowicz, *et al.*, [*Molecular Medicine,* 2, 624-636 (1996)] and shown to be overexpressed in carcinomas. A preliminary evaluation indicates that this enzyme is overexpressed in ovarian carcinoma (Figure 14).

Cofactor I and Complement factor B

Several serine proteases associated with the coagulation pathway were also subcloned. Examination of normal and ovarian carcinomas by quantitative PCR for expression of these enzymes, it was noticeable that this mRNA was not clearly overexpressed in ovarian carcinomas when compared to normal ovarian tissue. It should be noted that the same panel of tumors was used for the evaluation of each candidate protease.

EXAMPLE 16

Summary of previously unknown proteases detected herein

TADG-12

TADG-12 was identified from the primer pairs, sense-His/antisense-Asp (see Figure 1, Lanes 1 & 2). Upon subcloning both PCR products in lane 2, the 200 bp product had a unique protease-like sequence not included in GenBank. This 200 bp product contains many of the conserved amino acids common for the His-Asp domain of the family of serine proteins. The second and larger PCR product (300 bp) was shown to have a high degree of homology with TADG-12 (His-Asp sequence), but also contained approximately 100 bp of unique sequence. Synthesis of specific primers and the sequencing of the subsequent PCR products from three different tumors demonstrated that the larger PCR product (present in about 50% of ovarian carcinomas) includes an insert of about 100 bp near the 5' end (and near the histidine) of the sequence. This insert may be a retained genomic intron because of the appropriate position of splice sites and the fact that the insert does not contain an open reading frame (see Figure 15). This suggests the possibility of a splice site mutation which gives rise to retention of the intron, or a translocation of a sequence into the TADG-12 gene in as many as half of all ovarian carcinomas.

TADG-13 and TADG-14

Specific primers were synthesized for TADG-13 and TADG-14 to evaluate expression of genes in normal and ovarian carcinoma tissue. Northern blot analysis of ovarian tissues indicates the transcript for the TADG-14 gene is approximately 1.4 kb and is expressed in ovarian carcinoma tissues (Figure 16A) with no noticeable transcript presence in normal tissue. In quantitative PCR studies using specific primers, increased expression of TADG-14 in ovarian carcinoma tissues was noted compared to a normal ovary (Figure 16B). The presence of a specific PCR product for TADG-14 in both an HeLa library and an ovarian carcinoma library was also confirmed. Several candidate sequences corresponding to TADG-14 have been screened and isolated from the HeLa library.

Clearly from sequence homology, these genes fit into the family of serine proteases. TADG-13 and -14 are, however, heretofore undocumented genes which the specific primers of the invention allow to be evaluated in normal and tumor cells, and with which the presence or absence of expression of these genes is useful in the diagnosis or treatment selection for specific tumor types.

PUMP-1

In a similar strategy using redundant primers to metal binding domains and conserved histidine domains, a differentially expressed PCR product identical to matrix metallo-protease 7 (MMP-7) was identified, herein called PUMP-1. Using specific primers for PUMP-1, PCR produced a 250 bp product for Northern blot analysis.

PUMP-1 is differentially expressed in fetal lung and kidney tissues. Figure 17A shows the expression of PUMP-1 in human fetal tissue, while no transcript could be detected in either fetal brain or fetal liver. Figure 17B compares PUMP-1 expression in normal ovary and carcinoma subtypes using Northern blot analysis. Notably, PUMP-1 is expressed in ovarian carcinoma tissues, and again, the presence of a transcript in normal tissue was not detected. Quantitative PCR comparing normal versus ovarian carcinoma expression of the PUMP-1 mRNA indicates that this gene is highly expressed in serous carcinomas, including most low malignant serous tumors, and is, again, expressed to a lesser extent in mucinous tumors (see Figures 18A & 18B). PUMP-1, however, is so far the protease most frequently found overexpressed in mucinous tumors (See Table 7).

Cathepsin-L

Using redundant cysteine protease primers to conserved domains surrounding individual cysteine and histidine residues, the cathepsin-L protease was identified in several serous carcinomas. An initial examination of the expression of cathepsin L in normal and ovarian tumor tissue indicates that transcripts for the cathepsin-L protease are present in both normal and tumor tissues (Figure 19). However, its presence or absence in combination with other proteases of the present invention permits identification of specific tumor types and treatment choices.

Discussion

Redundant primers to conserved domains of serine, metallo-, and cysteine proteases have yielded a set of genes whose mRNAs are overexpressed in ovarian carcinoma. The genes which are clearly overexpressed include the serine proteases hepsin, stratum corneum chymotrypsin enzyme, protease M TADG12, TADG14 and the metallo-protease PUMP-1 (see Figure 19 and Table 7). Northern blot analysis of normal and ovarian carcinoma tissues, summarized in Figure 14, indicated overexpression of hepsin, stratum corneum chymotrypsin enzyme, PUMP-1 and TADG-14. A β-tubulin probe to control for loading levels was included.

TABLE 7

Overexpression of Proteases in Ovarian Tumors

Type	N	Hepsin	SCCE	Pump-1	Protease M
Normal	10	0% (0/10)	0% (0/10)	0% (0/10)	0% (0/10)
LMP	12	58.3% (7/12)	66.7% (8/12)	75.0% (9/12)	75% (9/12)
serous	7	85.7% (6/7)	85.7% (6/7)	85.7% (6/7)	100% (7/7)
mucinous	5	20.0% (1/5)	40.0% (2/5)	60% (3/5)	40.0%(2/5)
Carcinoma	32	84.4% (27/32)	78.1% (25/32)	81.3% (26/32)	90.6% (29/32
serous	19	94.7%(18/19)	89.5%(17/19)	78.9% (15/19)	94.7% (18/19
mucinous	7	42.9%(3/7)	28.6%(2/7)	71.4% (5/7)	85.7% (6/7)
endometr.	3	100% (3/3)	100%(3/3)	100% (3/3)	100% (3/3)
clear cell	3	100% (3/3)	100% (3/3)	100% (3/3)	67.7% (2/3)

For the most part, these proteins previously have not been associated with the extracellular matrix of ovarian carcinoma cells. No panel of proteases which might contribute to the growth, shedding, invasion and colony development of metastatic carcinoma has been previously described, including the three new candidate serine proteases which are herein disclosed. The establishment of an extracellular protease panel associated with either malignant growth or malignant potential offers the opportunity for the identification of diagnostic or prognostic markers and for therapeutic intervention through inhibition or down regulation of these proteases.

The availability of the instant gene-specific primers coding for the appropriate region of tumor specific proteases allows for the amplification of a specific cDNA probe using Northern and Southern analysis, and their use as markers to detect the presence of the cancer in tissue. The probes also allow more extensive evaluation of the expression of the gene in normal ovary versus low malignant potential tumor, as well as both high- and low-stage carcinomas. The evaluation of a panel of fresh frozen tissue from all the carcinoma subtypes (Table 4) allowed the determination of whether a protease is expressed predominantly in early stage disease or within specific carcinoma subtypes. It was also determined whether each gene's expression is confined to a particular stage in tumor progression and/or is associated with metastatic lesions. Detection of specific combinations of proteases is an identifying characteristic of the specific tumor types and yields valuable information for diagnoses and treatment selection. Particular tumor types may be more accurately diagnosed by the characteristic expression pattern of each specific tumor.

EXAMPLE 17

Peptide ranking

For vaccine or immune stimulation, individual 9-mers to 11-mers of the hepsin protein were examined to rank the binding of individual peptides to the top 8 haplotypes in the general population (Parker et al., (1994)). The computer program used for this analyses can be found at <http://www-bimas.dcrt.nih.gov/molbio/hla_bind/>. Table 8 shows the peptide ranking based upon the predicted half-life of each peptide's binding to a particular HLA allele. A larger half-life indicates a stronger association with that peptide and the particular HLA molecule. The hepsin peptides that strongly bind to an HLA allele are putative immunogens, and are used to innoculate an individual against hepsin.

TABLE 8

Hepsin peptide ranking

HLA Type & Ranking	Start	Peptide	Predicted Dissociation$_{1/2}$	SEQ ID No.
HLA A0201				
1	170	SLGRWPWQV	521.640	28
2	191	SLLSGDWVL	243.051	29
3	229	GLQLGVQAV	159.970	30
4	392	KVSDFREWI	134.154	31
5	308	VLQEARVPI	72.717	32
6	130	RLLEVISVC	71.069	33
7	98	ALTHSELDV	69.552	34
8	211	VLSRWRVFA	46.451	35
9	26	LLLLTAIGA	31.249	36
10	284	ALVDGKICT	30.553	37
11	145	FLAAICQDC	22.853	38
12	192	LLSGDWVLT	21.536	39
13	20	ALTAGTLLL	21.362	40

14	259	ALVHLSSPL	21.362	41
15	277	CLPAAGQAL	21.362	42
16	230	LQLGVQAVV	18.186	43
17	268	PLTEYIQPV	14.429	44
18	31	AIGAASWAI	10.759	45
19	285	LVDGKICTV	9.518	46
20	27	LLLTAIGAA	9.343	47
HLA A0205				
1	191	SLLSGDWVL	25.200	48
2	163	IVGGRDTSL	23.800	49
3	392	KVSDFREWI	18.000	50
4	64	MVFDKTEGT	15.300	51
5	236	AVVYHGGYL	14.000	52
6	55	QVSSADARL	14.000	53
7	130	RLLEVISVC	9.000	54
8	230	LQLGVQAVV	8.160	55
9	20	ALTAGTLLL	7.000	56
10	259	ALVHLSSPL	7.000	57
11	277	CLPAAGQAL	7.000	58
12	17	KVAALTAGT	6.000	59
13	285	LVDGKICTV	5.440	60
14	308	VLQEARVPI	5.100	61
15	27	LLLTAIGAA	5.100	62
16	229	GLQLGVQAV	4.000	63
17	313	RVPIISNDV	4.000	64
18	88	LSCEEMGFL	3.570	65
19	192	LLSGDWVLT	3.400	66
20	284	ALVDGKICT	3.000	67

HLA A1

1	89	SCEEMGFLR	45.000	68
2	58	SADARLMVF	25.000	69
3	393	VSDFREWIF	7.500	70
4	407	HSEASGMVT	6.750	71
5	137	VCDCPRGRF	5.000	72
6	269	LTEYIQPVC	4.500	73
7	47	DQEPLYPVQ	2.700	74
8	119	CVDEGRLPH	2.500	75
9	68	KTEGTWRLL	2.250	76
10	101	HSELDVRTA	1.350	77
11	250	NSEENSNDI	1.350	78
12	293	VTGWGNTQY	1.250	79
13	231	QLGVQAVVY	1.000	80
14	103	ELDVRTAGA	1.000	81
15	378	GTGCALAQK	1.000	82
16	358	VCEDSISRT	0.900	83
17	264	SSPLPLTEY	0.750	84
18	87	GLSCEEMGF	0.500	85
19	272	YIQPVCLPA	0.500	86
20	345	GIDACQGDS	0.500	87

HLA A24

1	301	YYGQQAGVL	200.000	88
2	238	VYHGGYLPF	100.000	89
3	204	CFPERNRVL	36.000	90
4	117	FFCVDEGRL	20.000	91
5	124	RLPHTQRLL	12.000	92
6	80	RSNARVAGL	12.000	93

	7	68	KTEGTWRLL1	2.000	94
	8	340	GYPEGGIDA	9.000	95
	9	242	GYLPFRDPN	9.000	96
	10	51	LYPVQVSSA	7.500	97
	11	259	ALVHLSSPL	7.200	98
	12	277	CLPAAGQAL	7.200	99
	13	191	SLLSGDWVL	6.000	100
	14	210	RVLSRWRVF	6.000	101
	15	222	VAQASPHGL	6.000	102
	16	236	AVVYHGGYL	6.000	103
	17	19	AALTAGTLL	6.000	104
	18	36	SWAIVAVLL	5.600	105
	19	35	ASWAIVAVL	5.600	106
	20	300	QYYGQQAGV	5.600	107
HLA B7					
	1	363	ISRTPRWRL	90.000	108
	2	366	TPRWRLCGI	80.000	109
	3	236	AVVYHGGYL	60.000	110
	4	13	CSRPKVAAL	40.000	111
	5	179	SLRYDGAHL	40.000	112
	6	43	LLRSDQEPL	40.000	113
	7	19	AALTAGTLL	36.000	114
	8	55	QVSSADARL	20.000	115
	9	163	IVGGRDTSL	20.000	116
	10	140	CPRGRFLAA	20.000	117
	11	20	ALTAGTLLL	12.000	118
	12	409	EASGMVTQL	12.000	119
	13	259	ALVHLSSPL	12.000	120

14	35	ASWAIVAVL	12.000	121
15	184	GAHLCGGSL	12.000	122
16	18	VAALTAGTL	12.000	123
17	222	VAQASPHGL	12.000	124
18	224	QASPHGLQL	12.000	125
19	265	SPLPLTEYI	8.000	126
20	355	GPFVCEDSI	8.00	127
HLA B8				
1	13	CSRPKVAAL	80.000	128
2	366	TPRWRLCGI	80.000	129
3	140	CPRGRFLAA	16.000	130
4	152	DCGRRKLPV	4.800	131
5	363	ISRTPRWRL	4.000	132
6	163	IVGGRDTSL	4.000	133
7	331	QIKPKMFCA	4.000	134
8	80	RSNARVAGL	2.000	135
9	179	SLRYDGAHL	1.600	136
10	43	LLRSDQEPL	1.600	137
11	409	EASGMVTQL	1.600	138
12	311	EARVPIISN	0.800	139
13	222	VAQASPHGL	0.800	140
14	19	AALTAGTLL	0.800	141
15	18	VAALTAGTL	0.800	142
16	184	GAHLCGGSL	0.800	143
17	224	QASPHGLQL	0.800	144
18	82	NARVAGLSC	0.800	145
19	204	CFPERNRVL	0.600	146
20	212	LSRWRVFAG	0.400	147

HLA B2702

1	172	GRWPWQVSL	300.000	148
2	44	LRSDQEPLY	200.00	149
3	155	RRKLPVDRI	180.000	150
4	213	SRWRVFAGA	100.000	151
5	166	GRDTSLGRW	100.000	152
6	369	WRLCGIVSW	100.000	153
7	180	LRYDGAHLC	100.000	154
8	96	LRALTHSEL	60.000	155
9	396	FREWIFQAI	60.000	156
10	123	GRLPHTQRL	60.000	157
11	207	ERNRVLSRW	30.000	158
12	209	NRVLSRWRV	20.000	159
13	14	SRPKVAALT	20.000	160
14	106	VRTAGANGT	20.000	161
15	129	QRLLEVISV	20.000	162
16	349	CQGDSGGPF	20.000	163
17	61	ARLMVFDKT	20.000	164
18	215	WRVFAGAVA	20.000	165
19	143	GRFLAAICQ	10.000	166
20	246	FRDPNSEEN	10.000	167

HLA B4403

1	132	LEVISVCDC	36.000	168
2	91	EEMGFLRAL	18.000	169
3	264	SSPLPLTEY	13.500	170
4	310	QEARVPIIS	12.000	171
5	319	NDVCNGADF	10.000	172
6	4	KEGGRTVPC	9.000	173

7	251	SEENSNDIA	8.000	174
8	256	NDIALVHLS	7.500	175
9	294	TGWGNTQYY	6.750	176
10	361	DSISRTPRW	6.750	177
11	235	QAVVYHGGY	6.000	178
12	109	AGANGTSGF	6.000	179
13	270	TEYIQPVCL	6.000	180
14	174	WPWQVSLRY	4.500	181
15	293	VTGWGNTQY	4.500	182
16	69	TEGTWRLLC	4.000	183
17	90	CEEMGFLRA	4.000	184
18	252	EENSNDIAL	4.000	185
19	48	QEPLYPVQV	4.000	186
20	102	SELDVRTAG	3.600	187

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. Further, these patents and publications are incorporated by reference herein to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

One skilled in the art will appreciate readily that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those objects, ends and advantages inherent herein. The present examples, along with the methods, procedures, treatments, molecules, and specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention as defined by the scope of the claims.

Where the terms "comprise", "comprises", "comprised" or "comprising" are used in this specification, they are to be interpreted as specifying the presence of the stated features, integers, steps or components referred to, but not to preclude the presence or addition of one or more other feature, integer, step, component or group thereof.

The claims defining the invention are as follows:

1. A method of vaccinating an individual against hepsin, comprising the step of:

inoculating an individual with a hepsin protein or fragment thereof, wherein said hepsin protein or fragment thereof lack hepsin protease activity, wherein said inoculation with said hepsin protein or fragment thereof elicits an immune response in said individual, thereby vaccinating said individual against hepsin.

2. The method of claim 1, wherein said individual has cancer, is suspected of having cancer or is at risk of getting cancer.

3. The method of claim 1 or claim 2, wherein said hepsin fragment is selected from the group consisting of a 9-residue fragment up to a 20-residue fragment.

4. The method of claim 3, wherein said 9-residue fragment is selected from the group consisting of SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 and 154.

5. A method of producing immune-activated cells directed towards hepsin, comprising the steps of:

exposing dendritic cells to a hepsin protein or fragment thereof, wherein said hepsin protein or fragment thereof lacks hepsin protease activity, wherein said exposure to said hepsin protein or fragment thereof activates said dendritic cells, thereby producing immune-activated cells directed toward hepsin.

6. The method of claim 5, wherein said immune-activated cells are selected from the group consisting of B-cells, T-cells and dendrites.

7. The method of claim 5 or claim 6, wherein said hepsin fragment is selected from the group consisting of a 9-residue fragment up to a 20-residue fragment.

8. The method of claim 7, wherein said 9-residue fragment is selected from the group consisting of SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 and 154.

9. The method of claim 5, wherein said dendritic cells are isolated from an individual prior to said exposure, wherein said activated dendritic cells are reintroduced into said individual subsequent to said exposure.

10. The method of claim 9, wherein said individual has a cancer, is suspected of having a cancer or is at risk of getting a cancer.

11. An immunogenic composition, comprising an immunogenic fragment of a hepsin protein and an appropriate adjuvant.

12. The immunogenic composition of claim 11, wherein said hepsin fragment is selected from the group consisting of a 9-residue fragment up to a 20-residue fragment.

13. The immunogenic composition of claim 12, wherein said 9-residue fragment is selected from the group consisting of SEQ ID Nos. 28, 29, 30, 31, 88, 89, 108, 109, 128, 129, 148, 149, 150, 151, 152, 153 and 154.

14. An oligonucleotide having a sequence complementary to SEQ ID No. 188.

15. A composition comprising the oligonucleotide of claim 14 and a physiologically acceptable carrier.

16. The method of claim 1 substantially as hereinbefore described in any one of the Examples.

DATED this 20th day of August, 2004

THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS
By their Patent Attorneys:
CALLINAN LAWRIE

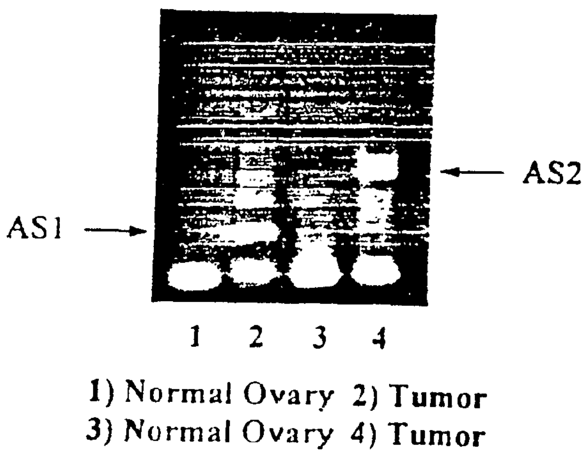

Fig. 1

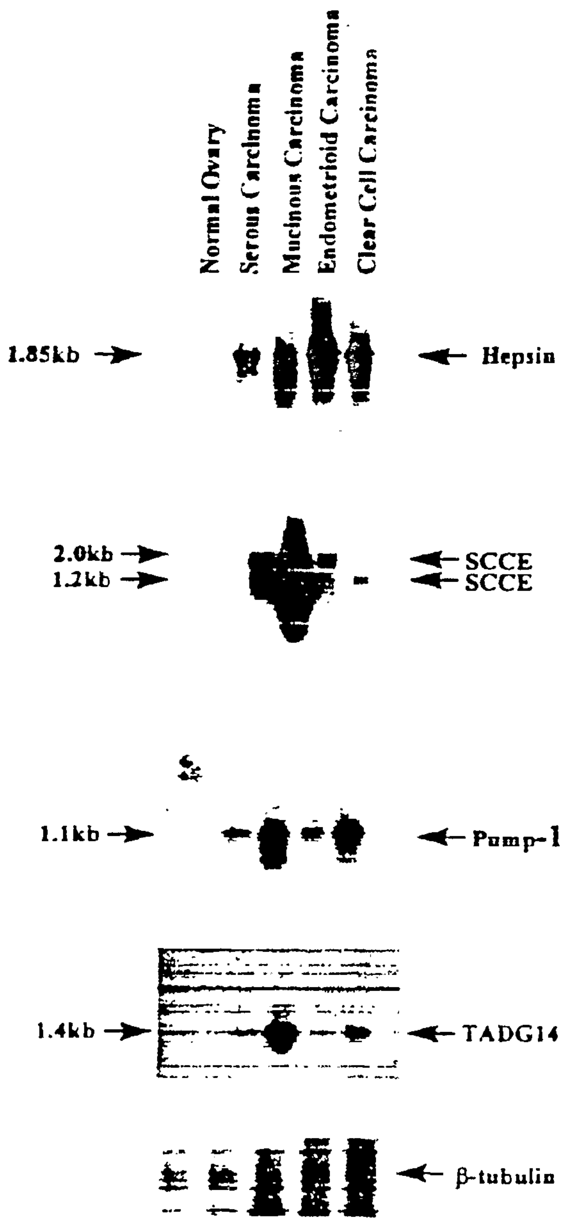
Normal Ovary
Serous Carcinoma
Mucinous Carcinoma
Endometrioid Carcinoma
Clear Cell Carcinoma
1.85kb
Hepsin
2.0kb
1.2kb
SCCE
SCCE
1.1kb
Pump-1
1.4kb
TADG14
β-tubulin

Fig. 2

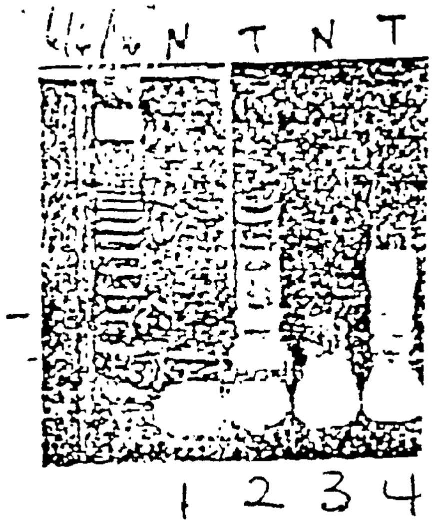

Fig. 3

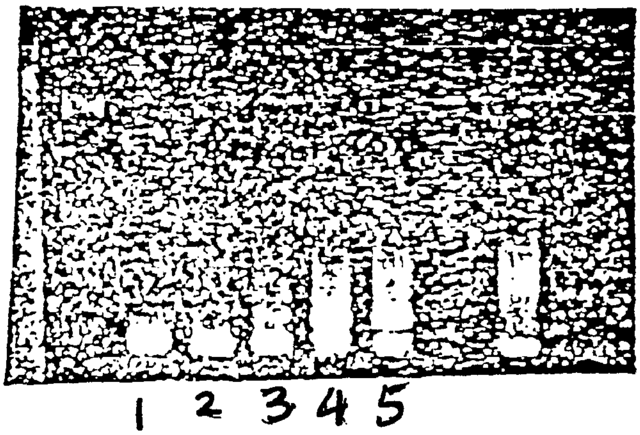

Fig. 4

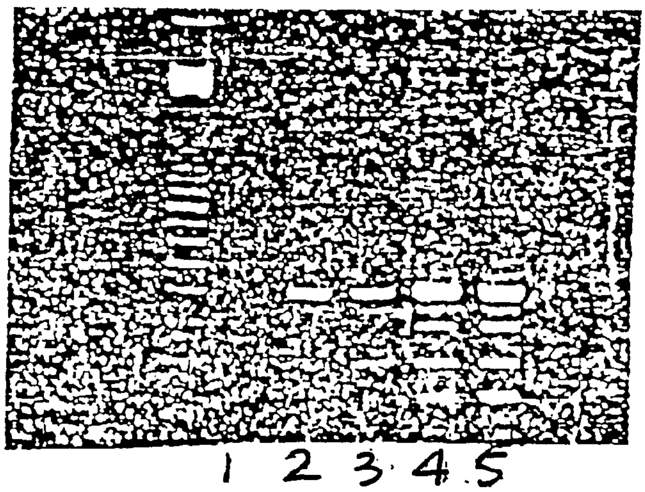

Fig. 5

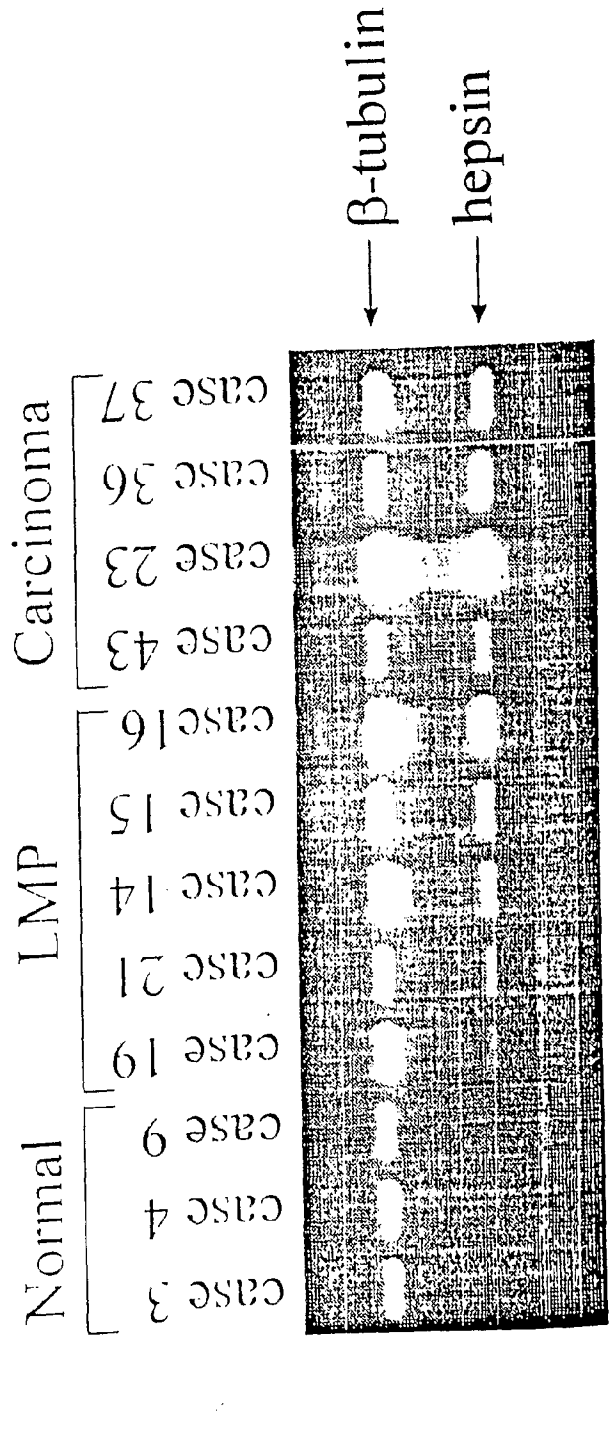
β-tubulin
hepsin
Carcinoma
case 37
case 36
case 23
case 43
LMP
case16
case 15
case 14
case 21
case 19
Normal
case 9
case 4
case 3

Fig. 6

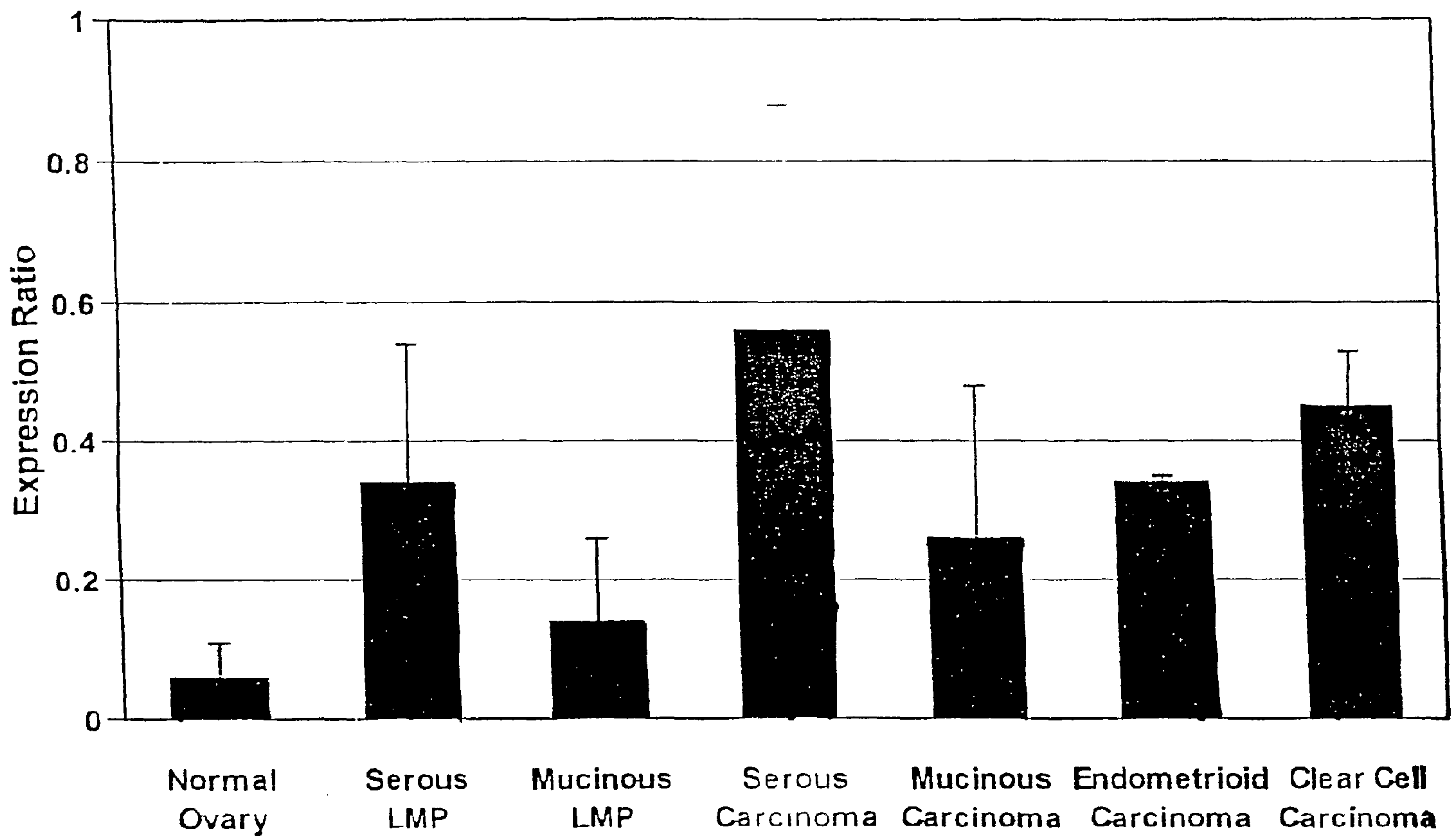
Expression Ratio
1
0.8
0.6
0.4
0.2
0
Normal Ovary
Serous LMP
Mucinous LMP
Serous Carcinoma
Mucinous Carcinoma
Endometrioid Carcinoma
Clear Cell Carcinoma

Fig. 7

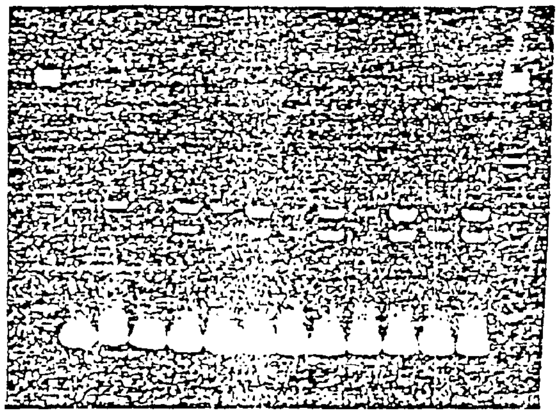

Fig. 8

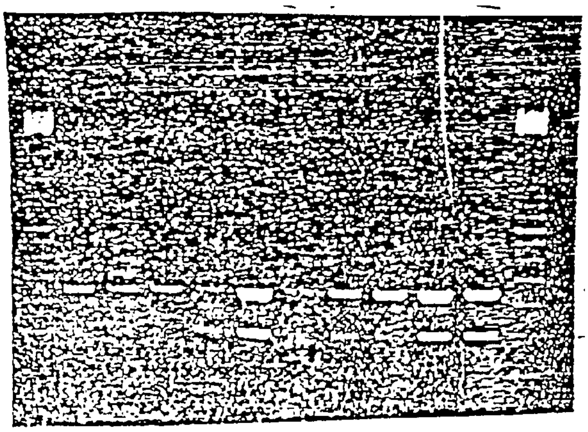

Fig. 9

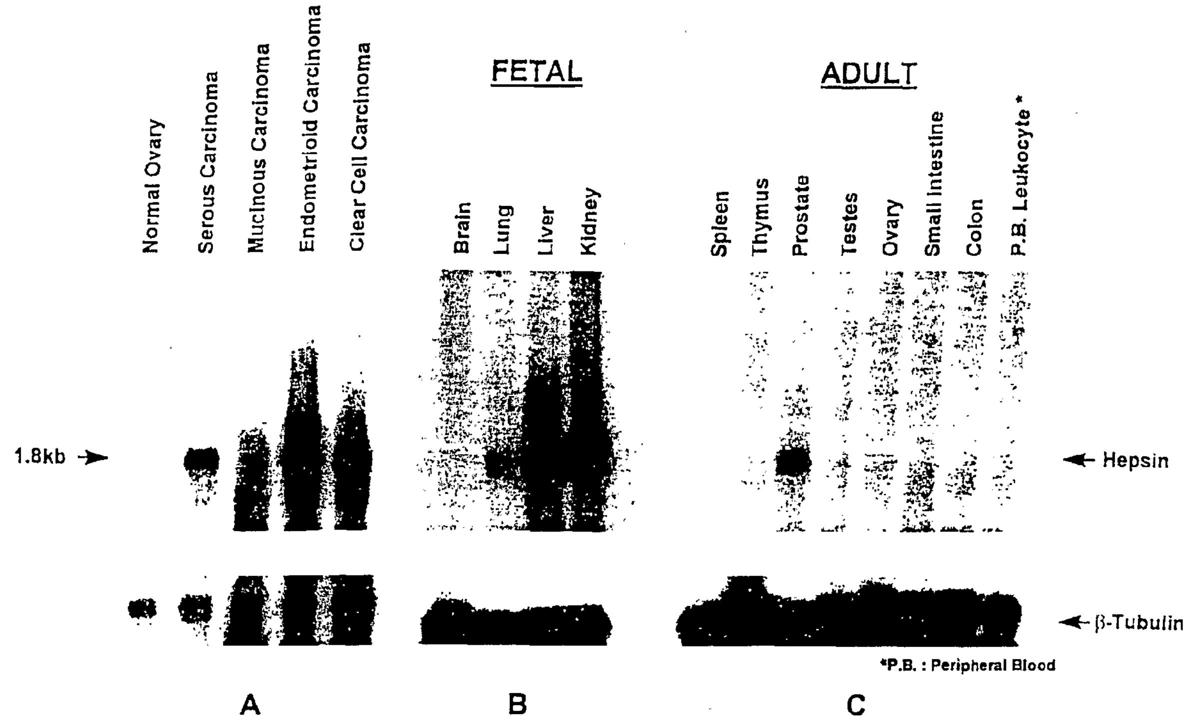
Normal Ovary
Serous Carcinoma
Mucinous Carcinoma
Endometrioid Carcinoma
Clear Cell Carcinoma
FETAL
Brain
Lung
Liver
Kidney
ADULT
Spleen
Thymus
Prostate
Testes
Ovary
Small Intestine
Colon
P.B. Leukocyte *
1.8kb
Hepsin
β-Tubulin
*P.B. : Peripheral Blood
A
B
C

Fig. 10

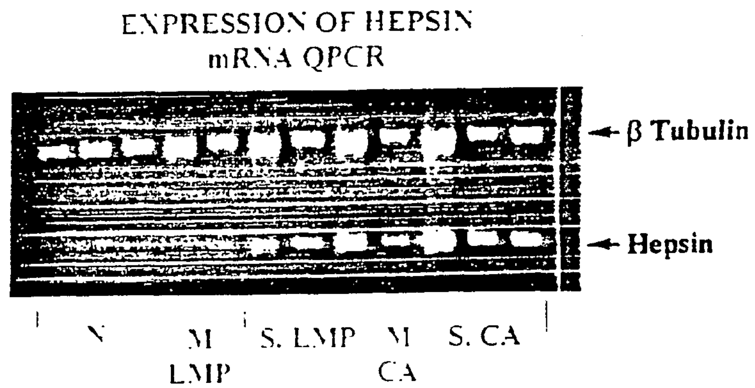

Fig. 11A

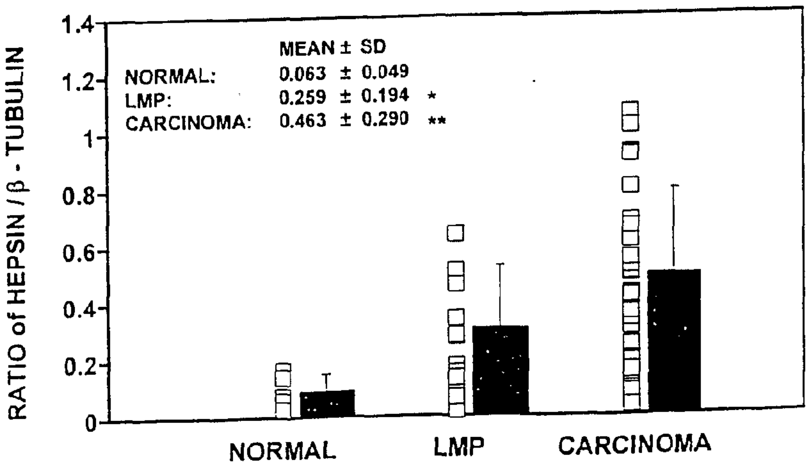

Fig. 11B

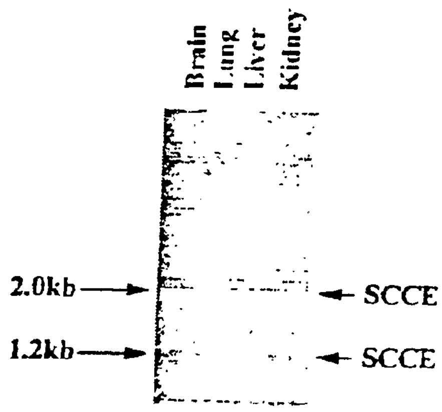

Fig. 12A

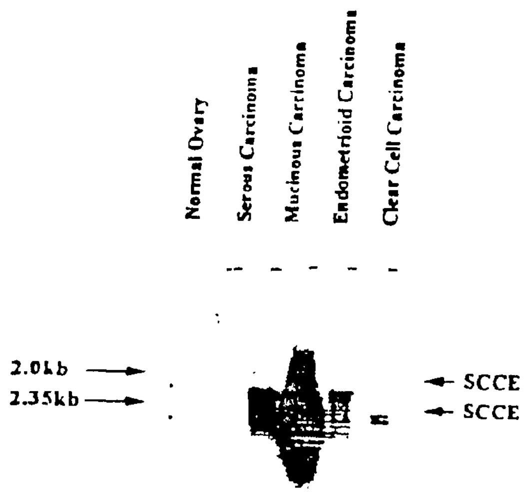

Fig. 12B

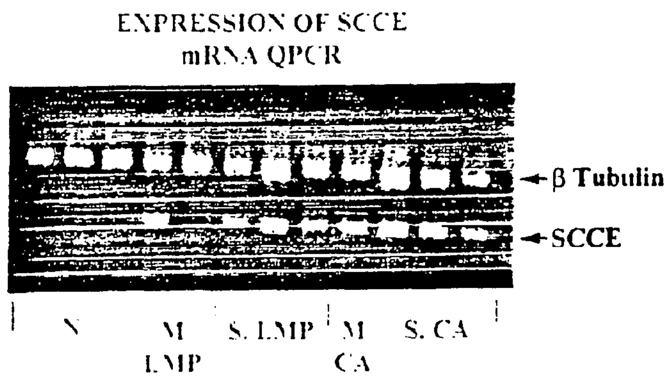

Fig. 13A

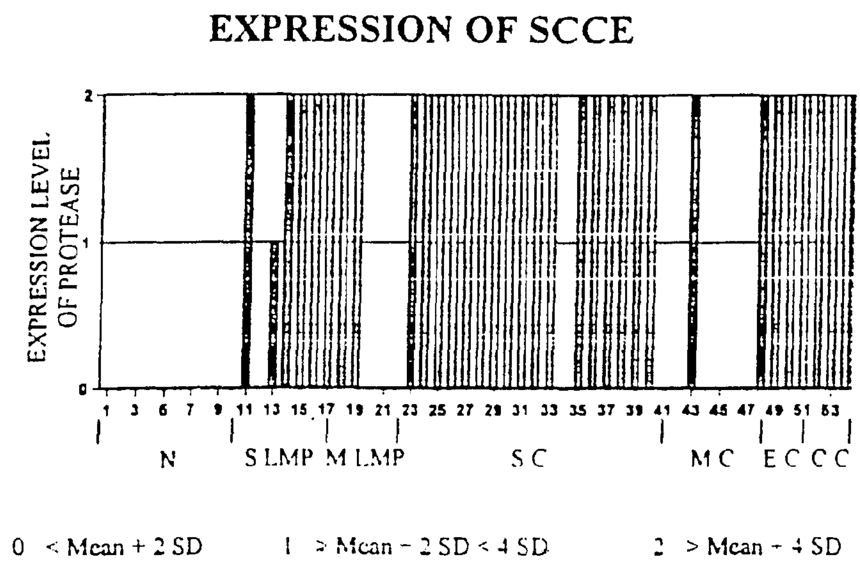

Fig. 13B

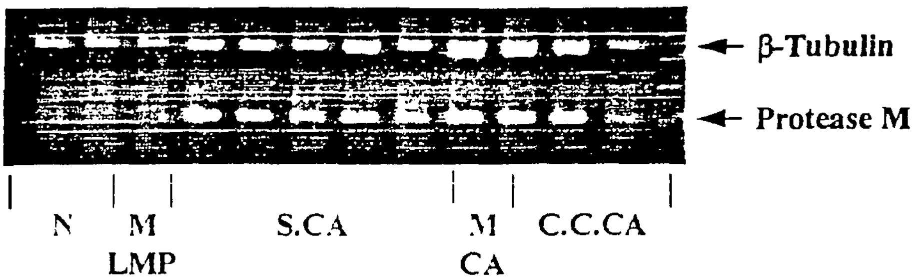

Fig. 14

1. .15

VVTAAHCVYDLYLPK

16 .30

SWTIQVGLVSLLDNP

↓ indicates the site of insert in TADG12

31 .45

H & D are the conserved regions of Serine protease.

APSHLVEKIVYHSKY

46 57

KPKRLGNDIALL

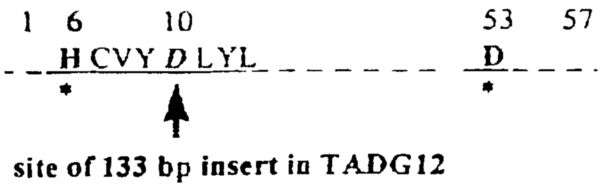

Fig. 15

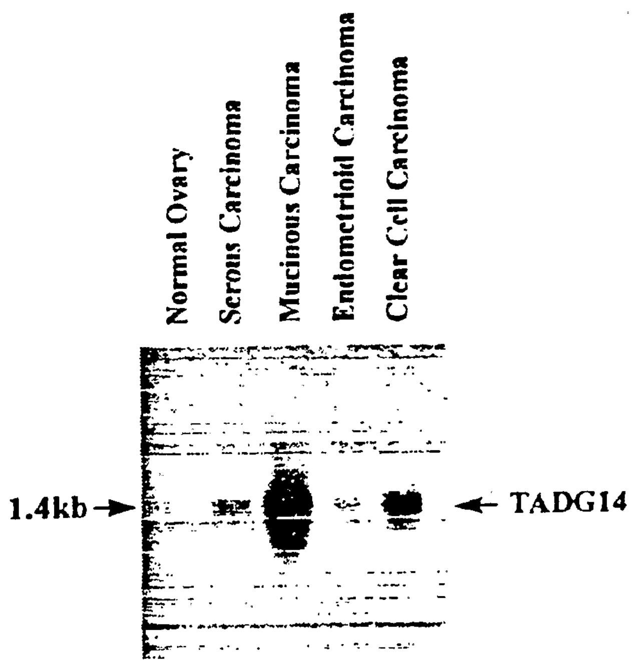

Fig. 16A

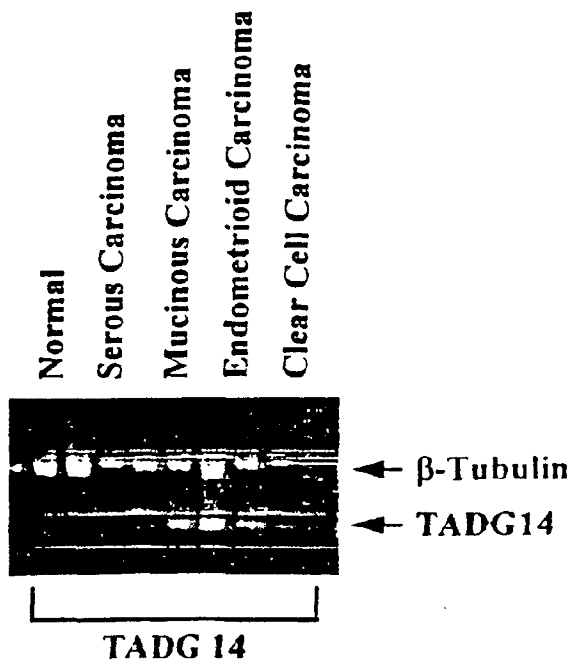

Fig. 16B

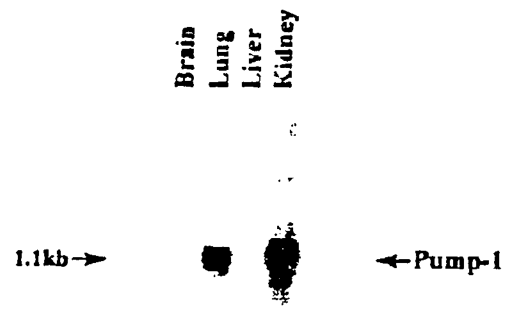

Fig. 17A

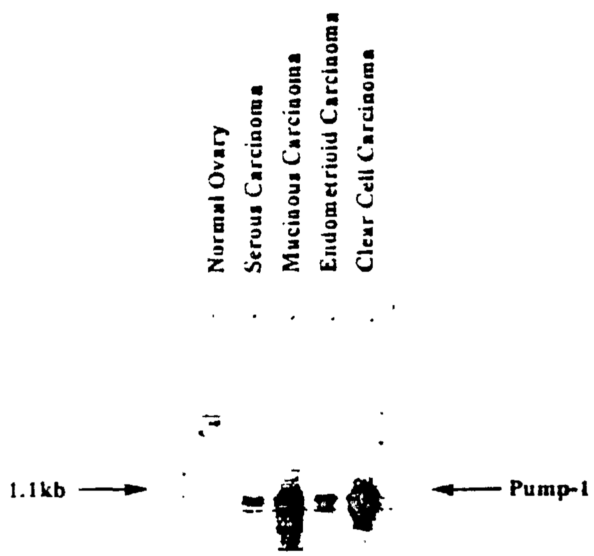

Fig. 17B

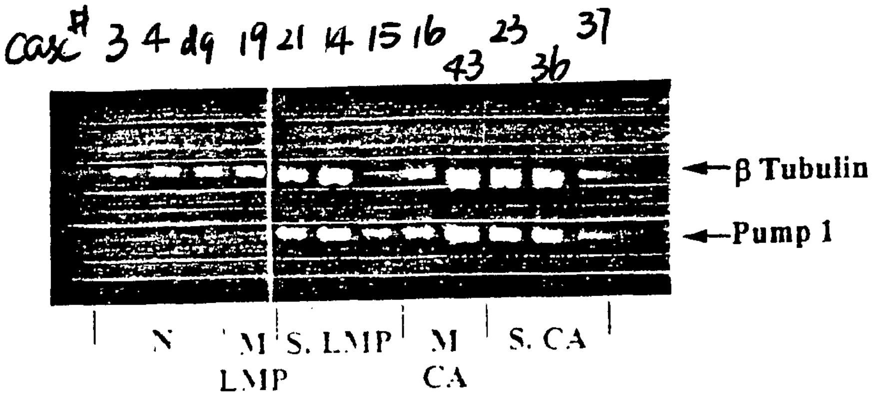

Fig. 18A

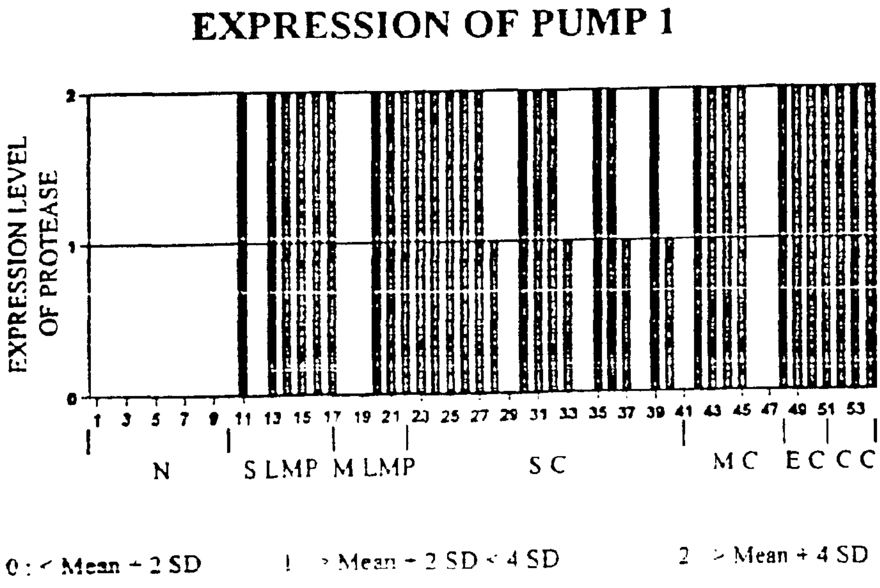

Fig. 18B

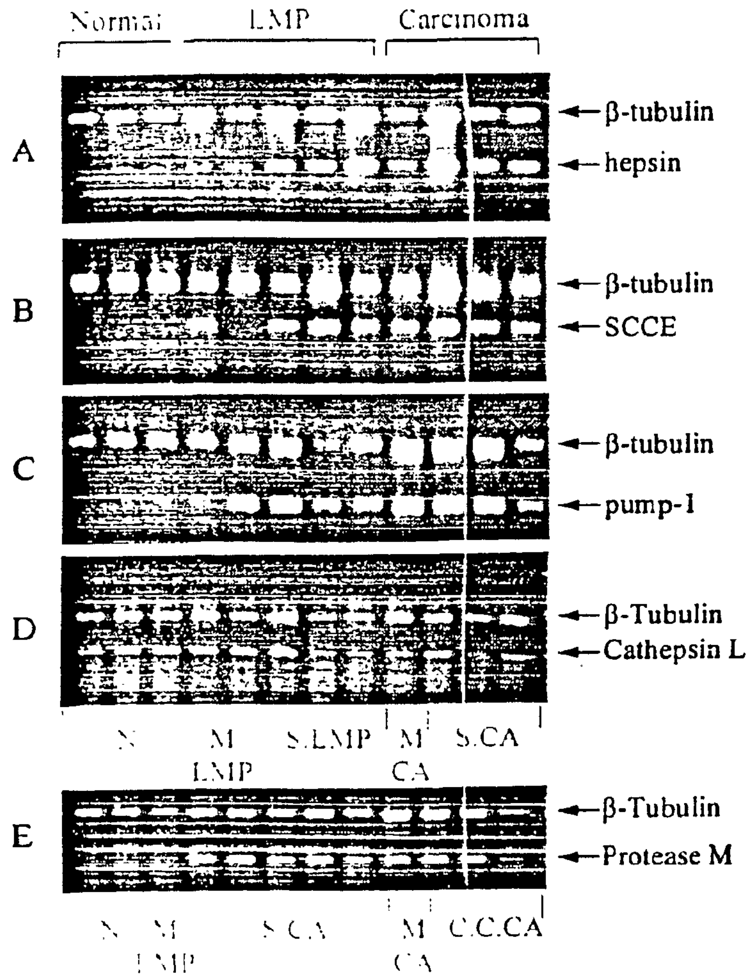
Normal
LMP
Carcinoma
A
B
C
D
E
β-tubulin
hepsin
β-tubulin
SCCE
β-tubulin
pump-1
β-Tubulin
Cathepsin L
N
M
LMP
S.LMP
M
CA
S.CA
β-Tubulin
Protease M
N
M
LMP
S.CA
M
CA
C.C.CA

Fig. 19

SEQUENCE LISTING

<110> O'Brien, Timothy J.

<120> Compositions and Methods for the Early Diagnosis
of Ovarian Cancer

<130> D6223CIPA/PCT

<141> 2001-02-20
<150> 09/510,738
<151> 2000-02-22

<160> 188

<210> 1
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<221> primer_bind
<222> 6, 9, 12, 15, 18
<223> sense oligonucleotide primer for amplifying serine
proteases, n = Inosine

<400> 1
tgggtngtna cngcngcnca ytg 23

<210> 2
<211> 20
<212> DNA
<213> Artificial sequence

<220>

<221> primer_bind
<222> 3, 6, 9, 12, 15, 18
<223> antisense oligonucleotide primer for amplifying serine
proteases, n = Inosine

<400> 2
arnarngcna tntcnttncc 20

<210> 3
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<221> primer_bind
<222> 3, 6, 9, 12, 18
<223> antisense oligonucleotide primer for amplifying serine
proteases, n = Inosine

<400> 3
arnggnccnc cnswrtcncc 20

<210> 4
<211> 24
<212> DNA
<213> Artificial sequence

<220>
<221> primer_bind
<222> 6, 15, 18
<223> sense oligonucleotide primer for amplifying cysteine proteases, n = Inosine

<400> 4
carggncart gyggnwsntg ytgg 24

<210> 5
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<221> primer_bind
<222> 3, 6, 15
<223> antisense oligonucleotide primer for amplifying cysteine proteases, n = Inosine

<400> 5
tanccnccrt trcanccytc 20

<210> 6
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<221> primer_bind
<222> 3, 6, 12, 15, 18
<223> sense oligonucleotide primer for amplifying metallo-proteases, n = Inosine

<400> 6
ccnmgntgyg gnrwnccnga 20

<210> 7
<211> 17
<212> DNA
<213> Artificial sequence

<220>

<221> primer_bind
<222> 6, 9, 11
<223> antisense oligonucleotide primer for amplifying metallo-
proteases, n = Inosine

<400> 7
ttrtgnccna nytcrtg 17

<210> 8
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for hepsin

<400> 8
tgtcccgatg gcgagtgttt 20

<210> 9
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for hepsin

<400> 9
cctgttggcc atagtactgc 20

<210> 10
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for SCCE

<400> 10
agatgaatga gtacaccgtg 20

<210> 11
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for SCCE

<400> 11
ccagtaagtc cttgtaaacc 20

<210> 12
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for CompB

<400> 12
aagggacacg agagctgtat 20

<210> 13
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for CompB

<400> 13
aagtggtagt tggaggaagc 20

<210> 14
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for Cath-L

<400> 14
attggagaga gaaaggctac 20

<210> 15
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for Cath-L

<400> 15
cttgggattg tacttacagg 20

<210> 16
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for PUMP-1

<400> 16
cttccaaagt ggtcacctac 20

<210> 17
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for PUMP-1

<400> 17
ctagactgct accatccgtc 20

<210> 18
<211> 17
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for β-tubulin

<400> 18
tgcattgaca acgaggc 17

<210> 19
<211> 17
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for β-tubulin

<400> 19
ctgtcttgac attgttg 17

<210> 20
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for Protease M

<400> 20
ctgtgatcca ccctgactat 20

<210> 21
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for Protease M

<400> 21
caggtggatg tatgcacact 20

<210> 22
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for TADG-12

<400> 22
gcgcactgtg tttatgagat 20

<210> 23
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for TADG-12

<400> 23
ctctttggct tgtacttgct 20

<210> 24
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for TADG-13

<400> 24
tgagggacat cattatgcac 20

<210> 25
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for TADG-13

<400> 25
caagttttcc ccataattgg 20

<210> 26
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> sense oligonucleotide primer specific for TADG-14

<400> 26
acagtacgcc tgggagacca 20

<210> 27
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> antisense oligonucleotide primer specific for TADG-14

<400> 27
ctgagacggt gcaattctgg 20

<210> 28
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 170-178 of the hepsin protein

<400> 28
Ser Leu Gly Arg Trp Pro Trp Gln Val
5

<210> 29
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 191-199 of the hepsin protein

<400> 29
Ser Leu Leu Ser Gly Asp Trp Val Leu
5

<210> 30

<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 229-237 of the hepsin protein

<400> 30
Gly Leu Gln Leu Gly Val Gln Ala Val
5

<210> 31
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 392-400 of the hepsin protein

<400> 31
Lys Val Ser Asp Phe Arg Glu Trp Ile
5

<210> 32
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 308-316 of the hepsin protein

<400> 32
Val Leu Gln Glu Ala Arg Val Pro Ile
5

<210> 33
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 130-138 of the hepsin protein

<400> 33
Arg Leu Leu Glu Val Ile Ser Val Cys
5

<210> 34
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 98-106 of the hepsin protein

<400> 34
Ala Leu Thr His Ser Glu Leu Asp Val
5

<210> 35
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 211-219 of the hepsin protein

<400> 35
Val Leu Ser Arg Trp Arg Val Phe Ala
5

<210> 36
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 26-34 of the hepsin protein

<400> 36
Leu Leu Leu Leu Thr Ala Ile Gly Ala
5

<210> 37
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 284-292 of the hepsin protein

<400> 37
Ala Leu Val Asp Gly Lys Ile Cys Thr
5

<210> 38
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 145-153 of the hepsin protein

<400> 38

Phe Leu Ala Ala Ile Cys Gln Asp Cys
5

<210> 39
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 192-200 of the hepsin protein

<400> 39
Leu Leu Ser Gly Asp Trp Val Leu Thr
5

<210> 40
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 20-28 of the hepsin protein

<400> 40
Ala Leu Thr Ala Gly Thr Leu Leu Leu
5

<210> 41
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 259-267 of the hepsin protein

<400> 41
Ala Leu Val His Leu Ser Ser Pro Leu
5

<210> 42
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 277-285 of the hepsin protein

<400> 42
Cys Leu Pro Ala Ala Gly Gln Ala Leu
5

<210> 43
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 230-238 of the hepsin protein

<400> 43
Leu Gln Leu Gly Val Gln Ala Val Val
5

<210> 44
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 268-276 of the hepsin protein

<400> 44
Pro Leu Thr Glu Tyr Ile Gln Pro Val
5

<210> 45
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 31-39 of the hepsin protein

<400> 45
Ala Ile Gly Ala Ala Ser Trp Ala Ile
5

<210> 46
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 285-293 of the hepsin protein

<400> 46
Leu Val Asp Gly Lys Ile Cys Thr Val
5

<210> 47
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 27-35 of the hepsin protein

<400> 47
Leu Leu Leu Thr Ala Ile Gly Ala Ala
5

<210> 48
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 191-199 of the hepsin protein

<400> 48
Ser Leu Leu Ser Gly Asp Trp Val Leu
5

<210> 49
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 163-171 of the hepsin protein

<400> 49
Ile Val Gly Gly Arg Asp Thr Ser Leu
5

<210> 50
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 392-400 of the hepsin protein

<400> 50
Lys Val Ser Asp Phe Arg Glu Trp Ile
5

<210> 51
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 64-72 of the hepsin protein

<400> 51
Met Val Phe Asp Lys Thr Glu Gly Thr
5

<210> 52
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 236-244 of the hepsin protein

<400> 52
Ala Val Val Tyr His Gly Gly Tyr Leu
5

<210> 53
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 55-63 of the hepsin protein

<400> 53
Gln Val Ser Ser Ala Asp Ala Arg Leu
5

<210> 54
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 130-138 of the hepsin protein

<400> 54
Arg Leu Leu Glu Val Ile Ser Val Cys
5

<210> 55
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 230-238 of the hepsin protein

<400> 55
Leu Gln Leu Gly Val Gln Ala Val Val
5

<210> 56
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 20-28 of the hepsin protein

<400> 56
Ala Leu Thr Ala Gly Thr Leu Leu Leu
5

<210> 57
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 259-267 of the hepsin protein

<400> 57
Ala Leu Val His Leu Ser Ser Pro Leu
5

<210> 58
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 277-285 of the hepsin protein

<400> 58
Cys Leu Pro Ala Ala Gly Gln Ala Leu
5

<210> 59
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 17-25 of the hepsin protein

<400> 59
Lys Val Ala Ala Leu Thr Ala Gly Thr
5

<210> 60
<211> 9
<212> PRT

<213> *Homo sapiens*

<220>
<223> Residues 285-293 of the hepsin protein

<400> 60
Leu Val Asp Gly Lys Ile Cys Thr Val
5

<210> 61
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 308-316 of the hepsin protein

<400> 61
Val Leu Gln Glu Ala Arg Val Pro Ile
5

<210> 62
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 27-35 of the hepsin protein

<400> 62
Leu Leu Leu Thr Ala Ile Gly Ala Ala
5

<210> 63
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 229-237 of the hepsin protein

<400> 63
Gly Leu Gln Leu Gly Val Gln Ala Val
5

<210> 64
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 313-321 of the hepsin protein

<400> 64
Arg Val Pro Ile Ile Ser Asn Asp Val
5

<210> 65
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 88-96 of the hepsin protein

<400> 65
Leu Ser Cys Glu Glu Met Gly Phe Leu
5

<210> 66
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 192-200 of the hepsin protein

<400> 66
Leu Leu Ser Gly Asp Trp Val Leu Thr
5

<210> 67
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 284-292 of the hepsin protein

<400> 67
Ala Leu Val Asp Gly Lys Ile Cys Thr
5

<210> 68
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 89-97 of the hepsin protein

<400> 68
Ser Cys Glu Glu Met Gly Phe Leu Arg
5

<210> 69
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 58-66 of the hepsin protein

<400> 69
Ser Ala Asp Ala Arg Leu Met Val Phe
5

<210> 70
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 393-401 of the hepsin protein

<400> 70
Val Ser Asp Phe Arg Glu Trp Ile Phe
5

<210> 71
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 407-415 of the hepsin protein

<400> 71
His Ser Glu Ala Ser Gly Met Val Thr
5

<210> 72
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 137-145 of the hepsin protein

<400> 72
Val Cys Asp Cys Pro Arg Gly Arg Phe
5

<210> 73
<211> 9
<212> PRT

<213> *Homo sapiens*

<220>
<223> Residues 269-277 of the hepsin protein

<400> 73
Leu Thr Glu Tyr Ile Gln Pro Val Cys
5

<210> 74
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 47-55 of the hepsin protein

<400> 74
Asp Gln Glu Pro Leu Tyr Pro Val Gln
5

<210> 75
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 119-127 of the hepsin protein

<400> 75
Cys Val Asp Glu Gly Arg Leu Pro His
5

<210> 76
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 68-76 of the hepsin protein

<400> 76
Lys Thr Glu Gly Thr Trp Arg Leu Leu
5

<210> 77
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 101-109 of the hepsin protein

<400> 77
His Ser Glu Leu Asp Val Arg Thr Ala
5

<210> 78
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 250-258 of the hepsin protein

<400> 78
Asn Ser Glu Glu Asn Ser Asn Asp Ile
5

<210> 79
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 293-301 of the hepsin protein

<400> 79
Val Thr Gly Trp Gly Asn Thr Gln Tyr
5

<210> 80
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 231-239 of the hepsin protein

<400> 80
Gln Leu Gly Val Gln Ala Val Val Tyr
5

<210> 81
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 103-111 of the hepsin protein

<400> 81
Glu Leu Asp Val Arg Thr Ala Gly Ala
5
<210> 82

<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 378-386 of the hepsin protein

<400> 82
Gly Thr Gly Cys Ala Leu Ala Gln Lys
5

<210> 83
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 358-366 of the hepsin protein

<400> 83
Val Cys Glu Asp Ser Ile Ser Arg Thr
5

<210> 84
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 264-272 of the hepsin protein

<400> 84
Ser Ser Pro Leu Pro Leu Thr Glu Tyr
5

<210> 85
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 87-95 of the hepsin protein

<400> 85
Gly Leu Ser Cys Glu Glu Met Gly Phe
5

<210> 86
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 272-280 of the hepsin protein

<400> 86
Tyr Ile Gln Pro Val Cys Leu Pro Ala
5

<210> 87
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 345-353 of the hepsin protein

<400> 87
Gly Ile Asp Ala Cys Gln Gly Asp Ser
5

<210> 88
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 301-309 of the hepsin protein

<400> 88
Tyr Tyr Gly Gln Gln Ala Gly Val Leu
5

<210> 89
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 238-246 of the hepsin protein

<400> 89
Val Tyr His Gly Gly Tyr Leu Pro Phe
5

<210> 90
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 204-212 of the hepsin protein

<400> 90

Cys Phe Pro Glu Arg Asn Arg Val Leu
5

<210> 91
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 117-125 of the hepsin protein

<400> 91
Phe Phe Cys Val Asp Glu Gly Arg Leu
5

<210> 92
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 124-132 of the hepsin protein

<400> 92
Arg Leu Pro His Thr Gln Arg Leu Leu
5

<210> 93
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 80-88 of the hepsin protein

<400> 93
Arg Ser Asn Ala Arg Val Ala Gly Leu
5

<210> 94
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 68-76 of the hepsin protein

<400> 94
Lys Thr Glu Gly Thr Trp Arg Leu Leu
5

<210> 95
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 340-348 of the hepsin protein

<400> 95
Gly Tyr Pro Glu Gly Gly Ile Asp Ala
5

<210> 96
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 242-250 of the hepsin protein

<400> 96
Gly Tyr Leu Pro Phe Arg Asp Pro Asn
5

<210> 97
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 51-59 of the hepsin protein

<400> 97
Leu Tyr Pro Val Gln Val Ser Ser Ala
5

<210> 98
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 259-267 of the hepsin protein

<400> 98
Ala Leu Val His Leu Ser Ser Pro Leu
5

<210> 99
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 277-285 of the hepsin protein

<400> 99
Cys Leu Pro Ala Ala Gly Gln Ala Leu
5

<210> 100
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 191-199 of the hepsin protein

<400> 100
Ser Leu Leu Ser Gly Asp Trp Val Leu
5

<210> 101
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 210-218 of the hepsin protein

<400> 101
Arg Val Leu Ser Arg Trp Arg Val Phe
5

<210> 102
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 222-230 of the hepsin protein

<400> 102
Val Ala Gln Ala Ser Pro His Gly Leu
5

<210> 103
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 236-244 of the hepsin protein

<400> 103
Ala Val Val Tyr His Gly Gly Tyr Leu
5

<210> 104
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 19-27 of the hepsin protein

<400> 104
Ala Ala Leu Thr Ala Gly Thr Leu Leu
5

<210> 105
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 36-44 of the hepsin protein

<400> 105
Ser Trp Ala Ile Val Ala Val Leu Leu
5

<210> 106
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 35-43 of the hepsin protein

<400> 106
Ala Ser Trp Ala Ile Val Ala Val Leu
5

<210> 107
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 300-308 of the hepsin protein

<400> 107
Gln Tyr Tyr Gly Gln Gln Ala Gly Val
5

<210> 108
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 363-371 of the hepsin protein

<400> 108
Ile Ser Arg Thr Pro Arg Trp Arg Leu
5

<210> 109
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 366-374 of the hepsin protein

<400> 109
Thr Pro Arg Trp Arg Leu Cys Gly Ile
5

<210> 110
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 236-244 of the hepsin protein

<400> 110
Ala Val Val Tyr His Gly Gly Tyr Leu
5

<210> 111
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 13-21 of the hepsin protein

<400> 111
Cys Ser Arg Pro Lys Val Ala Ala Leu
5

<210> 112
<211> 9
<212> PRT

<213> *Homo sapiens*

<220>
<223> Residues 179-187 of the hepsin protein

<400> 112
Ser Leu Arg Tyr Asp Gly Ala His Leu
5

<210> 113
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 43-51 of the hepsin protein

<400> 113
Leu Leu Arg Ser Asp Gln Glu Pro Leu
5

<210> 114
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 19-27 of the hepsin protein

<400> 114
Ala Ala Leu Thr Ala Gly Thr Leu Leu
5

<210> 115
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 55-63 of the hepsin protein

<400> 115
Gln Val Ser Ser Ala Asp Ala Arg Leu
5

<210> 116
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 163-171 of the hepsin protein

<400> 116
Ile Val Gly Gly Arg Asp Thr Ser Leu
5

<210> 117
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 140-148 of the hepsin protein

<400> 117
Cys Pro Arg Gly Arg Phe Leu Ala Ala
5

<210> 118
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 20-28 of the hepsin protein

<400> 118
Ala Leu Thr Ala Gly Thr Leu Leu Leu
5

<210> 119
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 409-417 of the hepsin protein

<400> 119
Glu Ala Ser Gly Met Val Thr Gln Leu
5

<210> 120
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 259-267 of the hepsin protein

<400> 120
Ala Leu Val His Leu Ser Ser Pro Leu

5

<210> 121
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 35-43 of the hepsin protein

<400> 121
Ala Ser Trp Ala Ile Val Ala Val Leu
5

<210> 122
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 184-192 of the hepsin protein

<400> 122
Gly Ala His Leu Cys Gly Gly Ser Leu
5

<210> 123
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 18-26 of the hepsin protein

<400> 123
Val Ala Ala Leu Thr Ala Gly Thr Leu
5

<210> 124
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 222-230 of the hepsin protein

<400> 124
Val Ala Gln Ala Ser Pro His Gly Leu
5

<210> 125

<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 224-232 of the hepsin protein

<400> 125
Gln Ala Ser Pro His Gly Leu Gln Leu
5

<210> 126
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 265-273 of the hepsin protein

<400> 126
Ser Pro Leu Pro Leu Thr Glu Tyr Ile
5

<210> 127
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 355-363 of the hepsin protein

<400> 127
Gly Pro Phe Val Cys Glu Asp Ser Ile
5

<210> 128
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 13-21 of the hepsin protein

<400> 128
Cys Ser Arg Pro Lys Val Ala Ala Leu
5

<210> 129
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 366-374 of the hepsin protein

<400> 129
Thr Pro Arg Trp Arg Leu Cys Gly Ile
5

<210> 130
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 140-148 of the hepsin protein

<400> 130
Cys Pro Arg Gly Arg Phe Leu Ala Ala
5

<210> 131
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 152-160 of the hepsin protein

<400> 131
Asp Cys Gly Arg Arg Lys Leu Pro Val
5

<210> 132
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 363-371 of the hepsin protein

<400> 132
Ile Ser Arg Thr Pro Arg Trp Arg Leu
5

<210> 133
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 133-141 of the hepsin protein

<400> 133

Ile Val Gly Gly Arg Asp Thr Ser Leu
5

<210> 134
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 331-339 of the hepsin protein

<400> 134
Gln Ile Lys Pro Lys Met Phe Cys Ala
5

<210> 135
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 80-88 of the hepsin protein

<400> 135
Arg Ser Asn Ala Arg Val Ala Gly Leu
5

<210> 136
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 179-187 of the hepsin protein

<400> 136
Ser Leu Arg Tyr Asp Gly Ala His Leu
5

<210> 137
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 43-51 of the hepsin protein

<400> 137
Leu Leu Arg Ser Asp Gln Glu Pro Leu
5

<210> 138
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 409-417 of the hepsin protein

<400> 138
Glu Ala Ser Gly Met Val Thr Gln Leu
5

<210> 139
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 311-319 of the hepsin protein

<400> 139
Glu Ala Arg Val Pro Ile Ile Ser Asn
5

<210> 140
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 222-230 of the hepsin protein

<400> 140
Val Ala Gln Ala Ser Pro His Gly Leu
5

<210> 141
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 19-27 of the hepsin protein

<400> 141
Ala Ala Leu Thr Ala Gly Thr Leu Leu
5

<210> 142
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 18-26 of the hepsin protein

<400> 142
Val Ala Ala Leu Thr Ala Gly Thr Leu
5

<210> 143
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 184-192 of the hepsin protein

<400> 143
Gly Ala His Leu Cys Gly Gly Ser Leu
5

<210> 144
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 224-232 of the hepsin protein

<400> 144
Gln Ala Ser Pro His Gly Leu Gln Leu
5

<210> 145
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 82-90 of the hepsin protein

<400> 145
Asn Ala Arg Val Ala Gly Leu Ser Cys
5

<210> 146
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 204-212 of the hepsin protein

<400> 146
Cys Phe Pro Glu Arg Asn Arg Val Leu
5

<210> 147
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 212-220 of the hepsin protein

<400> 147
Leu Ser Arg Trp Arg Val Phe Ala Gly
5

<210> 148
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 172-180 of the hepsin protein

<400> 148
Gly Arg Trp Pro Trp Gln Val Ser Leu
5

<210> 149
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 44-52 of the hepsin protein

<400> 149
Leu Arg Ser Asp Gln Glu Pro Leu Tyr
5

<210> 150
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 155-163 of the hepsin protein

<400> 150
Arg Arg Lys Leu Pro Val Asp Arg Ile
5

<210> 151
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 213-221 of the hepsin protein

<400> 151
Ser Arg Trp Arg Val Phe Ala Gly Ala
5

<210> 152
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 166-174 of the hepsin protein

<400> 152
Gly Arg Asp Thr Ser Leu Gly Arg Trp
5

<210> 153
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 369-377 of the hepsin protein

<400> 153
Trp Arg Leu Cys Gly Ile Val Ser Trp
5

<210> 154
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 180-188 of the hepsin protein

<400> 154
Leu Arg Tyr Asp Gly Ala His Leu Cys
5

<210> 155
<211> 9
<212> PRT

<213> *Homo sapiens*

<220>
<223> Residues 96-104 of the hepsin protein

<400> 155
Leu Arg Ala Leu Thr His Ser Glu Leu
5

<210> 156
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 396-404 of the hepsin protein

<400> 156
Phe Arg Glu Trp Ile Phe Gln Ala Ile
5

<210> 157
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 123-131 of the hepsin protein

<400> 157
Gly Arg Leu Pro His Thr Gln Arg Leu
5

<210> 158
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 207-215 of the hepsin protein

<400> 158
Glu Arg Asn Arg Val Leu Ser Arg Trp
5

<210> 159
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>

<223> Residues 209-217 of the hepsin protein

<400> 159
Asn Arg Val Leu Ser Arg Trp Arg Val
5

<210> 160
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 14-22 of the hepsin protein

<400> 160
Ser Arg Pro Lys Val Ala Ala Leu Thr
5

<210> 161
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 106-114 of the hepsin protein

<400> 161
Val Arg Thr Ala Gly Ala Asn Gly Thr
5

<210> 162
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 129-137 of the hepsin protein

<400> 162
Gln Arg Leu Leu Glu Val Ile Ser Val
5

<210> 163
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 349-357 of the hepsin protein

<400> 163
Cys Gln Gly Asp Ser Gly Gly Pro Phe

5

<210> 164
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 61-69 of the hepsin protein

<400> 164
Ala Arg Leu Met Val Phe Asp Lys Thr
5

<210> 165
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 215-223 of the hepsin protein

<400> 165
Trp Arg Val Phe Ala Gly Ala Val Ala
5

<210> 166
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 143-151 of the hepsin protein

<400> 166
Gly Arg Phe Leu Ala Ala Ile Cys Gln
5

<210> 167
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 246-254 of the hepsin protein

<400> 167
Phe Arg Asp Pro Asn Ser Glu Glu Asn
5

<210> 168

<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 132-140 of the hepsin protein

<400> 168
Leu Glu Val Ile Ser Val Cys Asp Cys
5

<210> 169
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 91-99 of the hepsin protein

<400> 169
Glu Glu Met Gly Phe Leu Arg Ala Leu
5

<210> 170
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 264-272 of the hepsin protein

<400> 170
Ser Ser Pro Leu Pro Leu Thr Glu Tyr
5

<210> 171
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 310-318 of the hepsin protein

<400> 171
Gln Glu Ala Arg Val Pro Ile Ile Ser
5

<210> 172
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>

<223> Residues 319-327 of the hepsin protein

<400> 172
Asn Asp Val Cys Asn Gly Ala Asp Phe
5

<210> 173
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 4-12 of the hepsin protein

<400> 173
Lys Glu Gly Gly Arg Thr Val Pro Cys
5

<210> 174
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 251-259 of the hepsin protein

<400> 174
Ser Glu Glu Asn Ser Asn Asp Ile Ala
5

<210> 175
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 256-264 of the hepsin protein

<400> 175
Asn Asp Ile Ala Leu Val His Leu Ser
5

<210> 176
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 294-302 of the hepsin protein

<400> 176
Thr Gly Trp Gly Asn Thr Gln Tyr Tyr
5

<210> 177
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 361-369 of the hepsin protein

<400> 177
Asp Ser Ile Ser Arg Thr Pro Arg Trp
5

<210> 178
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 235-243 of the hepsin protein

<400> 178
Gln Ala Val Val Tyr His Gly Gly Tyr
5

<210> 179
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 109-117 of the hepsin protein

<400> 179
Ala Gly Ala Asn Gly Thr Ser Gly Phe
5

<210> 180
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 270-278 of the hepsin protein

<400> 180
Thr Glu Tyr Ile Gln Pro Val Cys Leu
5

<210> 181
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 174-182 of the hepsin protein

<400> 181
Trp Pro Trp Gln Val Ser Leu Arg Tyr

<210> 182
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 293-301 of the hepsin protein

<400> 182
Val Thr Gly Trp Gly Asn Thr Gln Tyr
5

<210> 183
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 69-77 of the hepsin protein

<400> 183
Thr Glu Gly Thr Trp Arg Leu Leu Cys
5

<210> 184
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 90-98 of the hepsin protein

<400> 184
Cys Glu Glu Met Gly Phe Leu Arg Ala
5

<210> 185
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 252-260 of the hepsin protein

<400> 185
Glu Glu Asn Ser Asn Asp Ile Ala Leu
5

<210> 186
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 48-56 of the hepsin protein

<400> 186
Gln Glu Pro Leu Tyr Pro Val Gln Val
5

<210> 187
<211> 9
<212> PRT
<213> *Homo sapiens*

<220>
<223> Residues 102-110 of the hepsin protein

<400> 187
Ser Glu Leu Asp Val Arg Thr Ala Gly
5

<210> 188
<211> 1783
<212> DNA
<213> *Homo sapiens*

<220>
<223> full length cDNA of hepsin

<400> 188
tcgagcccgc tttccaggga ccctacctga gggcccacag gtgaggcagc 50
ctggcctagc aggccccacg ccaccgcctc tgcctccagg ccgcccgctg 100
ctgcggggcc accatgctcc tgcccaggcc tggagactga cccgaccccg 150
gcactacctc gaggctccgc cccacctgc tggaccccag ggtcccaccc 200
tggcccagga ggtcagccag ggaatcatta acaagaggca gtgacatggc 250
gcagaaggag ggtggccgga ctgtgccatg ctgctccaga cccaaggtgg 300
cagctctcac tgcggggacc ctgctacttc tgacagccat cggggcggca 350
tcctgggcca ttgtggctgt tctcctcagg agtgaccagg agccgctgta 400
cccagtgcag gtcagctctg cggacgctcg gctcatggtc tttgacaaga 450
cggaagggac gtggcggctg ctgtgctcct cgcgctccaa cgccagggta 500
gccggactca gctgcgagga gatgggcttc ctcagggcac tgacccactc 550
cgagctggac gtgcgaacgg cgggcgccaa tggcacgtcg ggcttcttct 600

```
gtgtggacga ggggaggctg ccccacaccc agaggctgct ggaggtcatc   650
tccgtgtgtg attgccccag aggccgtttc ttggccgcca tctgccaaga   700
ctgtggccgc aggaagctgc ccgtggaccg catcgtggga ggccgggaca   750
ccagcttggg ccggtggccg tggcaagtca gccttcgcta tgatggagca   800
cacctctgtg gggggatccct gctctccggg gactgggtgc tgacagccgc  850
ccactgcttc ccggagcgga accgggtcct gtcccgatgg cgagtgtttg   900
ccggtgccgt ggcccaggcc tctccccacg gtctgcagct gggggtgcag   950
gctgtggtct accacggggg ctatcttccc tttcgggacc ccaacagcga  1000
ggagaacagc aacgatattg ccctggtcca cctctccagt cccctgcccc  1050
tcacagaata catccagcct gtgtgcctcc cagctgccgg ccaggccctg  1100
gtggatggca agatctgtac cgtgacgggc tggggcaaca cgcagtacta  1150
tggccaacag gccggggtac tccaggaggc tcgagtcccc ataatcagca  1200
atgatgtctg caatggcgct gacttctatg gaaaccagat caagcccaag  1250
atgttctgtg ctggctaccc cgagggtggc attgatgcct gccagggcga  1300
cagcggtggt ccctttgtgt gtgaggacag catctctcgg acgccacgtt  1350
ggcggctgtg tggcattgtg agttggggca ctggctgtgc cctggcccag  1400
aagccaggcg tctacaccaa agtcagtgac ttccgggagt ggatcttcca  1450
ggccataaag actcactccg aagccagcgg catggtgacc cagctctgac  1500
cggtggcttc tcgctgcgca gcctccaggg cccgaggtga tcccggtggt  1550
gggatccacg ctgggccgag gatgggacgt tttcttctt gggcccggtc  1600
cacaggtcca aggacaccct ccctccaggg tcctctcttc cacagtggcg  1650
ggcccactca gccccgagac cacccaacct caccctcctg acccccatgt  1700
aaatattgtt ctgctgtctg ggactcctgt ctaggtgccc ctgatgatgg  1750
gatgctcttt aaataataaa gatggttttg att                    1783
```