FEBS Letters 580 (2006) 2216–2226

Hcc-2, a novel mammalian ER thioredoxin that is differentially expressed in hepatocellular carcinoma

Peter Morin Nissom[a], Siaw Ling Lo[b], Jennifer Chi Yi Lo[a], Peh Fern Ong[a], Justin Wee Eng Lim[c], Keli Ou[a,1], Rosa Cynthia Liang[b], Teck Keong Seow[b], Maxey Ching Ming Chung[b,c,*]

[a] *Bioprocessing Technology Institute, #06-01, Centros, 20 Biopolis Way, Singapore 138668, Singapore*
[b] *Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore*
[c] *Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Block MD7, 8 Medical Drive, Singapore 117597, Singapore*

Received 3 December 2005; revised 22 February 2006; accepted 8 March 2006

Available online 20 March 2006

Edited by Veli-Pekka Lehto

Abstract **Hepatocellular carcinoma (HCC) is the most common primary cancer of the liver. Thus there is great interest to identify novel HCC diagnostic markers for early detection of the disease and tumour specific associated proteins as potential therapeutic targets in the treatment of HCC. Currently, we are screening for early biomarkers as well as studying the development of HCC by identifying the differentially expressed proteins of HCC tissues during different stages of disease progression. We have isolated, by reverse transcriptase and polymerase chain reaction (RT-PCR), a 1741 bp cDNA encoding a protein that is differentially expressed in HCC. This novel protein was initially identified by proteome analysis and we designate it as Hcc-2. The protein is upregulated in poorly-differentiated HCC but unchanged in well-differentiated HCC. The full-length transcript encodes a protein of 363 amino acids that has three thioredoxin (Trx) (CGHC) domains and an ER retention signal motif (KDEL). Fluorescence GFP tagging to this protein confirmed that it is localized predominantly to the cytoplasm when expressed in mammalian cells. Protein alignment analysis shows that it is a variant of the TXNDC5 gene, and the human variants found in Genbank all show close similarity in protein sequence. Functionally, it exhibits the anticipated reductase activity in the insulin disulfide reduction assay, but its other biological role in cell function remains to be elucidated. This work demonstrates that an integrated proteomics and genomics approach can be a very powerful means of discovering potential diagnostic and therapeutic protein targets for cancer therapy.**

Keywords: Thioredoxin; Hcc-2; Hepatocellular carcinoma; Protein disulphide isomerase

1. Introduction

Hepatocellular carcinoma (HCC or hepatoma) is the most common primary cancer of the liver which is responsible for approximately one million deaths each year [1]. Persistent viral infection by the hepatitis B or C virus (HBV or HCV) is probably the most important cause of HCC worldwide which contributes to about 80% of HCCs in humans [2].

HCC has a high incidence rate in the underdeveloped and developing countries where the patients are often diagnosed with infiltrative or massive tumours [3]. The incidence of HCC is also increasing in the developed countries [4]. However, in these countries, HCC is mostly diagnosed at an asymptomatic stage by routine ultrasonography since the patient's underlying liver disease has been detected and monitored [3]. For most patients suffering from HCC, long-term survival is rare. The options for treatment are surgery, systematic chemotherapy, loco-regional treatment, and symptomatic relief, and of these, only surgery has the potential to cure. Unfortunately, liver resection is only feasible for 10–15% of the patients as they are often presented late. Thus, there is great interest to identify novel HCC diagnostic markers for early detection of the disease, and tumour specific associated proteins as potential therapeutic targets in the treatment of HCC.

We are currently screening for early biomarkers of HCC as well as studying the development of HCC by identifying the differentially expressed proteins of HCC tissues during different stages of disease progression. In this paper, we report the discovery of a protein, named Hcc-2, which was found to be upregulated in poorly-differentiated HCC but not in the well-differentiated HCC. The identification and characterization of Hcc-2 was performed based on an integrated proteomics, bioinformatics and genomics approach that was described earlier by us for the discovery of Hcc-1 [5].

2. Materials and methods

2.1. Tissue samples

The details of the tissue samples used have been described elsewhere [6]. Briefly, we used hepatitis B virus (HBV)-infected HCC samples and selected matched sample pairs (normal versus tumor) from the same liver for protein expression analyses. Two types of liver tissues were used, those that are derived from well-differentiated and poorly-differentiated HCC. For gene cloning of *Hcc-2*, total RNA was prepared from poorly-differentiated HCC tissues.

*Corresponding author. Fax: +65 6779 1453.
E-mail address: bchcm@nus.edu.sg (M.C.M. Chung).

[1] Present address: Agenica Research Pte Ltd, 11 Hospital Drive, Singapore 169610, Singapore.

Abbreviations: RACE, random amplification of cDNA ends; HCC, hepatocellular carcinoma; PCR, polymerase chain reaction; DMEM, Dulbecco's modified Eagle's medium; DTT, dithiothreitol; CHO, Chinese hamster ovary; TRX, thioredoxin; ORF, open reading frame; RT-PCR, reverse transcriptase and polymerase chain reaction

0014-5793/$32.00
doi:10.1016/j.febslet.2006.03.029

2.2. Proteome analysis and in silico protein assembly

The detailed procedure for proteome analysis by two-dimensional gel electrophoresis (2DE) and matrix-assisted laser desorption-time of flight (MALDI-TOF) MS has previously been published [7]. Selected peptide fragments were subjected to de novo sequencing on a QSTAR tandem hybrid quadrupole-TOF (QqTOF) MS system (MDS SCIEX, Ont., Canada) equipped with a nano-electrospray (nESI) ionization source (MDS Protana, Odense, Denmark). Details of the assembly method can be found in Choong et al. [5]. An additional step was introduced to validate the assembled sequence by using two random peptide sequences from the 5′ and 3′ ends of the assembled putative protein sequence to re-assemble and confirm that there was no more extension to the assembled sequence.

2.3. Functional and structural prediction for novel proteins

Various bioinformatics tools that are available from the internet were used to predict the putative structure and function of the novel proteins [5]. Besides that, PsiPred [8], Interpro [9] and GeneMine/LOOK [10] were also adopted in predicting secondary structure, protein domain and homology modeling.

2.4. Cloning and sequencing of Hcc-2

Initial cloning of *Hcc-2* was performed using degenerate primers targeting 21 bp of the 5′ and 3′ ends. These primers were designed based on the deduced sequence of *Hcc-2* as derived from the amino acid sequence assembly described in Section 2.2. Total RNA was prepared using Trizol reagent (Invitrogen, Carlsbad, CA) as per manufacturer's instructions. First strand cDNA was prepared by reverse transcription on total RNA isolated from tissue samples of poorly-differentiated HCC using an RT-PCR kit (Qiagen, Hilden). Polymerase chain reaction (PCR) was carried out on a PTC-100 Peltier Thermal Cycler (MJ Research). Each 50 μl reaction comprises 3 μl of cDNA (1 ng/μl), high fidelity polymerase, Bioxact (Bioline, UK) and primers at a concentration of 50 nM. Cycling parameters were 95 °C for 5 min followed by 35 cycles of 95 °C for 15 s, 60 °C for 1 min and 72 °C for 1 min. The RT-PCR product was cloned into pCR®2.1–TOPO® (Invitrogen) to give pTOPO-170-138. Upstream (5′UTR) untranslated sequences of *Hcc-2* were determined by random amplification of cDNA ends (RACE) using SMART RACE kit from Clontech (Clontech, Palo Alto, CA). Reactions were performed as recommended by manufacturer. Predicted *Hcc-2* sequence was confirmed by sequencing of numerous positive clones. Upon sequencing and confirmation of the cloned *Hcc-2*, the coding sequence of Hcc-2 was amplified using the primers:

Hcc-2 F 5′GGCGCAGGCCTGCCCATATGTTCACGCACGGGATCCAGAGCGCC Hcc-2 R 5′GGCCGAGATCTCCCGGGCTAAAGTTCGTCTTTCGCTTGGCTCAG and in addition, a *Stu*I and *Nde*I restriction site (underlined in Hcc-2 F) at the start codon (bold) and a *Sma*I and *Bgl*II restriction site (underlined in Hcc-2R) after the stop codon was introduced. The PCR product was cloned into pCR®2.1–TOPO® to give pTOPO-170-138NB.

2.5. Construction of vectors for baculovirus expression of recombinant Hcc-2

A recombinant baculovirus transfer vector was constructed using the Bac-to-Bac system (Life Technologies). The entire open reading frame was excised from pTOPO-170-138NB with *Stu*I and *Xba*I and ligated in frame into pFastbac™ HTB (Invitrogen) in *Stu*I/*Xba*I sites to generate pFastbac-170. The plasmid contained a His-tag at the N-terminus of Hcc-2. This plasmid was then used for transposition in DH10Bac *E. coli* cells. After selection for transposition, the bacmid was isolated and transfected into Sf9 cells using Cellfectin transfection reagent (Invitrogen), and the virus was isolated. All plasmids and bacmids were verified by DNA sequencing.

2.6. Expression and purification of Hcc-2 in Sf9

A shake flask culture of Sf9 insect cells in 500 ml of Grace's media supplemented with 10% fetal bovine serum and 1% pluronic was seeded at a density of 1×10^6/ml and then infected with the virus carrying *Hcc-2*. Cells were harvested 72 h after infection, and the cell pellet was resuspended in 5 ml of homogenization/binding buffer (20 mM Tris–HCl, pH 8.0, 5 mM imidazole, 0.5 M NaCl, Complete™ protease inhibitors without EDTA, Roche Molecular Biochemicals) and the mixture was incubated on ice for 30 min. Hcc-2 was then batch purified using Talon cobalt resin (Clontech) according to manufacturer's protocols. Purified protein was analysed on SDS–polyacrylamide gel, blotted onto PVDF membrane and subjected to Edman N-terminal sequencing on an ABI Procise 494 Protein Sequencer.

2.7. Insulin disulfide reduction assay

Purified Hcc-2 was tested for reductase activity using an insulin disulfide reducing assay [11]. A 5 μl reaction mixture contained 100 mM potassium phosphate, pH 7.0, 0.2 mM EDTA, 1 mg insulin, 1 μg Hcc-2 or recombinant *E. coli* thioredoxin (TRX) (Promega, Madison, WI). The reaction was initiated by addition of 0.33 mM dithiothreitol (DTT). Reactions proceeded at room temperature (22 °C) and were followed by measuring A_{595} on a Genequant spectrophotometer (Amersham Biosciences, Piscataway, NJ).

2.8. Construction and expression of GFP-tagged Hcc-2 in Chinese hamster ovary (CHO) cells

A *Bam*HI/*Eco*RI fragment was obtained from plasmid pTOPO-170-138NB and ligated in frame into pEGFP-C2 (Clontech) in the *Bgl*II/*Eco*RI site to generate construct, pEGFP-170, which contained a GFP at the N-terminus of Hcc-2. To establish stably transfected cell lines, CHO-K1 cells were transfected with plasmid pEGFP-170 using FuGENE 6 reagent (Roche Molecular Biochemicals). As a control, cells were also transfected with pEGFP-C2. After 24 h, cells were trypsinized and plated with different dilutions on 6-wells plates. After 10 days in selection medium containing 800 μg/ml neomycin G418 (Invitrogen), resulting colonies were maintained in 400 μg/ml neomycin for further analysis.

2.8.1. Transfection of CHO cells. CHO-K1 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) medium supplemented with 10% FBS and incubated at 37 °C with 5% CO_2. Transfection was done on 2×10^6 cells in 100 μl of Cell Line Nucleofector™ Solution T (Amaxa Biosystems) with 2 μg of pEFGP-170 or pEGFP-C2 plasmid DNA. The cell suspension was transferred into an Amaxa certified cuvette and cells were electroporated using cell type specific nucleofector program H14 (Amaxa manual). Cells were transferred into 6-wells plate containing 3 ml of pre-warm DMEM with 10% FBS medium. After 24 h of recovery, cells were transferred into selective medium (DMEM supplemented with 10% FBS and 800 μg/ml of G418 (Sigma)). After 3 weeks of selection, transfected cells were adapted to serum free HyQ PF-CHO MPS medium (Gibco) with 800 μg/ml of G418.

2.8.2. Microscopy. CHO-K1 cells transfected with pEGFP-170 fusion protein were seeded onto 6 well culture plates (Nunc Inc., Rochester, NY). At 24 h posttransfection, cells were washed twice with phosphate buffered saline and images were acquired with an Olympus IX70 inverted microscope. GFP was excited with a 450-nm laser, and a 515–565-nm bandpass filter was used for detection of the emission fluorescence. Multiple cells were inspected randomly, and only representative fields were presented in the figures.

2.9. Quantitative real time PCR and RT-PCR analysis

Human multiple tissue cDNA panels (Clontech, Catalog number 636742, 636743) containing first-strand cDNAs from 12 human tissues: heart, brain, placenta, lung, liver, skeletal muscle, kidney pancreas (636742) and spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocyte (636743) were subjected to analysis using quantitative real-time PCR (qRTPCR) on an ABI Prism 7000 Sequence Detector System (Applied Biosystems, Foster City, CA). Each 25 μl qRTPCR reaction comprises 3 μl of cDNA (1 ng/μl), 2× ITAQ SYBR I Master Mix (Biorad) and gene-specific primers at a concentration of 50 nM. Primers for *Hcc-2* (AJ440721) and *Homo sapiens* β-actin (*Actb*) mRNA (NM_001101) were used.

ACTB_F 5′-AGAGATGGCCACGGCTGCTT-3′
ACTB_R 5′-ATTTGCGGTGGACGATGGAG-3′
Hcc-2_F 5′-AGGCCAAGAAGCTGTGAAGT-3′
Hcc-2_R 5′-TCATACAGCCCTTGCTTGAG-3′

All reactions were performed in triplicate. Cycling parameters were 95 °C for 15 min followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Transcript level of Hcc-2 in different tissues was normalised to actin and expressed as ΔC_T, where $\Delta C_T = C_T(\text{Hcc-2}) - C_T$ (Actin). Relative transcript level (RTL) is obtained as $\text{RTL} = 1000 \times 2^{-}(\Delta C_t)$. RT-PCR analysis of transcript levels was performed

on first strand cDNAs generated from normal human liver and HCC tissues. Primers 170-A, 5′-GTGCCCCCGAGCTCAAGCAA-3′ and 170 R, 5′-CTAAAGTTCGTCTTTCGCTTGGCTCAG-3′, were used to generate a 751 bp product from *Hcc-2*. PCR reactions were performed as described in Section 2.4, over 30 cycles.

3. Results

3.1. Proteome analysis and in silico protein assembly

Fig. 1 showed a preparative 2DE map of HCC tumour tissue with Hcc-2 (pI of 5.2 and a relative molecular mass of 55 kDa) indicated by an arrow, while the gel insert showed that it was upregulated in poorly-differentiated HCC tissue. As this protein spot was not identified from the Swiss-Prot or NCBI databases based on its peptide mass fingerprinting from MALDI-TOF data, five of its tryptic peptides were subjected to de novo sequencing (Fig. 2). The amino acid sequences from the fragments were then used to search the EST database by using BLAST (tblastn) [12]. A total of 41 ESTs were found and this was assembled using the CAP3 [13] assembly software into a putative in silico open reading frame (ORF). The assembled ORF was confirmed by carrying out a theoretical tryptic digestion of the protein. Six peptides matching the rest of the significant peaks of the original peptide mass fingerprints of the protein were found (Fig. 2).

3.2. Functional and structural prediction

Using PSI-BLAST [12] on non-redundant database, Hcc-2 has 38% identity match with protein disulfide isomerase precursor. A search on Interpro [9] revealed that Hcc-2 has three

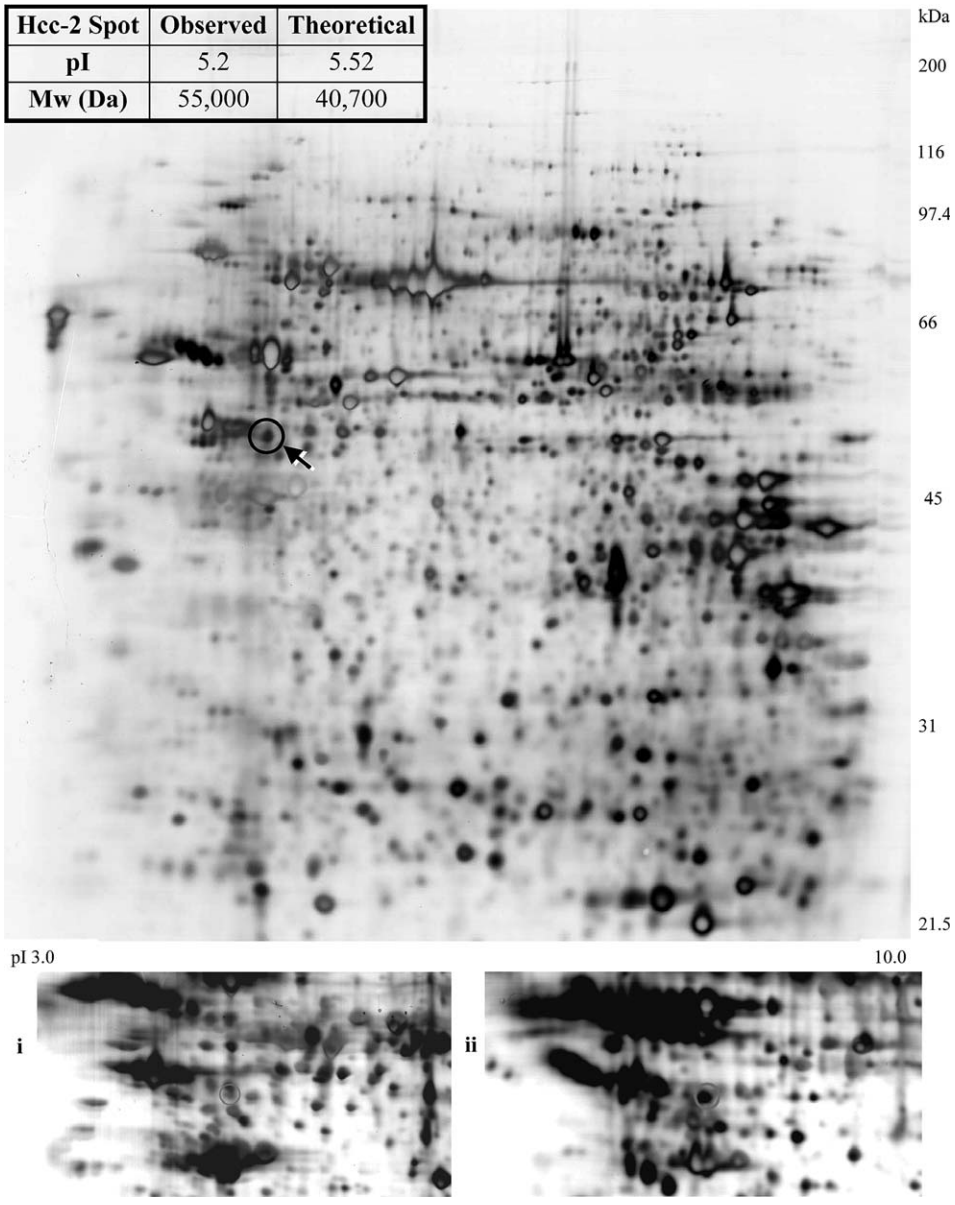

Fig. 1. Preparative 2DE map of poorly differentiated HCC tumour tissue. The novel Hcc-2 protein is circled, and its observed and theoretical isoelectric points and molecular weights are listed in the table as an insert. The gel inset represents the expression profile of Hcc-2 in the: (i) non-tumorous; and (ii) tumorous region of the liver.

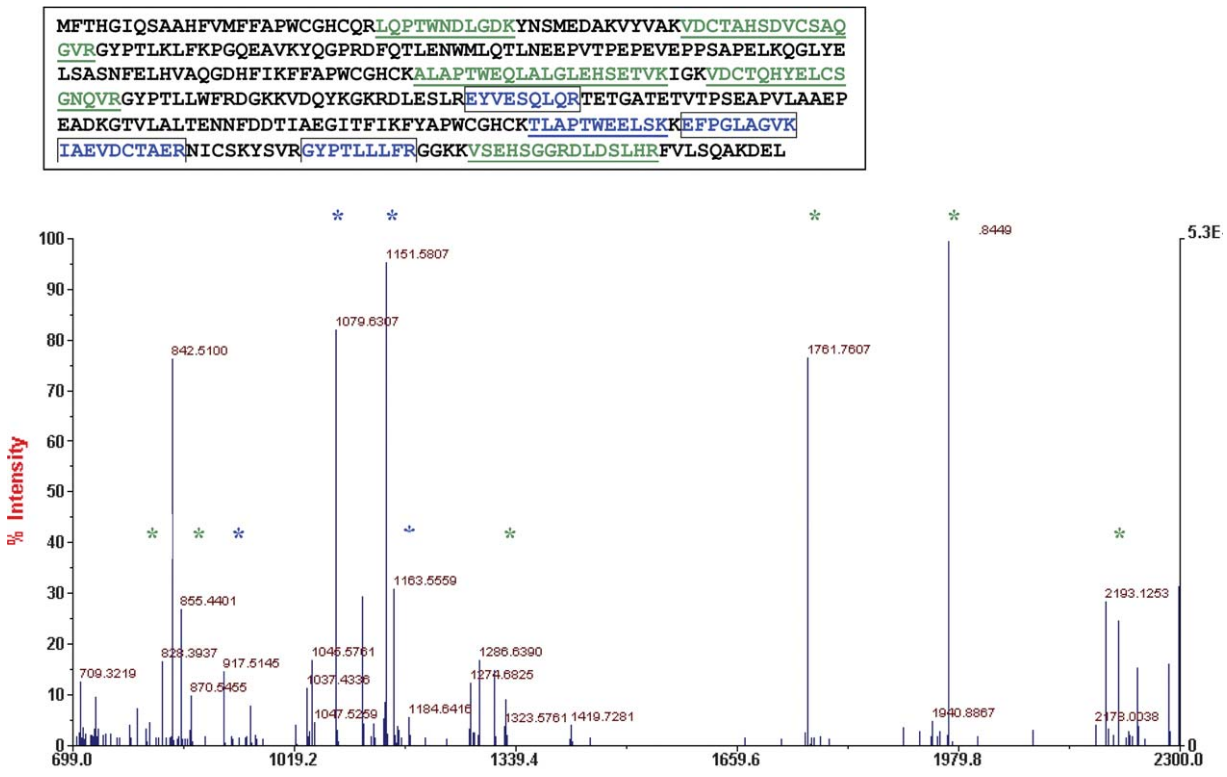

Fig. 2. MALDI spectrum and amino acid sequence (inset) of Hcc-2. The four peptides indicated by blue asterisks were subjected to de novo sequencing. The sequences of the peptides are boxed and in blue (inset), and one additional peptide in blue and underlined (inset), were utilised in the in silico assembly of the protein. The peptides indicated by green asterisks were the six additional peptides with masses that matched the underlined and green-coloured sequences (inset) of the assembled sequence.

Trx-like domains (sequence signature CXXC). Fig. 3 shows the location of the domains. The protein was predicted by PSORT [14] to be in the cytoplasm and it was found to have an endoplasmic reticulum retention motif (KDEL) in the C-terminus. The protein was found to have no predicted transmembrane segment (PredictProtein) [15] and no secretory signal (SignalP) [16]. Each of the three Trx-like domains of Hcc-2 was found to have a 40% identity match with the **a** domain of protein disulfide isomerase, PDI [17] but not the **b** domain [18]. The **a** domain of PDI shows significant sequence identity to Trx which has the characteristic α/β fold with the structure of βαβαβαββα [19,20]. Fig. 4 shows the result for the homology modeling of PDI's **a** domain and the first domain of Hcc-2 with the WCXXC active-site motif highlighted in yellow. All the three domains of Hcc-2 contain the active site with the sequence motif of CGHC which had been shown to be involved directly in thiol–disulfide exchange reactions [17]. It is interesting to note that Hcc-2 is a novel protein possessing three **a** domains, which is different from all the earlier known PDI-related proteins [19,21,22].

3.3. Cloning of human Hcc-2 cDNA

We used degenerate oligonucleotide primers to clone the *Hcc-2* gene from first strand cDNA derived from poorly-differentiated HCC tissues. RACE was performed to obtain information on the 5′ and 3′ untranslated regions of the gene. The subsequent PCR products were sequenced and the fully assembled DNA sequence were found to agree with the EST assembled sequence of the protein. Fig. 5 shows the fully assembled DNA sequence of *Hcc-2* after confirmation by DNA sequencing. The cDNA contained an ORF of 1092 bp. This ORF encodes a protein of 363 amino acids with a calculated molecular mass of 40.71 kDa. The amino acid sequence of the encoded protein agrees with the predicted sequence

```
1    MFTHGIQSAAHFVMFFAPWCGHCQRLQPTWNDLGDKYNSMEDAKVYVAKV
51   DCTAHSDVCSAQGVRGYPTLKLFKPGQEAVKYQGPRDFQTLENWMLQTLN
101  EEPVTPEPEVEPPSAPELKQGLYELSASNFELHVAQGDHFIKFFAPWCGH
151  CKALAPTWEQLALGLEHSETVKIGKVDCTQHYELCSGNQVRGYPTLLWFR
201  DGKKVDQYKGKRDLESLREYVESQLQRTETGATETVTPSEAPVLAAEPEA
251  DKGTVLALTENNFDDTIAEGITFIKFYAPWCGHCKTLAPTWEELSKKEFP
301  GLAGVKIAEVDCTAERNICSKYSVRGYPTLLLFRGGKKVSEHSGGRDLDS
351  LHRFVLSQAKDEL
```

Fig. 3. Predicted amino acid sequences of Hcc-2. Boxed peptides are the signature peptide for Trx (CXXC) defined by Interpro. The ER retention motif, KDEL, located at the C-terminal of the protein, is in bold.

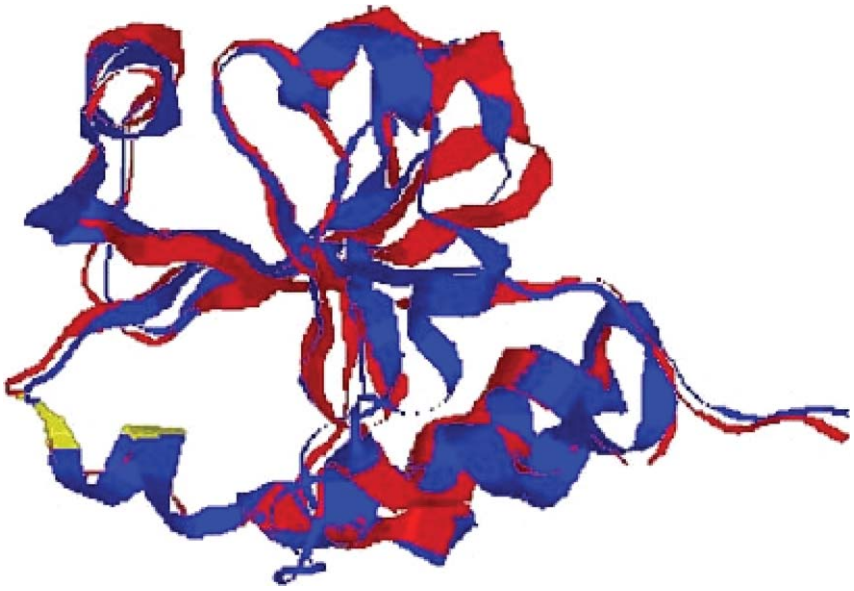

Fig. 4. Homology modeling of the overlapping regions of Hcc-2 (aa 1–125) and the PDB structure of human protein disulfide isomerase (1 MEK). The Hcc-2 structure is in blue and the 1 MEK structure is in red, while Cys20 and Cys23 in the Trx active site are in yellow.

(Fig. 3). This sequence has been deposited in Genbank under accession number AJ440721. We used this information to design primers spanning the putative ORF of the protein and cloned the ORF. PCR experiments generated a single 1 kb band on agarose gel electrophoresis (results not shown). PCR product was cloned into TOPO vectors and numerous positive clones were sequenced. This ORF was used for expression of Hcc-2 in baculovirus Sf9 cells.

3.4. Expression of recombinant Hcc-2

To confirm the identity of this protein as well as the predicted Trx-like activity associated with the CGHC motifs, we expressed Hcc-2 in baculovirus as a HIS-tagged protein. The full length Hcc-2 protein has a KDEL motif for ER retention at the C-terminus, hence we tagged a short HIS tag to the N-terminus. The purified protein ran as a single band and migrated at the range of 43–56 kDa on SDS–PAGE, higher than the predicted molecular mass of 40 kDa (Fig. 6). Edman N-terminal sequencing confirmed the first 19 amino acids as MFTHGIQSAAHFVMFFAPW, exactly as predicted. This confirmed that we had isolated and cloned the *Hcc-2* gene and expressed its product.

3.5. Reductase activity of Hcc-2

Hcc-2 was identified as a putative Trx protein because of the presence of the active site sequence CGHC. This protein is unique in that it has three such domains (Fig. 3). In order to determine if indeed these domains confer a Trx-like reducing activity, we performed an insulin disulfide reduction assay [11] using recombinant Hcc-2 which we expressed and purified as described earlier. The negative control, which is DTT and insulin did not exhibit any change in absorbance, indicating no activity. With the positive control, which is purified recombinant *E.coli* Trx, a rapid increase in turbidity was observed. Most importantly, the purified HIS-tagged Hcc-2 also exhibited an increase in turbidity indicating that this protein does have Trx-like reducing activity (Fig. 7). It is also apparent that it is a poorer reductase compared to *E. coli* thioredoxin. This is probably due to the difference in redox potentials as a result of the difference in the nature of the two intervening residues of the reactive CXXC sequence [19].

3.6. Localization of Hcc-2

Sequence analysis revealed a KDEL motif at the C-terminus of Hcc-2. This KDEL motif is known to signal the retention of proteins in the ER [23]. The presence of this motif in Hcc-2 indicates that it might be localized to the cytoplasm/ER. To confirm this, we constructed vectors for expression of Hcc-2 in mammalian cells. To facilitate single cell cloning of positive transfectants and subsequent localization of Hcc-2, a GFP was tagged to the N-terminus of Hcc-2. The expression of the GFP-*Hcc-2* gene is under the control of the constitutive cytomegalovirus promoter. Control plasmid expressing only GFP was also made. CHO-K1 cell line was transfected with the constructs. Positive fluorescing cells were examined under microscope. Control cells show fluorescence throughout the cells (Fig. 8A), while green fluorescence in GFP-Hcc-2 cells were enhanced in the cytoplasm but not in the nucleus (Fig. 8B). This observation is consistent with the Hcc-2 protein being retained in the cytoplasm due to the KDEL signal.

3.7. Tissue distribution of Hcc-2 mRNA

We used qRTPCR to study the distribution of the *Hcc-2* transcript in human tissues and cultured human cell lines. We designed real-time PCR primers to *Hcc-2*. The primers corresponded to 468–487 bp and 595–614 bp of *Hcc-2* (Fig. 5), respectively. Specificity of primer pair used was checked by sequence alignments and by analysis of dissociation curves generated by the real time analysis. Human multiple tissue cDNAs were used as templates. Fig. 9 shows the transcript level of *Hcc-2* in the different tissues. *Hcc-2* is ubiquitously expressed in all 16 tissue types. High expression was detected in the pancreas.

```
   1 CCTGGTGGCGAATTCGGCACGAGGAGCCCCGCCGCGATGCCCGCGCGCCCAGGACGCCTC
  61 CTCCTTCTGCTGGCCCGGCCGGCGGCCCTGACTGCGCTGCTGCTGCTGCTGCTGGGCCAT
 121 GGCGGCGGCGGGCGCTGGGGCGCTTTTTCCCAGGAGGCGGCGGCGGCGGCGGACGGG
 181 CCCCCCGCGGCAGACGGCGAGGACGGTCAGGACCCGCACAGCAAGCACCTGTACACGGCC
       M  F  T  H  G  I  Q  S  A  A  H  F  V  M  F  F  A  P  W
 241 GACATGTTCACGCACGGGATCCAGAGCGCCGCGCACTTCGTCATGTTCTTCGCGCCCTGG
     C  G  H  C  Q  R  L  Q  P  T  W  N  D  L  G  D  K  Y  N  S
 301 TGTGGACACTGCCAGCGGCTGCAGCCGACTTGGAATGACCTGGGAGACAAATACAACAGC
     M  E  D  A  K  V  Y  V  A  K  V  D  C  T  A  H  S  D  V  C
 361 ATGGAAGATGCCAAAGTCTATGTGGCTAAAGTGGACTGCACGGCCCACTCCGACGTGTGC
     S  A  Q  G  V  R  G  Y  P  T  L  K  L  F  K  P  G  Q  E  A
 421 TCCGCCCAGGGGGTGCGAGGATACCCCACCTTAAAGCTTTTCAAGCCAGGCCAAGAAGCT
     V  K  Y  Q  G  P  R  D  F  Q  T  L  E  N  W  M  L  Q  T  L
 481 GTGAAGTACCAGGGTCCTCGGGACTTCCAGACACTGGAAAACTGGATGCTGCAGACACTG
     N  E  E  P  V  T  P  E  P  E  V  E  P  P  S  A  P  E  L  K
 541 AACGAGGAGCCAGTGACACCAGAGCCGGAAGTGGAACCGCCCAGTGCCCCCGAGCTCAAG
     Q  G  L  Y  E  L  S  A  S  N  F  E  L  H  V  A  Q  G  D  H
 601 CAAGGGCTGTATGAGCTCTCAGCAAGCAACTTTGAGCTGCACGTTGCACAAGGCGACCAC
     F  I  K  F  F  A  P  W  C  G  H  C  K  A  L  A  P  T  W  E
 661 TTTATCAAGTTCTTCGCTCCGTGGTGTGGTCACTGCAAAGCCCTGGCTCCAACCTGGGAG
     Q  L  A  L  G  L  E  H  S  E  T  V  K  I  G  K  V  D  C  T
 721 CAGCTGGCTCTGGGCCTTGAACATTCCGAAACTGTCAAGATTGGCAAGGTTGATTGTACA
     Q  H  Y  E  L  C  S  G  N  Q  V  R  G  Y  P  T  L  L  W  F
 781 CAGCACTATGAACTCTGCTCCGGAAACCAGGTTCGTGGCTATCCCACTCTTCTCTGGTTC
     R  D  G  K  K  V  D  Q  Y  K  G  K  R  D  L  E  S  L  R  E
 841 CGAGATGGGAAAAAGGTGGATCAGTACAAGGGAAAGCGGGATTTGGAGTCACTGAGGGAG
     Y  V  E  S  Q  L  Q  R  T  E  T  G  A  T  E  T  V  T  P  S
 901 TACGTGGAGTCGCAGCTGCAGCGCACAGAGACTGGAGCGACGGAGACCGTCACGCCCTCA
     E  A  P  V  L  A  A  E  P  E  A  D  K  G  T  V  L  A  L  T
 961 GAGGCCCCGGTGCTGGCAGCTGAGCCCGAGGCTGACAAGGGCACTGTGTTGGCACTCACT
     E  N  N  F  D  D  T  I  A  E  G  I  T  F  I  K  F  Y  A  P
1021 GAAAATAACTTCGATGACACCATTGCAGAAGGAATAACCTTCATCAAGTTTTATGCTCCA
     W  C  G  H  C  K  T  L  A  P  T  W  E  E  L  S  K  K  E  F
1081 TGGTGTGGTCATTGTAAGACTCTGGCTCCTACTTGGGAGGAACTCTCTAAAAAGGAATTC
     P  G  L  A  G  V  K  I  A  E  V  D  C  T  A  E  R  N  I  C
1141 CCTGGTCTGGCGGGGGTCAAGATCGCCGAAGTAGACTGCACTGCTGAACGGAATATCTGC
     S  K  Y  S  V  R  G  Y  P  T  L  L  L  F  R  G  G  K  K  V
1201 AGCAAGTATTCGGTACGAGGCTACCCCACGTTATTGCTTTTCCGAGGAGGGAAGAAAGTC
     S  E  H  S  G  G  R  D  L  D  S  L  H  R  F  V  L  S  Q  A
1261 AGTGAGCACAGTGGAGGCAGAGACCTTGACTCGTTACACCGCTTTGTCCTGAGCCAAGCG
     K  D  E  L  *
1321 AAAGACGAACTTTAGGAACACAGTTGGAGGTCACCTCTCCTGCCCAGCTCCCGCACCCTG
1381 CGTTTAGGAGTTCAGTCCCACAGAGGCCACTGGGTTCCCAGTGGTGGCTGTTCAGAAAGC
1441 AGAACATACTAAGCGTGAGGTATCTTCTTTGTGTGTGTGTTTTCCAAGCCAACACACTCT
1501 ACAGATTCTTTATTAAGTTAAGTTTCTCTAAGTAAATGTGTAACTCATGGTCACTGTGTA
1561 AACATTTTCAGTGGCGATATATCCCCTTTGACCTTCTCTTGATGAAATTTACATGGTTTC
1621 CTTTGAGACTAAAATAGCGTTGAGGGAAATGAAATTGCTGGACTATTTGTGGCTCCTGAG
1681 TTGAGTGATTTTGGTGAAAGAAAGCACATCCAAAGCATAGTTTACCTGCCCACGAGTTCT
1741 GGAAAG
```

Fig. 5. Nucleic acid base sequence (and corresponding translation protein sequence) of *Hcc-2*. The sequence of cDNA clone *Hcc-2* (GenBank accession no. AJ440721) is shown. Nucleic acids in the open reading frame (ORF) and the noncoding regions are presented in uppercase and italicised letters. The thioredoxin (CGHC) domains are bold and underlined.

3.8. RT-PCR analysis of Hcc-2 mRNA expression between normal and HCC tissues

To determine the Hcc-2 mRNA expression level in normal liver tissues as compared to HCC tissues, we perform RT-PCR on the total RNA generated from normal human liver and HCC tissues. A primer pair that generates a 751bp product from the *Hcc-2* gene was designed. Primers targeting β-actin (Actb-F and Actb-R) was used as loading control. As Fig. 10 shows, the PCR product for *Hcc-2*, from HCC tissues is more than that from normal tissues indicating a higher level of Hcc-2 mRNA in the HCC tissues. An equal amount of cDNA had been used for both samples as shown by the loading control, actin. This result validates the protein expression pattern shown by proteomic analysis.

4. Discussion

In this work, we used an integrated approach of proteomics, bioinformatics and genomics to clone and characterize a novel

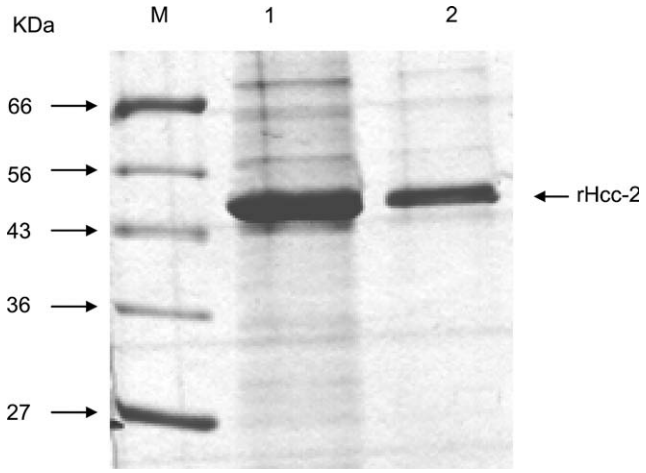

Fig. 6. SDS–PAGE of recombinant Hcc-2 expressed in baculovirus infected Sf9 insect cells. *Hcc-2* was cloned into a baculovirus protein expression system vector and expressed in Sf9 insect cells. Cells were harvested at 72 h , lysed and the protein was batch purified on Talon cobalt resins and analysed on a 12% SDS–polyacrylamide gel. The recombinant protein exhibited a molecular weight of between 43 and 56 kDa, higher than the predicted size of 40.71 kDa. (M) Broad range molecular weight marker (Biorad). (1) rHCC-2 bound to Talon resin, after washing. (2) Purified rHcc-2 eluted from Talon resins.

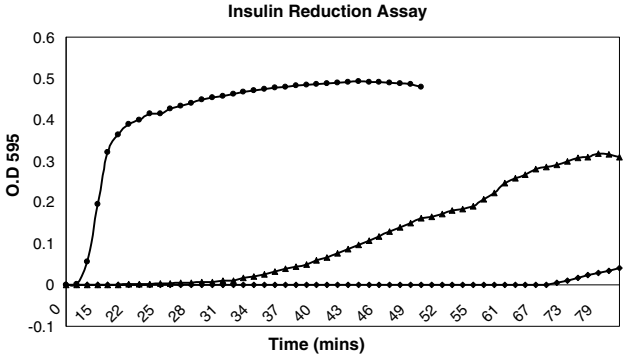

Fig. 7. Reductase activity of Hcc-2. Baculovirus expressed Hcc-2 and *E. coli* TRX proteins were incubated with insulin and the reduction of insulin disulfide bonds was measured by monitoring absorbance at 595 nm, every 2 min. ●, 0.2 mg/ml *E. coli* TRX; ▲, 0.2 mg/ml rHcc-2; ♦, Blank control. The experiments were carried out in duplicates.

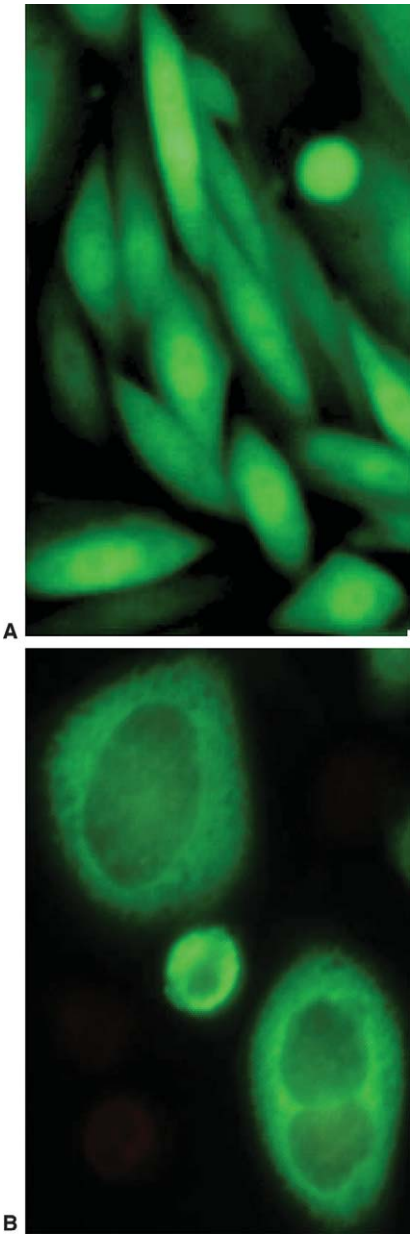

Fig. 8. Localization of Hcc-2 by confocal microscopy. Transfected CHO-K1 cell fluorescence images. Fluorescent images showing CHO-K1 cells transfected with pEGFP-C2 (A) and pEGFP-170 (B). After transient expression, cells were visualized with an Olympus IX70 inverted microscope. These data are representative of two separate experiments. Images show that the control, which is GFP, is distributed throughout the cell (A) whereas Hcc-2, which is tagged to GFP is localized predominantly to cytoplasm (B). The Hcc-2 image has been enlarged 400× to show the nuclei.

Trx-related protein, Hcc-2. The *Hcc-2* gene is ubiquitously expressed in normal human tissues (Fig. 9) and is upregulated in poorly differentiated HCC tissues (Fig. 10). The protein contains three redox active site and an ER retention signal sequence. We designate this protein, hepatocellular carcinoma 2 (Hcc-2), as it was the second upregulated protein identified by proteomics based on our studies on hepatocellular carcinoma [5,7]. Interestingly, the upregulation of Hcc-2 is characteristic of the expression pattern of Trx-like proteins in some cancers. For example, Trxs are known to be overexpressed in a variety of human primary tumours compared to levels in the corresponding normal tissues [24,25]. In patients with HCC, the plasma and serum levels of Trx have been reported to be elevated almost 2-fold and to decrease with surgical removal of the tumour [26].

The Hcc-2 protein migrated consistently at 55 kDa in 2DE despite that the protein has only 363 amino acids with a theoretical molecular mass of 40.71 kDa (Fig. 1A). The recombinant Hcc-2 protein has also migrated as greater than 43 kDa in SDS–PAGE (Fig. 6). N-terminal amino acid sequencing confirmed that the band at 55 kDa contained the same N-terminal sequence as the Hcc-2 protein. The deviation in migration may be due to the fact that Hcc-2 is a very hydrophilic protein since it consists of only 34.7% hydrophobic amino acids based on the Kyte–Doolittle plot and GRAVY (grand average of hydropathy) index of −0.447. This high surface

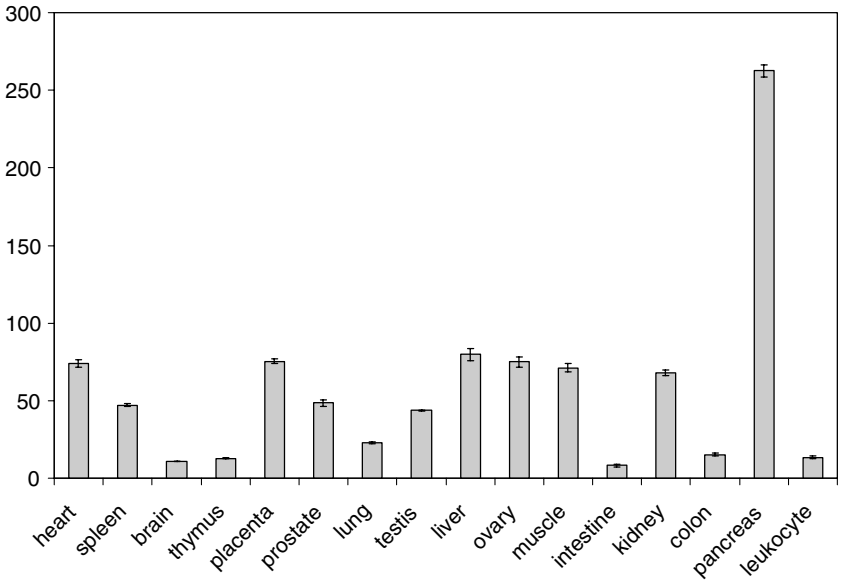

Fig. 9. Tissue distribution of Hcc-2 analysed by real time PCR. Relative transcript levels (RTL) of *Hcc-2* expressed as mean RTL ± S.E.M. Values were obtained from three technical replicates. The threshold cycle or the C_T value is the cycle at which a significant increase in the signal associated with an exponential growth of PCR product during the log-linear phase is detected. The higher the initial amount of transcript, the sooner accumulated product is detected in the PCR process, and the lower the C_T value. The relative levels of *Hcc-2* mRNA were determined by real time PCR in the tissues indicated on the bottom of the figure using β-actin mRNA levels as a reference and normalization standard. The data shown were obtained from three samples using the same preparations of tissue RNAs.

charge in Hcc-2 might have interfered with the binding of SDS to the protein in SDS–PAGE.

The predicted Hcc-2 protein (Genbank CAD29430.1) was subsequently found to overlap with other human homologues deposited in Genbank (GenBank accession no. BAC11526.1, AAQ89009.1, Q8NBS9, CAH56286.1, AAR99514.1, AAH01199.1, CAD39084.1, AAH52310.1). They all possess the exact active site sequence (CGHC) as Hcc-2, but some of them have different lengths in the N-terminal segments (Fig. 11). As a result, the last three proteins in the list above (AAH01199.1, CAD39084.1, AAH52310.1) have only two Trx-like domains, as compared with the other six proteins (BAC11526.1, AAQ89009.1, Q8NBS9, CAH56286.1, AAR99514.1 and Hcc-2) which have three Trx-like domains.

These homologues belonged to the TXNDC5 (thioredoxin domain containing 5) gene family of proteins, which consists of two variants, with variant 1 being the predominant transcript. Hcc-2 belongs to variant 1 but it is short of 69 amino acids at the N-terminus. On the basis of the Kyte–Doolittle hydropathy plot, there is a high possibility that a signal peptide and a transmembrane region may exist at this additional 69 amino acids (result not shown). Hitherto, all these homologous proteins do not have any biological functions assigned to them yet.

As indicated above and in the results section earlier, one interesting novel feature of Hcc-2 is that it contains three Trx domains. These Trx domains have a similar active site sequence (CGHC) as the **a** domain of PDI (Fig. 4). It has been well-established that proteins in the PDI superfamily consists of 2**a** (represented by **a** and **a**′) and 2**b** (represented by **b** and **b**′) domains, and a C-terminal acidic extension segment, and that only the **a** and **a**′ domains of PDI possess the Trx-like

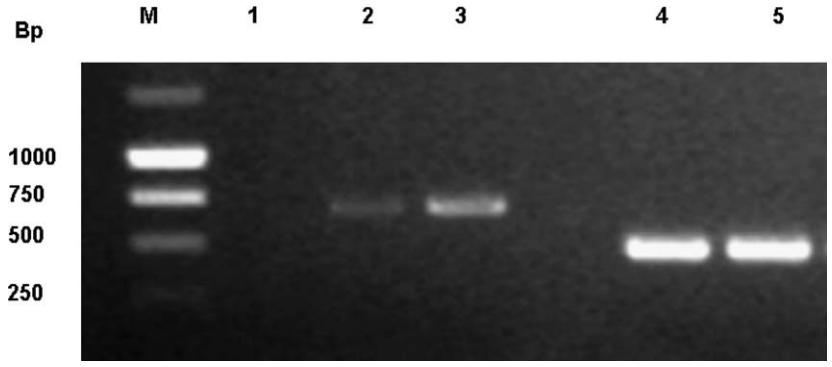

Fig. 10. Comparison of the mRNA expression level of Hcc-2 by RT-PCR analysis. 3 ng of first strand cDNA prepared from normal human liver and HCC tissues was used as template for PCR using primers to *Hcc-2* and β-actin. PCR reaction was carried out over 30 cycles. The products were separated by agarose gel electrophoresis and photographed using a gel documentation system (BioRad). *Hcc-2* primers generate a 751 bp product while β-actin primers produce a 500 bp product. Lanes M, 1 kb marker (Promega); 1, no template control; 2, Hcc-2 – normal liver tissue cDNA; 3, Hcc-2 – HCC tissue cDNA; 4, actin – normal liver tissue cDNA; 5, actin – HCC tissue cDNA.

```
BAC11526.1          MPARPGRLLPLLARPAALTALLLLLLGHGGGGRWGARAQEAAAAAADGPP 50
AAQ89009.1          MPARPGRLLPLLARPAALTALLLLLLGHGGGGRWGARAQEAAAAAADGPP 50
Q8NBS9              MPARPGRLLPLLARPAALTALLLLLLGHGGGGRWGARAQEAAAAAADGPP 50
CAH56286.1          ----------------------------------------AAAAAADGPP 10
AAR99514.1          --------------------------------------------------
CAD29430.1 (Hcc-2)  --------------------------------------------------
AAH01199.1          --------------------------------------------------
CAD39084.1          --------------------------------------------------
AAH52310.1          --------------------------------------------------

BAC11526.1          AADGEDGQDPHSKHLYTADMFTHGIQSAAHFVMFFAPWCGHCQRLQPTWN 100
AAQ89009.1          AADGEDGQDPHSKHLYTADMFTHGIQSAAHFVMFFAPWCGHCQRLQPTWN 100
Q8NBS9              AADGEDGQDPHSKHLYTADMFTHGIQSAAHFVMFFAPWCGHCQRLQPTWN 100
CAH56286.1          AADGEDGQDPHSKHLYTADMFTHGIQSAAHFVMFFAPWCGHCQRLQPTWN 60
AAR99514.1          ------------------MFTHGIQSAAHFVMFFAPWCGHCQRLQPTWN 31
CAD29430.1 (Hcc-2)  ------------------MFTHGIQSAAHFVMFFAPWCGHCQRLQPTWN 31
AAH01199.1          --------------------------------------------------
CAD39084.1          --------------------------------------------------
AAH52310.1          ------------------MTQSVDSNRGNRN--EKRCGHCQRLQPTWN 28

BAC11526.1          DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 150
AAQ89009.1          DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 150
Q8NBS9              DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 150
CAH56286.1          DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 110
AAR99514.1          DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 81
CAD29430.1 (Hcc-2)  DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 81
AAH01199.1          --------MEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 42
CAD39084.1          --------------------------------------------------
AAH52310.1          DLGDKYNSMEDAKVYVAKVDCTAHSDVCSAQGVRGYPTLKLFKPGQEAVK 78

BAC11526.1          YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 200
AAQ89009.1          YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 200
Q8NBS9              YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 200
CAH56286.1          YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 160
AAR99514.1          YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 131
CAD29430.1 (Hcc-2)  YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 131
AAH01199.1          YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 92
CAD39084.1          ------------------------------------QGLYELSASNFE 12
AAH52310.1          YQGPRDFQTLENWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFE 128
                                                        ************

BAC11526.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 250
AAQ89009.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 250
Q8NBS9              LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 250
CAH56286.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 210
AAR99514.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 181
CAD29430.1 (Hcc-2)  LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 181
AAH01199.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 142
CAD39084.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 62
AAH52310.1          LHVAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQH 178
                    **************************************************
```

Fig. 11. Clustalw output of Hcc-2 with the human variants of TXNDC5 deposited in Genbank. Hcc-2 (CAD29430.1) was aligned with human variants of TXNDC5. The conserved active site is in bold. Identical amino acids between the three proteins are marked with asterisks. The numbers next to the sequences represent NCBI Genbank accession number.

reducing activity [19,21]. The biochemical function of Hcc-2 was confirmed by the insulin disulfide reducing assay (Fig. 8) which detects the reduction of the two interchain disulfide bonds of insulin. In a recent paper, Knoblach et al. [27] identified ERp46 as a new member of the thioredoxin family of endoplasmic reticulum proteins. ERp46, having a sequence homology of 77.5% with Hcc-2, and consisting of 417 amino acids, also contained three CGHC domains. Functional studies using a complementation assay in the yeast *S. cerevisiae* demonstrated that it had a PDI-like activity.

Besides Hcc-2 and ERp46, the other proteins known to-date that have three CXXC domains are ERp72 and hPDIR (human protein–disulfide isomerase-related) protein. However, hPDIR protein has a domain structure of **b**–**a**–**a**′–**a**^{o}, while ERp72's domain structure is **c**–**a**^{o}–**a**–**b**–**b**′–**a**′. The **a** domain of ERp72 is similar to PDI, i.e., it consists of an active site sequence of CGHC. ERp72 is an abundant ER resident protein and it catalyzes the reduction, oxidation and isomerization of disulfide-bonded proteins in vitro [28]. It is interesting to note that the isomerase activity of ERp72 is less than PDI [28], considering that PDI has only two CGHC domains. This suggests that ERp72 may have possible multifunctional role other than those being involved in disulfide bond isomerization. For example, ERp72 did not exhibit peptide-binding capacity [29], but in conjunction with BiP, they co-precipitated over expressed human chorionic gonadotropin

```
BAC11526.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 300
AAQ89009.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 300
Q8NBS9               YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 300
CAH56286.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 260
AAR99514.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 231
CAD29430.1 (Hcc-2)   YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 231
AAH01199.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 192
CAD39084.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 112
AAH52310.1           YELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETG 228
                     ************************************************

BAC11526.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 350
AAQ89009.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 350
Q8NBS9               ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 350
CAH56286.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 310
AAR99514.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 281
CAD29430.1 (Hcc-2)   ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 281
AAH01199.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 242
CAD39084.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 162
AAH52310.1           ATETVTPSEAPVLAAEPEADKGTVLALTENNFDDTIAEGITFIKFYAPWC 278
                     ************************************************

BAC11526.1           GHCRTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 400
AAQ89009.1           GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 400
Q8NBS9               GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 400
CAH56286.1           GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 360
AAR99514.1           GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 331
CAD29430.1 (Hcc-2)   GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 331
AAH01199.1           GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 292
CAD39084.1           GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 212
AAH52310.1           GHCKTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICSKYSVRGYPTLL 328
                     ************************************************

BAC11526.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 432
AAQ89009.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 432
Q8NBS9               LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 432
CAH56286.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 392
AAR99514.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 363
CAD29430.1 (Hcc-2)   LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 363
AAH01199.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 324
CAD39084.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 244
AAH52310.1           LFRGGKKVSEHSGGRDLDSLHRFVLSQAKDEL 360
                     ********************************
```

Fig. 11 (*continued*)

b subunit [30]. Besides that, when ERp72 is in association with the molecular chaperones, PDI, BiP and GRp94, it has been shown to interact in vitro with denatured proteins [31] and apolipoprotein B [32]. This was also observed in vivo for thyroglobulin and thrombospondin [33]. In addition, ERp72 was shown to co-precipitate with laminin during differentiation of mouse F9 teratocarcinoma cells [34]. On the other hand, hPDIR protein has three different **a** domains with the following active site sequences: CSMC, CGHC, CPHC, respectively. It has been shown recently that it has isomerase and chaperone activities and that its expression is stress-inducible [35]. Furthermore, mutation studies showed that the three different Trx domains of hPDIR contribute to its isomerase activity with a rank order of CGHC > CPHC > CSMC. However, interestingly, both its isomerase and chaperone activities were found to be lower than those of PDI [35].

In summary, we have isolated and partially characterized a novel protein, Hcc-2, that was found to be upregulated in human hepatocellular carcinoma based on proteome analysis of liver HCC tissues. In addition to its disulfide isomerase function, Hcc-2, similar to the other known proteins with three active site CGHC domains (Fig. 11, ERp46, ERp72 and hPDIR protein) may have other functions that are unknown at present. For example it may act via complex formation in a manner similar to ERp72. Thus Hcc-2 could be involved in ER stress response by forming complexes with other molecular chaperones. This will form the basis of our future studies on this protein with the aim to unravel its role in HCC and tumour development. This work demonstrates that an integrated proteomics and genomics approach can be a very powerful means of discovering potential diagnostic and therapeutic protein targets for cancer therapy.

Acknowledgements: We thank Siow Qi Chang and Hong Guan Lim for excellent support in cell culture experiments, Lu Zheng for MS verification, C-terminal and N-terminal sequencing, Ally Lau and Gek San Tan for work done in 2-DE and Shao-En Ong for guidance in in silico sequence assembly. This work was generously supported by funding from the Agency for Science, Technology, and Research (A*STAR).

References

[1] Schafer, D.F. and Sorrell, M.F. (1999) Hepatocellular carcinoma. Lancet 353, 1253–1257.

[2] Thorgeirsson, S.S. and Grisham, J.W. (2002) Molecular pathogenesis of human hepatocellular carcinoma. Nat. Genet. 31, 339–346.
[3] Llovet, J.M. and Beaugrand, M. (2003) Hepatocellular carcinoma: present status and future prospects. J. Hepatol. 38 (Suppl. 1), S136–S149.
[4] McGlynn, K.A., Tsao, L., Hsing, A.W., Devesa, S.S. and Fraumeni Jr., J.F. (2001) International trends and patterns of primary liver cancer. Int. J. Cancer 94, 290–296.
[5] Choong, M.L., Tan, L.K., Lo, S.L., Ren, E.C., Ou, K., Ong, S.E., Liang, R.C., Seow, T.K. and Chung, M.C. (2001) An integrated approach in the discovery and characterization of a novel nuclear protein over-expressed in liver and pancreatic tumors. FEBS Lett. 496, 109–116.
[6] Liang, C.R., Leow, C.K., Neo, J.C., Tan, G.S., Lo, S.L., Lim, J.W., Seow, T.K., Lai, P.B. and Chung, M.C. (2005) Proteome analysis of human hepatocellular carcinoma tissues by two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics 5, 2258–2271.
[7] Ou, K., Seow, T.K., Liang, R.C., Ong, S.E. and Chung, M.C. (2001) Proteome analysis of a human hepatocellular carcinoma cell line, HCC-M: an update. Electrophoresis 22, 2804–2811.
[8] McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405.
[9] Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R.R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S.E., Pagni, M., Peyruc, D., Ponting, C.P., Selengut, J.D., Servant, F., Sigrist, C.J., Vaughan, R. and Zdobnov, E.M. (2003) The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res. 31, 315–318.
[10] Lee, C. and Irizarry, K. (2001) The GeneMine System for genome/proteome annotation and collaborative data mining. IBM Systems J. 40, 592–603.
[11] Holmgren, A. (1979) Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. J. Biol. Chem. 254, 9627–9632.
[12] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[13] Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. Genome Res. 9, 868–877.
[14] Nakai, K. and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24, 34–36.
[15] Rost, B., Yachdav, G. and Liu, J. (2004) The PredictProtein server. Nucleic Acids Res. 32, W321–W326.
[16] Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795.
[17] Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M. and Creighton, T.E. (1997) The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. Curr. Biol. 7, 239–245.
[18] Kemmink, J., Dijkstra, K., Mariani, M., Scheek, R.M., Penka, E., Nilges, M. and Darby, N.J. (1999) The structure in solution of the b domain of protein disulfide isomerase. J. Biomol. NMR 13, 357–368.
[19] Ferrari, D.M. and Soling, H.D. (1999) The protein disulphide-isomerase family: unravelling a string of folds. Biochem. J. 339 (Pt 1), 1–10.
[20] Martin, J.L. (1995) Thioredoxin – a fold for all reasons. Structure 3, 245–250.
[21] Alanen, H.I., Salo, K.E., Pekkala, M., Siekkinen, H.M., Pirneskoski, A. and Ruddock, L.W. (2003) Defining the domain boundaries of the human protein disulfide isomerases. Antioxid. Redox Signal 5, 367–374.
[22] Freedman, R.B., Klappa, P. and Ruddock, L.W. (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. EMBO Rep. 3, 136–140.
[23] Munro, S. and Pelham, H.R. (1987) A C-terminal signal prevents secretion of luminal ER proteins. Cell 48, 899–907.
[24] Powis, G., Mustacich, D. and Coon, A. (2000) The role of the redox protein thioredoxin in cell growth and cancer. Free Radic. Biol. Med. 29, 312–322.
[25] Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W. (1997) Gene expression profiles in normal and cancer cells. Science 276, 1268–1272.
[26] Miyazaki, K., Noda, N., Okada, S., Hagiwara, Y., Miyata, M., Sakurabayashi, I., Yamaguchi, N., Sugimura, T., Terada, M. and Wakasugi, H. (1998) Elevated serum level of thioredoxin in patients with hepatocellular carcinoma. Biotherapy 11, 277–288.
[27] Knoblach, B., Keller, B.O., Groenendyk, J., Aldred, S., Zheng, J., Lemire, B.D., Li, L. and Michalak, M. (2003) ERp19 and ERp46, new members of the thioredoxin family of endoplasmic reticulum proteins. Mol. Cell Proteom. 2, 1104–1119.
[28] Mazzarella, R.A., Srinivasan, M., Haugejorden, S.M. and Green, M. (1990) ERp72, an abundant luminal endoplasmic reticulum protein, contains three copies of the active site sequences of protein disulfide isomerase. J. Biol. Chem. 265, 1094–1101.
[29] Klappa, P., Stromer, T., Zimmermann, R., Ruddock, L.W. and Freedman, R.B. (1998) A pancreas-specific glycosylated protein disulphide-isomerase binds to misfolded proteins and peptides with an interaction inhibited by oestrogens. Eur. J. Biochem. 254, 63–69.
[30] Feng, W., Bedows, E., Norton, S.E. and Ruddon, R.W. (1996) Novel covalent chaperone complexes associated with human chorionic gonadotropin beta subunit folding intermediates. J. Biol. Chem. 271, 18543–18548.
[31] Nigam, S.K., Goldberg, A.L., Ho, S., Rohde, M.F., Bush, K.T. and Sherman, M. (1994) A set of endoplasmic reticulum proteins possessing properties of molecular chaperones includes Ca(2+)-binding proteins and members of the thioredoxin superfamily. J. Biol. Chem. 269, 1744–1749.
[32] Linnik, K.M. and Herscovitz, H. (1998) Multiple molecular chaperones interact with apolipoprotein B during its maturation. The network of endoplasmic reticulum-resident chaperones (ERp72, GRP94, calreticulin, and BiP) interacts with apolipoprotein b regardless of its lipidation state. J. Biol. Chem. 273, 21368–213673.
[33] Kuznetsov, G., Chen, L.B. and Nigam, S.K. (1994) Several endoplasmic reticulum stress proteins, including ERp72, interact with thyroglobulin during its maturation. J. Biol. Chem. 269, 22990–22995.
[34] Miyaishi, O., Kozaki, K., Iida, K., Isobe, K., Hashizume, Y. and Saga, S. (1998) Elevated expression of PDI family proteins during differentiation of mouse F9 teratocarcinoma cells. J. Cell Biochem. 68, 436–445.
[35] Horibe, T., Gomi, M., Iguchi, D., Ito, H., Kitamura, Y., Masuoka, T., Tsujimoto, I., Kimura, T. and Kikuchi, M. (2004) Different contributions of the three CXXC motifs of human protein–disulfide isomerase-related protein to isomerase activity and oxidative refolding. J. Biol. Chem. 279, 4604–4611.