20000475-7

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/574, C07K 16/30, 2/00 — A1

(11) International Publication Number: WO 99/41612

(43) International Publication Date: 19 August 1999 (19.08.99)

(21) International Application Number: PCT/GB99/00458

(22) International Filing Date: 15 February 1999 (15.02.99)

(30) Priority Data:
1. 9803138.8 — 13 February 1998 (13.02.98) — GB
2. 9803269.1 — 16 February 1998 (16.02.98) — GB

(71) Applicant *(for all designated States except US)*: OXFORD GLYCOSCIENCES (UK) LTD [GB/GB]; 10 The Quadrant, Barton Lane, Abingdon Science Park, Abingdon, Oxon OX14 3YS (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: PAREKH, Rajesh, Bhikhu [GB/GB]; Oxford GlycoSciences (UK) Ltd., 10 The Quadrant, Barton Lane, Abingdon Science Park, Abingdon, Oxon OX14 3YS (GB). PATEL, Thakorbhai, Parshotambhai [GB/GB]; Oxford GlycoSciences (UK) Ltd., 10 The Quadrant, Barton Lane, Abingdon Science Park, Abingdon, Oxon OX14 3YS (GB). TOWNSEND, Robert, Reid [US/GB]; Oxford GlycoSciences (UK) Ltd., 10 The Quadrant, Barton Lane, Abingdon Science Park, Abingdon, Oxon OX14 3YS (GB). MOYSES, Christopher [GB/GB]; Oxford GlycoSciences (UK) Ltd., 10 The Quadrant, Barton Lane, Abingdon Science Park, Abingdon, Oxon OX14 3YS (GB). JOHNSON, Philip, J. [GB/CN]; The Chinese University of Hong Kong, Dept. of Clinical Oncology, Prince of Wales Hospital, Shatin, N.T. (CN).

(74) Agents: CHAPMAN, Paul, William et al.; Kilburn & Strode, 20 Red Lion Street, London WC1R 4PJ (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, 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, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND COMPOSITIONS FOR DIAGNOSIS OF HEPATOMA

(57) Abstract

The present invention provides methods and compositions for screening, diagnosis and prognosis of hepatoma, for monitoring the effectiveness of hepatoma treatment, and for drug development. Hepatoma–Diagnostic Features (HFs), detectable by two–dimensional electrophoresis of serum or plasma are described. The invention further provides Hepatoma–Diagnostic Protein Isoforms (HPIs) detectable in serum or plasma, preparations comprising isolated HPIs, antibodies immunospecific for HPIs, and kits comprising the aforesaid.

2 PDs.
1 applⁿ. (changed addr, see PF23 of 6/10/00)
5 Invs.
IPC = G01N 33/574
IPER = 13/9/99

METHODS AND COMPOSITIONS FOR DIAGNOSIS OF HEPATOMA

1. INTRODUCTION

The present invention relates to the identification of proteins and protein isoforms that are associated with hepatocellular carcinoma (hepatoma) and to their use for screening, diagnosis, prognosis, therapy and drug development.

2. BACKGROUND OF THE INVENTION

Hepatoma is an increasingly prevalent cancer, affecting several hundred thousand patients worldwide. Methods for its treatment are still relatively poor, and the main hope for more effective therapy lies in earlier and more accurate diagnosis.

Measurement of serum alphafetoprotein (AFP) is widely used in the diagnosis of hepatoma. While AFP is a useful marker when levels are markedly elevated, poor specificity and poor positive predictive value (PPV) for hepatoma at lower levels severely limit its practical application. Johnson et al., 1997, Br. J. Cancer 75: 236-240. Certain lectins show differential binding to AFP from hepatoma serum compared with AFP in benign conditions. Aoyagi et al., 1985, Biochim. Biophys. Acta 830: 217-223; Aoyagi et al., 1993, Br. J. Cancer 67: 486-492. However, these approaches have not found broad clinical acceptance. Johnson et al., *supra*. More recently, isoelectric focusing has been used to detect isoforms of AFP that appear relatively specific for hepatoma. Johnson et al., *supra*. These reports illustrate an ongoing need for identification of markers in serum for diagnosis of hepatoma.

3. SUMMARY OF THE INVENTION

The present invention provides methods and compositions for screening, diagnosis and prognosis of hepatoma, for monitoring the effectiveness of hepatoma treatment, and for drug development.

A first aspect of the invention provides methods for diagnosis of hepatoma that comprise analyzing a sample of plasma or serum by two-dimensional electrophoresis to detect the level of at least one Hepatoma-Diagnostic Feature (HF), *e.g.*, an HF selected from the group of HFs disclosed herein. These methods are also suitable for screening, prognosis, monitoring the results of therapy, and drug development.

A second aspect of the invention provides methods for diagnosis of hepatoma that comprise detecting in a sample of plasma or serum the level of at least one Hepatoma-Diagnostic Protein Isoform (HPI), *e.g.*, an HPI selected from the group of HPIs disclosed herein. These methods are also suitable for screening, prognosis, monitoring the results of therapy, and drug development.

A third aspect of the invention provides monoclonal and polyclonal antibodies capable of immunospecific binding to an HPI, *e.g.*, an HPI disclosed herein.

A fourth aspect of the invention provides a preparation comprising an isolated HPI, *i.e.* an HPI free from proteins or protein isoforms having a significantly different isoelectric point or a significantly different apparent molecular weight from the HPI.

4. BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is an image obtained from 2-dimensional electrophoresis of normal human serum, which has been annotated to identify 14 landmark features, designated PL1 to PL12 and PL15 to PL16.

5. DETAILED DESCRIPTION OF THE INVENTION

The invention described in detail below encompasses methods and compositions for screening, diagnosis and prognosis of hepatoma in a mammalian subject, methods for monitoring the results of hepatoma therapy, and methods for drug development. Preferably, the subject is human, more preferably a human adult.

For clarity of disclosure, and not by way of limitation, the invention will be described with respect to the analysis of serum samples. However, as one skilled in the art will appreciate, the assays and techniques described herein can be applied to other types of patient samples, including a body fluid (*e.g.* plasma, urine, bile, ascitic fluid), a tissue sample suspected of containing material derived from a hepatoma (*e.g.* a biopsy such as a liver biopsy or a biopsy of a suspicious mass) or homogenate thereof.

5.1. Hepatoma-Diagnostic Features (HFs)

In one aspect of the invention, two-dimensional electrophoresis is used to analyze serum from a subject in order to measure the abundance of one or more Hepatoma-Diagnostic Features (HFs) for screening or diagnosis of hepatoma, to determine the prognosis of a hepatoma patient, to monitor the effectiveness of hepatoma therapy, or for drug development. As used herein, "two-dimensional electrophoresis" (2D-electrophoresis) means a technique comprising isoelectric focusing, followed by denaturing electrophoresis; this generates a two-dimensional gel (2D-gel) containing a plurality of separated proteins. Preferably, the step of denaturing electrophoresis uses polyacrylamide electrophoresis in the presence of sodium dodecyl sulfate (SDS-PAGE). Especially preferred are the highly accurate and automatable methods and apparatus ("the Preferred Technology") described in U.S. Application No. 08/980,574 and WO 98/23950 which are incorporated herein by reference in its entirety. Briefly, the Preferred Technology provides efficient, computer-assisted methods and apparatus for identifying, selecting and characterizing biomolecules in a biological sample. A two-dimensional array is generated by separating biomolecules in a two-dimensional gel according to their electrophoretic

mobility and isoelectric point. A computer-generated digital profile of the array is generated, representing the identity, apparent molecular weight, isoelectric point, and relative abundance of a plurality of biomolecules detected in the two-dimensional array, thereby permitting computer-mediated comparison of profiles from multiple biological samples, as well as computer aided excision of separated proteins of interest.

As used herein, the term "Hepatoma-Diagnostic Feature" (HF) refers to a feature (e.g. a spot in a 2D gel), detectable by 2D electrophoresis of a biological sample, that is differentially present in one sample compared to another, relevant sample, *e.g.*, in serum from a subject with hepatoma compared with serum from a subject without hepatoma. As used herein, a feature (or a protein isoform) is "differentially present" in a first sample with respect to a second sample when a method for detecting the feature or isoform (e.g. 2D electrophoresis or an immunoassay) reveals that the feature (or protein isoform) is present at a different relative abundance in the first sample as compared with the second sample. If the measured feature in the first sample is at a higher abundance than in the second sample, the feature or isoform is "increased" in the first sample with respect to the second; conversely, if the measured feature in the first sample is at a lower abundance than in the second sample, the feature or isoform is "decreased" in the first sample with respect to the second.

Preferably, the relative abundance of a feature in two samples is determined in two steps. First, the signal obtained upon detecting the feature in a sample is normalized by reference to a suitable background parameter, *e.g.*, to the total protein in the sample being analyzed (*e.g.*, total protein loaded onto a gel), to an invariant feature, *i.e.*, a feature whose abundance is known to be similar in the samples being compared, *e.g.*, one or more Expression Reference Features (ERFs), such as the ERFs disclosed below, or to the total signal detected from all proteins in the sample.

Secondly, the normalized signal for the feature in one sample or sample set is compared with the normalized signal for the same feature in another sample or sample set in order to identify features that are "differentially present" in the first sample (or sample set) with respect to the second.

Two groups of HFs have been identified through the methods and apparatus of the Preferred Technology. The first group consists of HFs that are decreased in the serum of subjects with hepatoma as compared with the serum of subjects without hepatoma (*e.g.* subjects with cirrhosis). These HFs can be described by apparent molecular weight (MW) and isoelectric point (pI) as follows:

Table I. HFs Decreased In Hepatoma Serum

Name	MCI	Fold decrease	pI	MW (kd)	p
HF-1	7537	11.7	8.12	44,093	0.0003
HF-2	7341	4.9	6.02	62,370	0.0239
HF-3	7232	4.6	7.60	103,154	0.0094
HF-4	7534	3.1	7.41	44,764	0.0278
HF-5	7763	2.7	6.93	67,555	0.0362
HF-6	7945	2.7	5.73	57,295	0.0377
HF-7	7760	2.5	6.37	69,194	0.0109
HF-8	7179	2.5	5.82	166,395	0.0138
HF-9	8031	2.1	6.15	24,467	0.0258
HF-10	7831	2.0	7.39	37,863	0.0025
HF-11	7910	2.0	7.79	93,302	0.1078
HF-12	7562	2.0	5.37	40,672	0.0166
HF-13	7762	2.0	6.41	67,826	0.1015
HF-14	7641	2.0	6.80	28,169	0.0364

The second group consists of HFs that are increased in the serum of subjects with hepatoma as compared with the serum of subjects without hepatoma (*e.g.* subjects with cirrhosis). These HFs can be described by apparent molecular weight (MW) and isoelectric point (pI) as follows:

Table II. HFs Increased In Hepatoma Serum

Name	MCI	Fold increase	pI	MW (kd)	p
HF-15	7308	3.9	5.55	67,555	0.0013
HF-16	7463	3.8	7.84	50,625	0.0055
HF-17	7629	3.1	7.27	31,244	0.0000
HF-18	7366	3.0	4.60	59,928	0.0000
HF-19	7759	3.0	5.40	69,194	0.0021
HF-20	7581	3.0	4.94	38,465	0.0000
HF-21	7320	2.9	7.63	66,485	0.0009
HF-22	7690	2.9	5.72	16,707	0.0136
HF-23	7241	2.8	5.27	96,988	0.0000
HF-24	7242	2.7	5.31	96,988	0.0000
HF-25	7240	2.6	5.14	97,732	0.0000
HF-26	7734	2.6	6.04	108,311	0.0011

HF-27	7469	2.6	8.12	50,134	0.0344
HF-28	7431	2.6	5.34	52,432	0.0009
HF-29	7553	2.5	4.84	41,288	0.0000
HF-30	7457	2.5	7.61	50,822	0.0255
HF-31	7330	2.4	7.44	63,882	0.0003
HF-32	7460	2.4	7.25	50,625	0.0085
HF-33	7237	2.4	5.10	97,732	0.0000
HF-34	7243	2.4	5.21	96,988	0.0000
HF-35	7368	2.3	4.57	60168	0.0168
HF-36	7730	2.3	5.10	136,409	0.0007
HF-37	7288	2.3	6.79	71,726	0.0035
HF-38	7568	2.3	5.34	39,869	0.0034
HF-39	7238	2.3	5.04	97,732	0.0000
HF-40	7691	2.3	5.40	16,551	0.0049
HF-41	7317	2.2	6.06	66,220	0.0974
HF-42	7465	2.2	7.44	50,329	0.0097
HF-43	7380	2.2	6.01	58,044	0.0144
HF-44	7239	2.2	5.07	97,732	0.0000
HF-45	7692	2.2	6.10	16,447	0.0412
HF-46	7379	2.2	4.66	57,582	0.0040
HF-47	7180	2.2	5.36	164,341	0.0059
HF-48	7633	2.2	7.26	30,638	0.0016
HF-49	7392	2.1	5.10	56,444	0.0014
HF-50	7572	2.1	5.71	39,239	0.0227
HF-51	7190	2.0	5.26	161,308	0.0001
HF-52	7186	2.0	5.32	163,324	0.0008
HF-53	7556	2.0	5.15	40,753	0.0020
HF-54	7540	2.0	5.29	43,541	0.0004
HF-55	7185	2.0	5.22	163,324	0.0000
HF-56	7822	2.0	5.12	42,416	0.0030

For any given HF, the ratio obtained upon comparing the normalized signal observed upon analyzing serum from subjects with hepatoma relative to the normalized signal obtained upon analyzing serum from subjects without hepatoma will depend upon the particular analytical protocol and detection technique that is used. Accordingly, the present invention contemplates that each laboratory will establish a reference range for each HF in hepatoma-free subjects according to the analytical protocol and detection technique in use, as is conventional in the diagnostic art. Preferably, at least one positive control serum sample from a subject known to have hepatoma or at least one negative control serum sample from a subject known to be free of hepatoma (and more preferably at least one positive and at least one negative control sample) is included in each batch of test samples analyzed.

As the skilled artisan will readily appreciate, the measured MW and pI of a given feature or protein isoform will vary to some extent depending on the precise protocol used for each step of the 2D electrophoresis and for landmark matching. As used herein, the terms "MW" and "pI" are defined, respectively, to mean the apparent molecular weight and the isoelectric point of a feature or protein isoform as measured in exact accordance with the experimental protocol set forth in Section 6 below ("the Reference Protocol"). When the Reference Protocol is followed and when samples are run in duplicate or a higher number of replicates, variation in the measured mean pI of an HF or HPI is typically less than $\pm$ 1% and variation in the measured mean MW of an HF or HPI is typically less than $\pm$ 5%. Where the skilled artisan wishes to deviate from the Reference Protocol, calibration experiments should be performed to compare the MW and pI for each HF or protein isoform as detected (a) by the Reference Protocol and (b) by the deviant protocol.

HFs can be used for detection, prognosis, diagnosis, or monitoring of hepatoma or for drug development. In one embodiment of the invention, serum from a subject is analyzed by 2D electrophoresis for quantitative detection of one or more HFs selected from the group consisting of HF-1 to HF-14 and HF-57 to HF-59, wherein a decreased abundance of an HF in serum from the subject relative to serum from a subject (or subjects) without hepatoma (*e.g.*, a control sample or a previously determined reference range) indicates the presence of hepatoma; preferably, the one or more HFs are selected from the group consisting of HF-1 and HF-59. In another embodiment of the invention, serum from a subject is analyzed by 2D electrophoresis for quantitative detection of one or more HFs selected from the group consisting of HF-15 to HF-56 and HF-60 to HF-61, wherein an increased abundance of an HF in serum from the subject relative to serum from a subject (or subjects) without hepatoma (*e.g.*, a control sample or a previously determined reference range) indicates the presence of hepatoma; preferably, the one or more HFs are selected from the group consisting of HF-17, HF-18, HF-20, HF-21, HF-23, HF-24, HF-25, HF-28, HF-29, HF-31, HF-33, HF-34, HF-36, HF-39, HF-44, HF-51, HF-52, HF-54 and HF-55; still more preferably, the one or more HFs are selected from the group consisting of HF-17, HF-18, HF-20, HF-23, HF-24, HF-25, HF-29, HF-33, HF-34, HF-39, HF-44 and HF-55.

5.2. Hepatoma-Diagnostic Protein Isoforms (HPIs)

In another aspect of the invention, serum from a subject is analyzed for quantitative detection of one or more Hepatoma-Diagnostic Protein Isoforms (HPIs) for screening or diagnosis of hepatoma, to determine the prognosis of a hepatoma patient, to monitor the effectiveness of hepatoma therapy, or for drug development. As used herein, the term "Hepatoma-Diagnostic Protein Isoform" refers to a protein isoform that is differentially present in serum subjects with hepatoma compared with serum from subjects without hepatoma. As is well known in the art, the protein product of a single gene may be expressed as variants (isoforms) (a) that differ as a result of differential post-translational modification (e.g. glycosylation, phosphorylation or acylation), so that proteins of identical amino acid sequence can differ in their pI, MW or both, and/or (b) that differ in their amino acid composition (e.g. as a result of alternative mRNA splicing or limited proteolysis). It follows that differential presence of a protein isoform does not require differential expression of the gene encoding the protein in question.

Two groups of HPIs have been identified by partial amino acid sequencing of HFs, using the methods and apparatus of the Preferred Technology. The first group consists of HPIs that are decreased in the serum of subjects with hepatoma as compared with the serum of subjects without hepatoma, where the differential presence is very highly significant ($p \leq 0.001$). The MW, pI and partial amino acid sequence of these HPIs are presented in Table III, as follows:

Table III. HPI Decreased In Hepatoma Serum

HF #	HPI	Known homologous protein	Partial amino acid sequence	pI	MW (kd)
1	HPI-1	Immunoglobulin G - H_1, H_2, H_3	REPQVYTLPPSR	8.12	44,093

The second group comprises HPIs that are increased in the serum of subjects with hepatoma as compared with the serum of subjects without hepatoma, where the differential presence is very highly significant ($p \leq 0.001$). The MWs, pIs and partial amino acid sequences of these HPIs are presented in Table IV, as follows:

Table IV. HPIs Increased In Hepatoma Serum

HF #	HPI	Known homologous protein	Partial amino acid sequence	pI	MW (kd)
17	HPI-2	complement factor 4	RGLQDEDGYR	7.27	31,244
18	HPI-3	α_1-anti-chymotrypsin	KITLLSALVETR	4.60	59,928
20	HPI-4	complement factor 3	KGYTQQLAFR	4.94	38,465
21	HPI-5	complement factor 3	RIPIEDGSGEVVLSR	7.63	66,485
23	HPI-6	complement factor 4	RTYNVLDMK	5.27	96,988
25	HPI-8	complement factor 4	KAEMADQASAWLTR	5.14	97,732
28	HPI-9	vitamin D-binding protein	KHLSLLTTLSNR	5.34	52,432
29	HPI-10	haptoglobin-1	RVGYVSGWGR	4.84	41,288
29	HPI-11	haptoglobin-2	KYVMLPVADQDQCIR	4.84	41,288
31	HPI-12	complement factor 3	KTIYTPGSTVLYR	7.44	63,882

31	HPI-13	NOVEL	AETTDF	7.44	63,882
33	HPI-14	complement factor 4	RQGSFQGGFR	5.10	97,732
34	HPI-15	complement factor 4	RQGSFQGGFR	5.21	96,988
36	HPI-16	inter-α-trypsin inhibitor family heavy chain related protein	RFAHTVVTSR	5.10	136,409
39	HPI-17	complement factor 4	RQGSFQGGFR	5.04	97,732
44	HPI-18	complement factor 4	RTYNVLDMK	5.07	97,732
51	HPI-19	Ceruloplasmin	KGAYPLSIEPIGVR	5.26	161,308
52	HPI-20	Ceruloplasmin	KALYLQYTDETFR	5.32	163,324
54	HPI-21	NOVEL	[I/L]W[I/L]GTTR	5.29	43,541
55	HPI-22	Ceruloplasmin	RQSEDSTFYLGER	5.22	163,324

In one embodiment of the invention, serum from a subject is analyzed for quantitative detection of one or more HPIs selected from the group consisting of HPI-1 and HPI-23, wherein a decreased level of the one or more HPIs in serum from the subject relative to serum from a subject or subjects without hepatoma (e.g. a control sample or a previously determined reference range) indicates the presence of hepatoma. In another embodiment of the invention, serum from a subject is analyzed by 2D electrophoresis for quantitative detection of one or more HPIs selected from the group consisting of HPI-2 to HPI-22 and HPI-24, wherein an increased level of the one or more HPIs in serum from the subject relative to serum from a subject or subjects without hepatoma (e.g. a control sample or a previously determined reference range) indicates the presence of hepatoma.

As shown above, the HPIs described herein include previously unknown proteins as well as isoforms of known proteins where the isoforms were not previously known to be associated with hepatoma. For each HPI, the present invention additionally provides a preparation comprising the isolated HPI or fragments thereof, and further provides antibodies that bind to said HPI, to said fragments, or to both said HPI and said fragments. As used herein, an "isolated" HPI is an HPI free of proteins or protein isoforms having a significantly different pI or MW from those of the HPI, as determined by 2D electrophoresis. As used herein, a "significantly different" pI or MW is one that causes the contaminating protein isoform to be resolved from the HPI on 2D electrophoresis, performed according to the Reference Protocol.

In one embodiment, an isolated protein is provided, said protein comprising a peptide with the amino acid sequence identified in Table III or IV for an HPI, said protein having a pI and MW within 10% (preferably within 5%, more preferably within 1%) of the values identified in Tables III and IV for that HPI.

The HPIs of the invention can be assayed by any method known to those skilled in the art. In one embodiment, the HPIs are separated on a 2-D gel by virtue of their MWs and pIs and visualized by staining the gel.

Alternatively, HPIs can be detected in assays, such as immunoassays, for detection, prognosis, diagnosis or monitoring of hepatoma or for drug development. In one embodiment, an immunoassay is performed by contacting a sample, derived from a subject to be tested, with an anti-HPI antibody under conditions such that immunospecific binding can occur, and detecting or measuring the amount of any immunospecific binding by the antibody. Preferably, the anti-HPI antibody preferentially binds to the HPI rather than to other isoforms of the same protein. In a preferred embodiment, the anti-HPI antibody binds to the HPI with at least 2-fold greater affinity, more preferably at least 5-fold greater affinity, still more preferably at least 10-fold greater affinity, than to other isoforms of the same protein.

In one embodiment, binding of antibody in tissue sections can be used to detect aberrant HPI localization or aberrant (*e.g.*, high, low, absent) levels of an HPI. In a specific embodiment, antibody to an HPI can be used to assay a patient tissue (e.g. a liver biopsy) or serum sample for the presence of the HPI where an aberrant level of HPI is indicative of hepatoma. As used herein, an "aberrant level" means an increased or decreased level relative to that present, or a standard level representing that present, in an analogous sample from a portion of the body or from a subject not having hepatoma.

The immunoassays which can be used include without limitation competitive and non-competitive assay systems using techniques such as western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, to name but a few.

If desired, an HPI can be detected by means of a two-step sandwich assay. Where an HPI represents a particular glycoform of a protein, the first step can employ an anti-HPI antibody (which can optionally be immobilized on a solid phase) to capture the HPI; in the second step, a directly or indirectly labelled lectin can be used to detect the captured HPI. Any lectin can be used for this purpose that preferentially binds

to the HPI rather than (a) to other glycoforms that have the same core protein as the HPI or (b) to other isoforms that share the antigenic determinant recognized by the antibody. In a preferred embodiment, the chosen lectin binds to the HPI with at least 2-fold greater affinity, more preferably at least 5-fold greater affinity, still more preferably at least 10-fold greater affinity, than to said other glycoforms that have the same core protein as the HPI or to said other isoforms that share the same antigenic determinant recognized by the antibody. A lectin that is suitable for detecting a given HPI can readily be identified by methods well known in the art, for instance upon testing one or more lectins enumerated in Table I on pages 158-159 of Sumar et al., Lectins as Indicators of Disease-Associated Glycoforms, *In*: Gabius H-J & Gabius S (eds.), 1993, Lectins and Glycobiology, at pp. 158-174 (which is incorporated herein by reference in its entirety).

If desired, a gene encoding an HPI, a related gene and related nucleic acid sequences and subsequences, including complementary sequences, can also be used in hybridization assays. A nucleotide encoding an HPI, or subsequences thereof comprising about at least 8 nucleotides (or the complement of the foregoing) can be used as hybridization probes. Hybridization assays can be used for detection, prognosis, diagnosis, or monitoring of conditions, disorders, or disease states, associated with aberrant changes in HPI gene expression, in particular hepatoma or recrudescence of hepatoma following surgical or other therapy. In a particular embodiment, such a hybridization assay is carried out by a method comprising contacting a patient sample containing nucleic acid with a nucleic acid probe capable of hybridizing to a DNA or RNA encoding an HPI, under conditions such that hybridization can occur, and detecting or measuring any resulting hybridization.

The invention also provides diagnostic kits, comprising in one or more containers an anti-HPI antibody. In addition, such a kit may optionally comprise one or more of the following: (1) instructions for using the anti-HPI antibody for diagnosis, prognosis, therapeutic monitoring, drug development or any combination of these applications; (2) a regulatory notice, i.e. a notice in the form prescribed or approved by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human testing or administration (3) a labeled binding partner to the antibody; and (4) a solid phase (such as a reagent strip) upon which the anti-HPI antibody is immobilized. If no labeled binding partner to the antibody is provided, the anti-HPI antibody itself can be labeled with a detectable marker, *e.g.*, a chemiluminescent, enzymatic, fluorescent, or radioactive moiety.

The invention also provides a kit comprising in one or more

containers a nucleic acid probe capable of hybridizing to RNA encoding a distinct HPI. In a specific embodiment, a kit can comprise in one or more containers a pair of primers (*e.g.*, each in the size range of 6-30 nucleotides, more preferably 10-20 nucleotides) that are capable of priming amplification -- such as by polymerase chain reaction (see *e.g.*, Innis et al., 1990, PCR Protocols, Academic Press, Inc., San Diego, CA), ligase chain reaction (see EP 320,308) use of Qβ replicase, cyclic probe reaction, or other methods known in the art -- under appropriate reaction conditions of at least a portion of a nucleic acid encoding an HPI.

Kits are also provided which allow for the detection of a plurality of HPIs or a plurality of nucleic acids each encoding an HPI. A kit can optionally further comprise a predetermined amount of an isolated HPI protein or a nucleic acid encoding an HPI, *e.g.*, for use as a standard or control.

5.3. Use in Clinical Studies 5.3. Use in Clinical Studies

The diagnostic methods and compositions of the present invention can assist in conducting or monitoring a clinical study, *e.g.*, for testing drugs for therapy of hepatoma. In one embodiment, candidate molecules are tested for their ability to restore HF or HPI levels in a patient suffering from hepatoma towards levels found in subjects not suffering from hepatoma or, in a treated patient (*e.g.* after surgery) to maintain HF or HPI levels at or near non-hepatoma values. The levels of one or more HFs or HPIs can be assayed.

In another embodiment, the methods and compositions of the present invention are used to identify individuals with hepatoma when screening candidates for a clinical study; such individuals can then be included in or excluded from the study or can be placed in a separate cohort for treatment or analysis. If desired, the candidates can concurrently be screened to identify individuals with Hepatitis B and/or Hepatitis C; procedures for these screens are well known in the art and include, for instance, serological studies to detect antibodies to one or more Hepatitis B or Hepatitis C antigens, and PCR studies to identify one or more oligonucleotide sequences from the Hepatitis B or Hepatitis C genome.

5.4. Purification of HPIs

In particular aspects, the invention provides isolated HPIs, preferably human HPIs, and fragments and derivatives thereof which comprise an antigenic determinant (*i.e.*, can be recognized by an antibody) or which are otherwise functionally active, as well as nucleic acid sequences encoding the foregoing. "Functionally active" HPI as used herein refers to that material displaying one or more known functional activities associated with a full-length (wild-type) HPI, *e.g.*, binding

to an HPI substrate or HPI binding partner, antigenicity (binding to an anti-target antibody), immunogenicity, etc.

In specific embodiments, the invention provides fragments of an HPI comprising at least 6 amino acids, 10 amino acids, 50 amino acids, or at least 75 amino acids. Fragments, or proteins comprising fragments, lacking some or all of the regions of an HPI are also provided. Nucleic acids encoding the foregoing are provided.

Once a recombinant nucleic acid which expresses the HPI gene sequence is identified, the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labelling of the product followed by analysis by gel electrophoresis, immunoassay, etc.

Once the HPI is identified, it can be isolated and purified by standard methods including chromatography (*e.g.*, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins.

Alternatively, once an HPI produced by a recombinant nucleic acid is identified, the entire amino acid sequence of the HPI can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, see Hunkapiller, et al., 1984, Nature 310:105-111).

In another alternative embodiment, native HPIs can be purified from natural sources, by standard methods such as those described above (*e.g.*, immunoaffinity purification).

In a preferred embodiment, HPIs are isolated by the Preferred Technology described in U.S. Application No. 08/980,574 and WO 98/23950 which are incorporated herein by reference. For preparative-scale runs, a narrow-range "zoom gel" having a pH range of 2 pH units or less is preferred for the isoelectric focusing step, according to the method described in Westermeier, 1993, Electrophoresis in Practice (VCH, Weinheim, Germany), pp. 197-209 (which is incorporated herein by reference in its entirety); this modification permits a larger quantity of a target protein to be loaded onto the gel, and thereby increases the quantity of isolated HPI that can be recovered from the gel. When used in this way for preparative-scale runs, the Preferred Technology typically provides up to 100 ng, and can provide up to 1000 ng, of an isolated HPI in a single run. Those of skill in the art will appreciate that a zoom gel can be used in any separation strategy which employs gel isoelectric focusing.

In a specific embodiment of the present invention, such HPIs, whether produced by recombinant DNA techniques or by chemical synthetic methods or by purification of native proteins, include (but are not

limited to) those containing, as a primary amino acid sequence, all or part of the amino acid sequence of the HPI, as well as fragments and other derivatives, and analogs thereof, including proteins homologous thereto.

5.5. Production of Antibodies to HPIs

According to the invention, an HPI, its fragments or other derivatives, or analogs thereof, may be used as an immunogen to generate antibodies which immunospecifically bind such an immunogen. Such proteins, fragments, derivatives, or analogs can be isolated by any convenient means, including the methods described in the preceding section of this application. The antibodies generated include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In a specific embodiment, antibodies to a human HPI are produced. In another embodiment, antibodies to a domain of an HPI are produced. In a specific embodiment, hydrophilic fragments of an HPI are used as immunogens for antibody production.

Various procedures known in the art may be used for the production of polyclonal antibodies to an HPI or derivative or analog. In a particular embodiment, rabbit polyclonal antibodies to an epitope of an HPI, or a subsequence thereof, can be obtained. For the production of antibody, various host animals can be immunized by injection with the native HPI, or a synthetic version, or derivative (*e.g.*, fragment) thereof, including but not limited to rabbits, mice, rats, horses, goats, chickens, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to complete or incomplete Freund's adjuvant, mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and Corynebacterium parvum.

For preparation of monoclonal antibodies directed toward an HPI sequence or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). (Each of the foregoing references in incorporated herein by reference.) In an additional embodiment of the invention, monoclonal antibodies can be produced in

germ-free animals as described in PCT/US90/02545, which is incorporated herein by reference. According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote et al., 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030) or by transforming human B cells with EBV virus *in vitro* (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). In fact, according to the invention, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger et al., 1984, Nature 312:604-608; Takeda et al., 1985, Nature 314:452-454) by splicing the genes from a mouse antibody molecule specific for an HPI together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention. (Each of the foregoing references is incorporated herein by reference.)

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778, incorporated herein by reference) can be adapted to produce HPI-specific single-chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for HPIs, derivatives, or analogs.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent, and Fv fragments.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, *e.g.*, ELISA (enzyme-linked immunosorbent assay). For example, to select antibodies which recognize a specific domain of an HPI, one may assay generated hybridomas for a product which binds to an HPI fragment containing such a domain. For selection of an antibody that specifically binds a first HPI homolog but which does not specifically bind a different HPI homolog, one can select on the basis of positive binding to the first HPI homolog and a lack of binding to the second HPI homolog. Similarly, for selection of an antibody that specifically binds an HPI but which does not specifically bind a different isoform of the same protein (*e.g.*, a different glycoform having the same core peptide as the HPI), one can select on the basis of positive binding to the HPI and a lack of binding to the different isoform (*e.g.*, glycoform).

Antibodies specific to a domain of an HPI are also provided.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the HPIs of the invention, *e.g.*, for imaging these proteins, measuring levels thereof in appropriate physiological samples, in diagnostic methods, etc.

5.6. Isolation Of DNA Encoding An HPI

Specific embodiments for the cloning of an HPI gene, are presented below by way of example and not of limitation.

The nucleotide sequences of the present invention, including DNA and RNA, and comprising a sequence encoding the HPI or a fragment or analog thereof, may be synthesized using methods known in the art, such as using conventional chemical approaches or polymerase chain reaction (PCR) amplification of overlapping oligonucleotides. The sequences also provide for the identification and cloning of the HPI gene from any species, for instance for screening cDNA libraries, genomic libraries or expression libraries.

The nucleotide sequences comprising a sequence encoding an HPI of the present invention are useful for their ability to selectively form duplex molecules with complementary stretches of other protein genes. Depending on the application, a variety of hybridization conditions may be employed to achieve varying sequence identities.

For a high degree of selectivity, relatively stringent conditions are used to form the duplexes, such as low salt or high temperature conditions. As used herein, "highly stringent conditions" means hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3; incorporated herein by reference in its entirety.) For some applications, less stringent hybridization conditions are required. As used herein "moderately stringent conditions" means washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*). Hybridization conditions can also be rendered more stringent by the addition of increasing amounts of formamide, to destabilize the hybrid duplex. Thus, particular hybridization conditions can be readily manipulated, and will generally be chosen depending on the desired results. For example, convenient hybridization temperatures in the presence of 50% formamide are: 42°C for a probe which is 95 to 100% homologous to the HPI gene fragment, 37°C for 90 to 95% homology and 32°C for 70 to 90% homology.

In the preparation of genomic libraries, DNA fragments are generated, some of which will encode a part or the whole of an HPI. The DNA may be cleaved at specific sites using various restriction enzymes.

Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis, column chromatography and sucrose gradient centrifugation. The DNA fragments can then be inserted into suitable vectors, including but not limited to plasmids, cosmids, bacteriophages lambda or T_4, and yeast artificial chromosome (YAC). (See, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II; Ausubel F.M. et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York). The genomic library may be screened by nucleic acid hybridization to labeled probe (Benton and Davis, 1977, Science 196:180; Grunstein and Hogness, 1975, Proc. Natl. Acad. Sci. USA 72:3961).

The genomic libraries may be screened with labeled degenerate oligonucleotide probes corresponding to the amino acid sequence of any peptide of the HPI using optimal approaches well known in the art. Any probe used preferably is at least 10 nucleotides (more preferably 15 nucleotides, still more preferably 20 nucleotides) in length.

As shown in Tables III and IV above, some HPIs disclosed herein correspond to previously identified proteins encoded by genes whose sequences are publicly known. To screen such a gene, any probe may be used that is complementary to the gene or its complement; preferably the probe is 10 nucleotides or longer, more preferably 15 nucleotides or longer. The Entrez database held by the National Center for Biotechnology Information (NCBI) -- which is accessible at http://www.ncbi.nlm.nih.gov/ -- (and, for HPI-16, the NCBI non-redundant database) provides gene sequences for these HPIs under the following accession numbers, and each sequence is incorporated herein by reference:

Table V. Gene sequences of HPI-related proteins

HF #	HPI	Accession numbers
1	HPI-1	AA617854, AA630254, AA580356, AA715907, AA614684, AA580429, AA523377
17	HPI-2	U24578, M59815, M59816, K02403, AF019413
18	HPI-3	T40940, T40002
20	HPI-4	T19152, H73939, T40182, T40167, T40158, Z20894
21	HPI-5	T19152, H73939, T40182, T40167, T40158, Z20894

23	HPI-6	U24578, M59815, M59816, K02403, AF019413
25	HPI-8	U24578, M59815, M59816, K02403, AF019413
28	HPI-9	T41010, T40102, T40058, T39954
29	HPI-10	T41056, T40108, Z21017, Z20888, Z19984, Z19971
29	HPI-11	T40178, Z21027, Z19988
31	HPI-12	T19152, H73939, T40182, T40167, T40158, Z20894
33	HPI-14	U24578, M59815, M59816, K02403, AF019413
34	HPI-15	U24578, M59815, M59816, K02403, AF019413
36	HPI-16	NCBI non-redundant database published 11/22/97, Accession # 1082547
39	HPI-17	U24578, M59815, M59816, K02403, AF019413
44	HPI-18	U24578, M59815, M59816, K02403, AF019413
51	HPI-19	AA269874
52	HPI-20	AA269874
55	HPI-22	AA269874

For each of HPI-13 and HPI-21, a degenerate set of probes is provided, as follows:

(a) Probes for HPI-21

```
5'- A T A T G G A T A G G A A C A A C A A G A -3'
       T         T   G   G   G   G     G
       C         C   T   T   T
                     C   C   C

5'- A T A T G G A T A G G A A C A A C A C G A -3'
       T         T   G   G   G   G     G
       C         C   T   T   T   T     T
                     C   C   C   C     C

5'- A T A T G G T T A G G A A C A A C A A G A -3'
       T         G   G   G   G   G     G
       C             T   T   T
                     C   C   C

5'- A T A T G G T T A G G A A C A A C A C G A -3'
       T         G   G   G   G   G     G
       C             T   T   T   T     T
                     C   C   C   C     C

5'- A T A T G G C T A G G A A C A A C A A G A -3'
       T         G   G   G   G   G     G
       C         T   T   T   T
                 C   C   C   C
```

```
5'- A T A T G G C T A G G A A C A A C A C G A -3'
       T           G     G     G     G     G
       C           T     T     T     T     T
                   C     C     C     C     C

5'- T T A T G G A T A G G A A C A A C A A G A -3'
       G           T     G     G     G     G
                   C     T     T     T
                         C     C     C

5'- T T A T G G A T A G G A A C A A C A C G A -3'
       G           T     G     G     G     G
                   C     T     T     T     T
                         C     C     C     C

5'- T T A T G G T T A G G A A C A A C A A G A -3'
       G           G     G     G     G     G
                         T     T     T
                         C     C     C

5'- T T A T G G T T A G G A A C A A C A C G A -3'
       G           G     G     G     G     G
                         T     T     T     T
                         C     C     C     C

5'- T T A T G G C T A G G A A C A A C A A G A -3'
       G           G     G     G     G     G
                   T     T     T     T
                   C     C     C     C

5'- T T A T G G C T A G G A A C A A C A C G A -3'
       G           G     G     G     G     G
                   T     T     T     T     T
                   C     C     C     C     C

5'- C T A T G G A T A G G A A C A A C A A G A -3'
       G           T     G     G     G     G
       T           C     T     T     T
       C                 C     C     C

5'- C T A T G G A T A G G A A C A A C A C G A -3'
       G           T     G     G     G     G
       T           C     T     T     T     T
       C                 C     C     C     C

5'- C T A T G G T T A G G A A C A A C A A G A -3'
       G           G     G     G     G     G
       T                 T     T     T
       C                 C     C     C

5'- C T A T G G T T A G G A A C A A C A C G A -3'
       G           G     G     G     G     G
       T                 T     T     T     T
       C                 C     C     C     C
```

```
5'- C T A T G G C T A G G A A C A A C A A G A -3'
       G         G   G   G   G   G
       T         T   T   T   T
       C         C   C   C   C

5'- C T A T G G C T A G G A A C A A C A C G A -3'
       G         G   G   G   G   G
       T         T   T   T   T   T
       C         C   C   C   C   C
```

(b) Probes for HPI-13

```
5'-G C A G A A A C A A C A G A T T T T -3'
     G   G   G   G   C   C
     T       T   T
     C       C   C
```

Clones in libraries with insert DNA encoding the HPI or fragments thereof will hybridize to one or more of the degenerate oligonucleotide probes (or their complement). Hybridization of such oligonucleotide probes to genomic libraries are carried out using methods known in the art. For example, hybridization with one of the above-mentioned degenerate sets of oligonucleotide probes, or their complement (or with any member of such a set, or its complement) can be performed under highly stringent or moderately stringent conditions as defined above, or can be carried out in 2X SSC, 1.0% SDS at 50°C and washed using the same conditions.

In yet another aspect, clones of nucleotide sequences encoding a part or the entire HPI or HPI-derived polypeptides may also be obtained by screening expression libraries. For example, DNA from the relevant source is isolated and random fragments are prepared and ligated into an expression vector (*e.g.*, a bacteriophage, plasmid, phagemid or cosmid) such that the inserted sequence in the vector is capable of being expressed by the host cell into which the vector is then introduced. Various screening assays can then be used to select for the expressed HPI or HPI-derived polypeptides. In one embodiment, the various anti-HPI antibodies of the invention can be used to identify the desired clones using methods known in the art. See, for example, Harlow and Lane, 1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Appendix IV. Clones or plaques from the library are brought into contact with the antibodies to identify those clones that bind.

In an embodiment, colonies or plaques containing DNA that encodes an HPI or HPI-derived polypeptide can be detected using DYNA Beads according to Olsvick et al., 29th ICAAC, Houston, Tex. 1989, incorporated herein by reference. Anti-HPI antibodies are crosslinked to tosylated DYNA Beads M280, and these antibody-containing beads would then be used

to adsorb to colonies or plaques expressing HPI or HPI-derived polypeptide. Colonies or plaques expressing an HPI or HPI-derived polypeptide are identified as any of those that bind the beads.

Alternatively, the anti-HPI antibodies can be nonspecifically immobilized to a suitable support, such as silica or Celite™ resin. This material would then be used to adsorb to bacterial colonies expressing the HPI protein or HPI-derived polypeptide as described in the preceding paragraph.

In another aspect, PCR amplification may be used to produce substantially pure DNA encoding a part of or the whole of an HPI from genomic DNA. Oligonucleotide primers, degenerate or otherwise, corresponding to known HPI sequences can be used as primers.

PCR can be carried out, *e.g.*, by use of a Perkin-Elmer Cetus thermal cycler and a thermostable DNA polymerase such as *Thermus aquaticus* DNA polymerase (Gene Amp™ or AmpliTaq® DNA polymerase). One can choose to synthesize several different degenerate primers, for use in the PCR reactions. It is also possible to vary the stringency of hybridization conditions used in priming the PCR reactions, to allow for greater or lesser degrees of nucleotide sequence similarity between the degenerate primers and the corresponding sequences in the DNA. After successful amplification of a segment of the sequence encoding an HPI, that segment may be molecularly cloned and sequenced, and utilized as a probe to isolate a complete genomic clone. This, in turn, will permit the determination of the gene's complete nucleotide sequence, the analysis of its expression, and the production of its protein product for functional analysis, as described *infra*.

The HPI gene can also be identified by mRNA selection by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified HPI DNA of another species (*e.g.*, mouse, human). Immunoprecipitation analysis or functional assays (*e.g.*, aggregation ability *in vitro*; binding to receptor) of the *in vitro* translation products of the isolated products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against an HPI. A radiolabelled HPI cDNA can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabelled mRNA or cDNA may then be used as a probe to identify the HPI DNA fragments from among other genomic DNA fragments.

Alternatives to isolating HPI genomic DNA include, but are not limited to, chemically synthesizing the gene sequence itself from a known

sequence or making cDNA to the mRNA which encodes the HPI. For example, RNA for cDNA cloning of the HPI gene can be isolated from cells which express the HPI. Other methods are possible and within the scope of the invention.

Any eukaryotic cell potentially can serve as the nucleic acid source for the molecular cloning of the HPI gene. The nucleic acid sequences encoding the HPI can be isolated from vertebrate, mammalian, human, porcine, bovine, feline, avian, equine, canine, as well as additional primate sources, insects, plants, etc. The DNA may be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell. (See, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will contain only exon sequences. Whatever the source, the HPI gene should be molecularly cloned into a suitable vector for propagation.

The identified and isolated gene or cDNA can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include, but are not limited to, bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene). The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and HPI gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, etc., so that many copies of the gene sequence are generated.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate the isolated HPI gene, cDNA, or synthesized DNA sequence enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing

transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

The HPI sequences provided by the instant invention include those nucleotide sequences encoding substantially the same amino acid sequences as found in native HPIs, and those encoded amino acid sequences with functionally equivalent amino acids, as well as those encoding other target derivatives or analogs.

5.7. Expression of DNA Encoding HPIs

The nucleotide sequence encoding an HPI or a functionally active analog or fragment or other derivative thereof can be inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. The necessary transcriptional and translational signals can also be supplied by the native HPI gene or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.*, vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (*e.g.*, baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used. In specific embodiments, the human HPI gene is expressed, or a sequence encoding a functionally active portion of the human HPI. In yet another embodiment, a fragment of target comprising a domain of the HPI is expressed.

Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional and translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of nucleic acid sequence encoding an HPI or peptide fragment may be regulated by a second nucleic acid sequence so that the HPI or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of an HPI gene may be controlled by any promoter or enhancer element known in the art. Promoters which may be used to control HPI gene expression include, but are not limited to, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. USA 78:1441-1445), the

regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42); prokaryotic expression vectors such as the β-lactamase promoter (Villa-Kamaroff, et al., 1978, Proc. Natl. Acad. Sci. USA 75:3727-3731), or the *tac* promoter (DeBoer, et al., 1983, Proc. Natl. Acad. Sci. USA 80:21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38:647-658; Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer et al., 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

In a specific embodiment, a vector is used that comprises a promoter operably linked to an HPI-encoding nucleic acid, one or more origins of replication, and, optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene).

In a specific embodiment, an expression construct is made by subcloning an HPI coding sequence into the *EcoRI* restriction site of each

of the three pGEX vectors (Glutathione S-Transferase expression vectors; Smith and Johnson, 1988, Gene 7:31-40). This allows for the expression of the HPI product from the subclone in the correct reading frame.

Expression vectors containing HPI gene inserts can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or absence of "marker" gene functions, and (c) expression of inserted sequences. In the first approach, the presence of an HPI gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted HPI gene. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, etc.) caused by the insertion of an HPI gene in the vector. For example, if the HPI gene is inserted within the marker gene sequence of the vector, recombinants containing the HPI gene insert can be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying the HPI gene product expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the HPI gene product in *in vitro* assay systems, *e.g.*, binding with anti-HPI antibody.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As previously explained, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (*e.g.*, lambda), and plasmid and cosmid DNA vectors, to name but a few.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered HPI may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, phosphorylation of proteins). Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a

glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may affect processing reactions to different extents.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the differentially expressed or pathway gene protein may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the differentially expressed or pathway gene protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the differentially expressed or pathway gene protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, et al., 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy et al., 1980, Cell 22:817) genes can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, et al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare, et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, et al., 1981, J. Mol. Biol. 150:1); and hygro, which confers resistance to hygromycin (Santerre, et al., 1984, Gene 30:147) genes.

In other specific embodiments, the HPI, fragment, analog, or derivative may be expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence (of a different protein)). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in

the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer.

Both cDNA and genomic sequences can be cloned and expressed.

6. EXAMPLE: PROTEINS FROM SERUM OF PATIENTS WITH AND WITHOUT HEPATOMA

Using the following Reference Protocol, proteins in serum from 33 patients with hepatoma and 19 patients with cirrhosis of the liver were separated by isoelectric focusing followed by SDS-PAGE and analyzed. Each sample was run in duplicate.

6.1. Sample preparation

A protein assay was carried out on the serum sample as received (Pierce BCA Cat # 23225). A volume of serum corresponding to 300µg of total protein was aliquoted and an equal volume of 10% (w/v) SDS (Fluka 71729), 2.3% (w/v) dithiothreitol (BDH 443852A) was added. The sample was heated at 95°C for 5 mins, and then allowed to cool to 20°C . 125µl of the following buffer was then added to the sample:

8M urea (BDH 452043w)
4% CHAPS (Sigma C3023)
65mM dithiotheitol (DTT)
2% (v/v) Resolytes 3.5-10 (BDH 44338 2x)

This mixture was vortexed, and centrifuged at 13000 rpm for 5 mins at 15°C, and the supernatant was analyzed by isoelectric focusing.

6.2. Isoelectric Focusing

Isoelectric focusing (IEF), was performed using the Immobiline® DryStrip Kit (Pharmacia BioTech), following the procedure described in the manufacturer's instructions, *see* Instructions for Immobiline® DryStrip Kit, Pharmacia, # 18-1038-63, Edition AB (incorporated herein by reference in its entirety). Immobilized pH Gradient (IPG) strips (18cm, pH 3-10 non-linear strips; Pharmacia Cat. # 17-1235-01) were rehydrated overnight at 20°C in a solution of 8M urea, 2% (w/v) CHAPS, 10mM DTT, 2% (v/v) Resolytes 3.5-10, as described in the Immobiline DryStrip Users Manual. For IEF, 50µl of supernatant (prepared as above) was loaded onto a strip, with the cup-loading units being placed at the basic end of the strip. The loaded gels were then covered with mineral oil (Pharmacia 17-3335-01) and a voltage was immediately applied to the strips according to the following profile, using a Pharmacia EPS3500XL power supply (Cat 19-3500-01):

Initial voltage = 300V for 2 hrs
Linear Ramp from 300V to 3500V over 3hrs
Hold at 3500V for 19hrs

For all stages of the process, the current limit was set to 10mA for 12 gels, and the wattage limit to 5W. The temperature was held at 20°C

throughout the run.

6.3. Gel Equilibration and SDS-PAGE 6.3. Gel Equilibration and SDS-PAGE

After the final 19hr step, the strips were immediately removed and immersed for 10 mins at 20°C in a first solution of the following composition: 6M urea; 2% (w/v) DTT; 2% (w/v) SDS; 30% (v/v) glycerol (Fluka 49767); 0.05M Tris/HCl, pH 6.8 (Sigma Cat T-1503). The strips were removed from the first solution and immersed for 10 mins at 20°C in a second solution of the following composition: 6M urea; 2% (w/v) iodoacetamide (Sigma I-6125); 2% (w/v) SDS; 30% (v/v) glycerol; 0.05M Tris/HCl, pH 6.8. After removal from the second solution, the strips were loaded onto supported gels for SDS-PAGE according to Hochstrasser et al., 1988, Analytical Biochemistry 173: 412-423 (incorporated herein by reference in its entirety), with modifications as specified below.

6.4. Preparation of supported gels 6.4. Preparation of supported gels

The gels were cast between two glass plates of the following dimensions: 23cm wide x 24cm long (back plate); 23cm wide x 24cm long with a 2cm deep notch in the central 19cm (front plate). To promote covalent attachment of SDS-PAGE gels, the back plate was treated with a 0.4% solution of γ-methacryl-oxypropyltrimethoxysilane in ethanol (BindSilane™; Pharmacia Cat. # 17-1330-01). The front plate was treated with RepelSilane™ (Pharmacia Cat. # 17-1332-01) to reduce adhesion of the gel. Excess reagent was removed by washing with water, and the plates were allowed to dry. At this stage, both as identification for the gel, and as a marker to identify the coated face of the plate, an adhesive bar-code was attached to the back plate in a position such that it would not come into contact with the gel matrix.

The dried plates were assembled into a casting box with a capacity of 13 gel sandwiches. The top and bottom plates of each sandwich were spaced by means of 1mm thick spacers, 2.5 cm wide. The sandwiches were interleaved with acetate sheets to facilitate separation of the sandwiches after gel polymerization. Casting was then carried out according to Hochstrasser et al., *op. cit.*

A 9-16% linear polyacrylamide gradient was cast, extending up to a point 2cm below the level of the notch in the front plate, using the Angelique gradient casting system (Large Scale Biology). Stock solutions were as follows. Acrylamide (40% in water) was from Serva (Cat. # 10677). The cross-linking agent was PDA (BioRad 161-0202), at a concentration of 2.6% (w/w) of the total starting monomer content. The gel buffer was 0.375M Tris/HCl, pH 8.8. The polymerization catalyst was 0.05% (v/v)

TEMED (BioRad 161-0801), and the initiator was 0.1% (w/v) APS (BioRad 161-0700). No SDS was included in the gel and no stacking gel was used. The cast gels were allowed to polymerize at 20°C overnight, and then stored at 4°C in sealed polyethylene bags with 6ml of gel buffer, and were used within 4 weeks.

6.5. SDS-PAGE 6.5. SDS-PAGE

A solution of 0.5% (w/v) agarose (Fluka Cat 05075) was prepared in running buffer (0.025M Tris, 0.198M glycine (Fluka 50050), 1% (w/v) SDS, supplemented by a trace of bromophenol blue). The agarose suspension was heated to 70°C with stirring, until the agarose had dissolved. The top of the supported 2nd D gel was filled with the agarose solution, and the equilibrated strip was placed into the agarose, and tapped gently with a palette knife until the gel was intimately in contact with the 2nd D gel. The gels were placed in the 2nd D running tank, as described by Amess et al., 1995, Electrophoresis 16: 1255-1267 (incorporated herein by reference in its entirety). The tank was filled with running buffer (as above) until the level of the buffer was just higher than the top of the region of the 2nd D gels which contained polyacrylamide, so as to achieve efficient cooling of the active gel area. Running buffer was added to the top buffer compartments formed by the gels, and then voltage was applied immediately to the gels using a Consort E-833 power supply. For 1 hour, the gels were run at 20mA/gel. The wattage limit was set to 150W for a tank containing 6 gels, and the voltage limit was set to 600V. After 1 hour, the gels were then run at 40mA/gel, with the same voltage and wattage limits as before, until the bromophenol blue line was 0.5cm from the bottom of the gel. The temperature of the buffer was held at 10°C throughout the run.

6.6. Staining 6.6. Staining

Upon completion of the electrophoresis run, the gels were immediately removed from the tank for fixation. The top plate of the gel cassette was carefully removed, leaving the gel bonded to the bottom plate. The bottom plate with its attached gel was then placed into a staining apparatus, which can accommodate 12 gels. The gels were completely immersed in fixative solution of 40% (v/v) ethanol (BDH 28719), 10% (v/v) acetic acid (BDH 100016X), 50% (v/v) water (MilliQ-Millipore), which was continuously circulated over the gels. After an overnight incubation, the fixative was drained from the tank, and the gels were primed by immersion in 7.5% (v/v) acetic acid, 0.05% (w/v) SDS, 92.5% (v/v) water for 30 mins. The priming solution was then drained, and the gels were stained by complete immersion in a staining solution for 4 hours. A solution of fluorescent dye was prepared by diluting

Sypro Red (Molecular Probes, Inc., Eugene, Oregon) according to the manufacturer's instructions; this diluted solution was filtered under vacuum though a 0.4um filter.

6.7. Imaging of the gel 6.7. Imaging of the gel

A computer-readable output was produced by imaging the fluorescently stained gels with a Storm scanner (Molecular Dynamics, Sunnyvale, California) according to the manufacturer's instructions, (see Storm User's Guide, 1995, Version 4.0, Part No. 149-355, incorporated herein by reference in its entirety) with modifications as described below. The gels were removed from the stain, rinsed with water briefly, and imaged on the Storm Scanner, in Red Fluorescence mode with a PMT setting of 1000V, and a resolution of 200 µm. Since the gel was rigidly bonded to a glass plate, the gel was held in contact with the scanner bed during imaging. To avoid interference patterns arising from non-uniform contact between the gel and the scanner bed, a film of water was introduced under the gel, taking care to avoid air pockets. Moreover, the gel was placed in a frame provided with two fluorescent buttons that were imaged together with the gel to provide reference points (designated M1 and M2) for determining the x,y coordinates of other features detected in the gel. A matched frame was provided on a robotic gel excisor in order to preserve accurate alignment of the gel. After imaging, the gels were sealed in polyethylene bags containing a small volume of staining solution, and then stored at 4°C.

6.8. Digital Analysis of the Data 6.8. Digital Analysis of the Data

The data were processed as described in U.S. Application No. 08/980,574, Sections 5.4 and 5.5 (incorporated herein by reference), as set forth more particularly below.

6.8.1. Computer Analysis Of The Detector Output

The output from the scanner was first processed using the MELANIE® II 2D PAGE analysis program (Release 2.2, 1997, BioRad Laboratories, Hercules, California, Cat. # 170-7566) to autodetect the registration points, M1 and M2; to autocrop the images (*i.e.*, to eliminate signals originating from areas of the scanned image lying outside the boundaries of the gel, *e.g.*, the reference frame); to filter out artifacts due to dust; to detect and quantify features; and to create image files in GIF format. Features were detected using the following parameters:

Smooths =2
Laplacian threshold 50
Partials threshold 1
Saturation = 100
Peakedness = 0

Minimum Perimeter = 10

6.9. Assignment of pI and MW Values

Images were evaluated to reject images which had gross abnormalities, or were of too low a loading or overall image intensity, or were of too poor a resolution, or where duplicates were too dissimilar. If one image of a duplicate was rejected then the other image belonging to the duplicate was also rejected regardless of image quality. Samples that were rejected were scheduled for repeat analysis.

Landmark identification was used to determine the pI and MW values of features detected in the images. This process involves the identification of certain proteins which are expected to be found in any given biological sample. As these common proteins exhibit an identical isoelectric point and molecular weight from sample to sample, they can be used as standards; this process also corrects for any possible gel variation or distortion.

From the dataset of normal serum gels, a gel was arbitrarily chosen as the Primary Master Gel. Landmark features were then identified by comparing the features detected in this Primary Master Gel with features previously identified on 2D electrophoresis of normal human serum. (*see* Bjellqvist et al., 1993, Electrophoresis 14: 1357-1365; incorporated herein by reference in its entirety).

Fourteen landmark features, designated PL1 to PL12 and PL15 to PL16, were identified in the Primary Master Gel. These landmark features are identified in Figure 1 and were assigned the pI and/or MW values indicated in Table V.

Table V. Landmark Features used in this study

Name	pI	MW (kd)	Name	pI	MW (kd)
PL1	None	186,073	PL8	6.47	47,195
PL2	6.20	100,000	PL9	5.29	43,541
PL3	4.73	93,708	PL10	5.22	23,000
PL4	5.13	73,465	PL11	4.47	25,183
PL5	4.97	52,739	PL12	5.52	13,800
PL6	4.10	None	PL15	7.80	36,962
PL7	4.80	40,997	PL16	8.58	None

As many of these landmarks as possible were identified in each gel image in the dataset.

All features in the Master gel were then assigned a pI value by linear interpolation/extrapolation (using the MELANIE II software) to the pI of the two nearest landmarks that had been assigned a pI value, and were assigned a MW value by linear interpolation/extrapolation (using the MELANIE II software) to the MW of the two nearest landmarks that had been assigned a MW value. Each feature was also labelled with a unique number known as its Molecular Cluster Index (or MCI).

Secondary Master gels were chosen for both the cirrhosis gels and the HCC gels. Features in these gels were paired with common features in the Master gel, using the algorithm supplied with the MELANIE II software, as described at Section A, pp. 8-10 of the MELANIE II 2D PAGE (Release 2.2) User Manual (The Melanie Group, Geneva, Switzerland). Features that have been paired are linked to the corresponding MCI, and hence to an associated pI and MW value. Unpaired features present in these secondary master gels were assigned pI and MW values by linear interpolation/extrapolation (using the MELANIE II software) with respect to pI and MW of the landmarks. Additional unique entries were then created in the MCI for these features.

6.9.1. Construction of Profiles 6.9.1.

Construction of Profiles

All gels in the dataset were now matched to the Primary and Secondary Master Gels, and paired features were linked to the corresponding entries in the Molecular Cluster Index.

Duplicate gels were then aligned via the landmarks and a matching process performed so as to pair identical spots on the duplicate gels. This provided increased assurance that subsequently measured isoelectric points and molecular weights were accurate, as paired spots demonstrated the reproducibility of the separation and also filtered out artefacts.

A measurement of the intensity of each protein spot was taken and stored. Each protein spot was assigned an identification code and matched to a spot on the Master gel.

The end result of this aspect of the analysis was the generation, for each duplicate set of gels representing a single serum sample, of a digital profile which contained, for each identified spot: 1) a unique arbitrary identification code, 2) the x,y coordinates, 3) the isoelectric point, 4) the molecular weight, 5) the signal value, 6) the standard deviation for each of the preceding measurements, and 7) a pointer to the MCI of the spot on the master gel to which this spot was matched. By virtue of the Laboratory Information Management System (LIMS), this

profile was traceable to the actual stored gel from which it was generated, so that proteins identified by computer analysis of gel profile databases could be retrieved. The LIMS also permitted the profile to be traced back to the original sample or patient.

6.9.2. Cross-Matching Between Samples 6.9.2. Cross-Matching Between Samples

Once the profile was generated, analysis was directed toward the selection of interesting proteins. Each significant feature in a profile was assigned an index, the "Molecular Cluster Index" (MCI) that identifies the feature in all gels and that serves as a pointer to parameters (1) to (7) above of the feature. A molecular cluster table was generated from the master gel for each sample type (*i.e.* hepatoma serum and cirrhosis serum). Gels from all other samples of the same type were matched with the relevant primary and secondary master gels. The digital profiles for each sample were then annotated by adding, for each matched feature, the MCI assigned to that feature in the master profile.

6.9.3. Differential Analysis of the Profiles

Within each sample set (hepatoma serum or cirrhosis serum), the profiles were analyzed to identify and select those features present in at least 50% of the profiles. These selected features were then assembled into a hepatoma serum feature set and a cirrhosis serum feature set. Matching features of each feature set were then compared to identify those features showing at least a 2-fold difference in mean intensity between hepatoma serum and cirrhosis serum. Differentially present features were identified as Hepatoma-Diagnostic Features (HFs).

6.9.4. Statistical Analysis Of HFs 6.9.4. Statistical Analysis Of HFs

For each HF, a statistical analysis was performed using Student's t-test to compare the distribution of signal intensities for that feature in the 33 hepatoma serum profiles and in the 19 cirrhosis serum profiles. A p value was derived for each HF.

6.10. Recovery and analysis of selected proteins 6.10. Recovery and analysis of selected proteins

Proteins in HFs were robotically excised and processed to generate tryptic peptides using the Preferred Technology as described in U.S. Application No. 08/980,574, which is incorporated herein by reference in its entirety. Partial amino acid sequences of these peptides were determined by mass spectroscopy, using de novo sequencing.

6.11. Results 6.11. Results

These initial experiments identified 17 features that were decreased and 44 features that were increased in hepatoma serum as compared with cirrhosis serum. Details of these HFs and a p value for each are provided in Tables I and II. Each HF was differentially present in hepatoma serum as compared with non-hepatoma serum ($p < 0.1$). For some preferred HFs (HF-1, HF-3, HF-10, HF-15, HF-16, HF-17, HF-18, HF-19, HF-20, HF-21, HF-23, HF-24, HF-25, HF-26, HF-28, HF-29, HF-31, HF-32, HF-33, HF-34, HF-36, HF-37, HF-38, HF-39, HF-40, HF-42, HF-44, HF-46, HF-47, HF-48, HF-49, HF-51, HF-52, HF-53, HF-54, HF-55, HF-56, HF-59 and HF-60), the difference was more significant ($p \leq 0.01$), and for certain highly preferred HFs (HF-1, HF-17, HF-18, HF-20, HF-21, HF-23, HF-24, HF-25, HF-28, HF-29, HF-31, HF-33, HF-34, HF-36, HF-39, HF-44, HF-51, HF-52, HF-54, HF-55, HF-59 and HF-60), the difference was still more significant ($p \leq 0.001$).

Partial amino acid sequences were determined for the differentially present HPIs in these HFs. Details of these HPIs are provided in Tables III and IV. Computer searches of public databases identified at least 2 HPIs for which neither the partial amino acid sequence, nor any oligonucleotide encoding such a peptide sequence, was described in any public database examined. Table IV illustrates that several HPIs are isoforms of the same protein. For example, HPI-2, HPI-6, HPI-8, HPI-14, HPI-15, HPI-17 and HPI-18 are isoforms of complement factor 4; HPI-4, HPI-5 and HPI-12 are isoforms of complement factor 3; and HPI-19, HPI-20 and HPI-21 are isoforms of ceruloplasmin. These isoforms are thought to arise from differences in post-translational processing (*e.g.*, glycosylation, phosphorylation, acylation or minimal proteolysis).

The present invention is not to be limited in scope by the particular exemplified embodiments, which are intended as illustrations of single aspects of the invention. Indeed, various modifications of the invention in addition to those enumerated herein will be apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

All publications cited herein are incorporated by reference in their entirety.

WHAT IS CLAIMED IS:

1. A method for screening, diagnosis or prognosis of hepatoma in a human subject or for monitoring the effect of an anti-hepatoma drug or therapy administered to a human subject, comprising:

(a) analyzing a sample of serum or plasma from the subject by two-dimensional electrophoresis to generate a two-dimensional array of features;

(b) for at least two chosen features whose intensity correlates with the presence or absence of hepatoma, comparing the intensity of each chosen feature in the sample with the intensity of that chosen feature in serum or plasma from one or more persons without hepatoma,

wherein the intensity of the chosen features in the sample indicates the presence or absence of hepatoma in the subject.

2. The method according to claim 1, wherein the chosen features do not include alphafetoprotein (AFP).

3. The method according to claim 1, wherein alphafetoprotein (AFP) is a chosen feature.

4. A method for screening, diagnosis or prognosis of hepatoma in a human subject or for monitoring the effect of an anti-hepatoma drug or therapy administered to a human subject, comprising:

(a) analyzing a sample of serum or plasma from the subject by two-dimensional electrophoresis to separate a plurality of proteins according to isoelectric point and electrophoretic mobility; and

(b) quantitatively detecting at least one of the following Hepatoma-Diagnostic Features (HFs): HF-1, HF-2, HF-3, HF-4, HF-5, HF-6, HF-7, HF-8, HF-9, HF-10, HF-11, HF-12, HF-13, HF-14, HF-15, HF-16, HF-17, HF-18, HF-19, HF-20, HF-21, HF-22, HF-23, HF-24, HF-25, HF-26, HF-27, HF-28, HF-29,

HF-30, HF-31, HF-32, HF-33, HF-34, HF-35, HF-36, HF-37, HF-38, HF-39, HF-40, HF-41, HF-42, HF-43, HF-44, HF-45, HF-46, HF-47, HF-48, HF-49, HF-50, HF-51, HF-52, HF-53, HF-54, HF-55, HF-56.

5. The method according to claim 4, wherein step (b) comprises quantitatively detecting at least one of the following HFs: HF-17, HF-18, HF-20, HF-21, HF-23, HF-24, HF-25, HF-28, HF-29, HF-31, HF-33, HF-34, HF-36, HF-39, HF-44, HF-51, HF-52, HF-54 and HF-55.

6. The method according to claim 5, wherein step (b) comprises quantitatively detecting at least one of the following HFs: HF-17, HF-18, HF-20, HF-23, HF-24, HF-25, HF-29, HF-33, HF-34, HF-39, HF-44 and HF-55.

7. The method according to claims 4, 5 or 6, wherein step (a) comprises isoelectric focussing followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

8. A method for screening, diagnosis or prognosis of hepatoma in a subject or for monitoring the effect of an anti-hepatoma drug or therapy administered to a subject, comprising:

(a) in a sample of serum or plasma from the subject, quantitatively detecting at least one of the following Hepatoma-Diagnostic Protein Isoforms (HPIs): HPI-1, HPI-2, HPI-3, HPI-4, HPI-5, HPI-6, HPI-8, HPI-9, HPI-10, HPI-11, HPI-12, HPI-13, HPI-14, HPI-15, HPI-16, HPI-17, HPI-18, HPI-19, HPI-20, HPI-21 and HPI-22.

9. The method according to claim 8, wherein the step of quantitatively detecting comprises testing at least one aliquot of the sample, said step of testing comprising:

(a) contacting the aliquot with an antibody that is

immunospecific for a preselected HPI; and

(b) detecting whether binding has occurred between the antibody and at least one species in the aliquot.

10. The method according to claim 9, wherein the antibody is a monoclonal antibody.

11. The method according to claim 9, wherein the step of quantitatively detecting comprises testing a plurality of aliquots with a plurality of antibodies.

12. The method according to claim 11, wherein the antibodies are monoclonal antibodies.

13. A preparation comprising one of the following isolated Hepatoma-Diagnostic Protein Isoforms (HPIs): HPI-1, HPI-2, HPI-3, HPI-4, HPI-5, HPI-6, HPI-8, HPI-9, HPI-10, HPI-11, HPI-12, HPI-13, HPI-14, HPI-15, HPI-16, HPI-17, HPI-18, HPI-19, HPI-20, HPI-21 and HPI-22.

14. A kit comprising the preparation of claim 13.

15. A kit comprising a plurality of preparations of claim 13.

16. A preparation comprising an isolated human protein, said protein comprising a peptide having the sequence AETTDF.

17. The preparation according to claim 16, wherein the protein has an isoelectric point (pI) of about 7.44 and an apparent molecular weight (MW) of about 63,882.

18. The preparation according to claim 17, wherein the pI is within 10% of 7.44 and the MW is within 10% of 63,882.

19. The preparation according to claim 18, wherein the pI is

within 5% of 7.44 and the MW is within 5% of 63,882.

20. The preparation according to claim 19, wherein the pI is within 1% of 7.44 and the MW is within 1% of 63,882.

21. A preparation comprising an isolated human protein, said protein comprising a peptide having one of the following sequences: IWIGTTR, IWLGTTR, LWIGTTR or LWLGTTR.

22. The preparation according to claim 21, wherein the protein has an isoelectric point (pI) of about 5.29 and an apparent molecular weight (MW) of about 43,541.

23. The preparation according to claim 22, wherein the pI is within 10% of 5.29 and the MW is within 10% of 43,541.

24. The preparation according to claim 23, wherein the pI is within 5% of 5.29 and the MW is within 5% of 43,541.

25. The preparation according to claim 24, wherein the pI is within 1% of 5.29 and the MW is within 1% of 43,541.

26. An antibody capable of immunospecific binding to one of the following Hepatoma-Diagnostic Protein Isoforms (HPIs): HPI-1, HPI-2, HPI-3, HPI-4, HPI-5, HPI-6, HPI-8, HPI-9, HPI-10, HPI-11, HPI-12, HPI-13, HPI-14, HPI-15, HPI-16, HPI-17, HPI-18, HPI-19, HPI-20, HPI-21 and HPI-22.

27. A kit comprising the antibody of claim 26.

28. A kit comprising a plurality of antibodies of claim 26.

29. The use of one or more HFs as defined in claim 4 in screening, diagnosis or prognosis of hepatoma in a human subject or for monitoring the effect of an anti-hepatoma drug.

30. The use of one or more HPIs as defined in claim 8 in screening, diagnosis or prognosis of hepatoma in a human subject or for monitoring the effect of an anti-hepatoma drug.

31. The use of a protein as defined in any one of claims 16 to 25 in screening, diagnosis or prognosis of hepatoma in a human subject or for monitoring the effect of an anti-hepatoma drug.

32. The use of one or more antibodies that is/are immunospecific for one or more of the HPIs defined in claim 8 in screening, diagnosis or prognosis of hepatoma in a human subject or for monitoring the effect of an anti-hepatoma drug.

33. The use as claimed in claim 32 wherein the one or more antibodies is/are as defined in claim 26.

34. The use as claimed in any one of claims 29 to 33 which is used in combination with an assay or assessment for Hepatitis B and/or Hepatitis C.

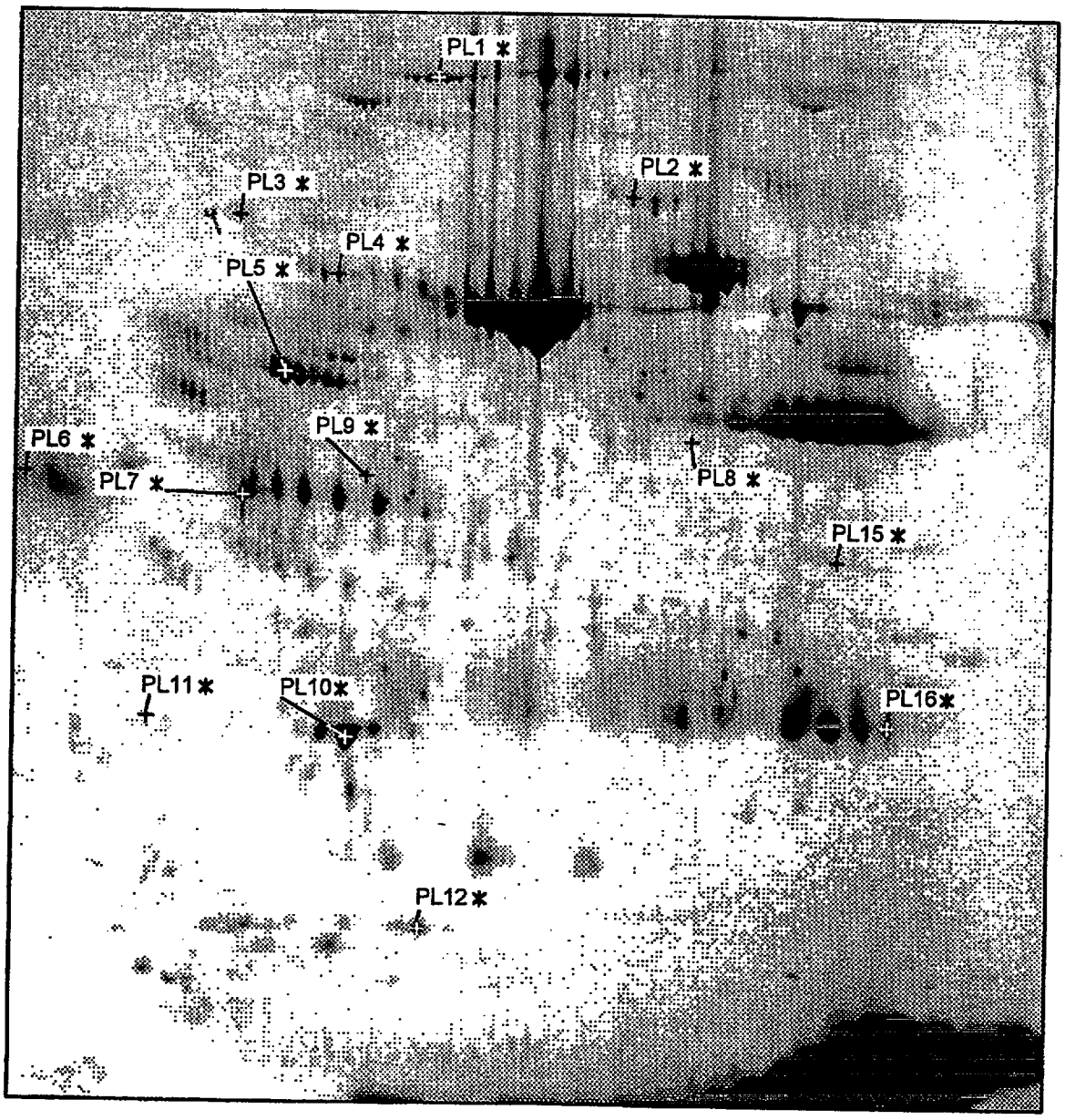


FIG. 1