Amino Acids (2004) 26: 9–18
DOI 10.1007/s00726-003-0060-6

Amino Acids
© Springer-Verlag 2004
Printed in Austria

# Evidence for the existence of hypothetical proteins in human bronchial epithelial, fibroblast, amnion, lymphocyte, mesothelial and kidney cell lines

**J. E. Oh**[1], **K. Krapfenbauer**[2], **M. Fountoulakis**[2], **Th. Frischer**[1], and **G. Lubec**[1]

[1] Department of Pediatrics, University of Vienna, Vienna, Austria
[2] Center for Medical Genomics, F. Hoffmann-La Roche, Basel, Switzerland

Received October 3, 2003
Accepted November 20, 2003

**Summary.** The human genome maybe limited to about 30 000 genes whereas the human proteome may be represented by a rough estimate of one million proteins. A legion of proteins have been described and information about these structures are readily available in data banks. There remains, however, a large series of unknown or hypothetical proteins (HPs). Many of them have been predicted from nucleic acid sequences only and are therefore named predicted or HPs. Carrying out "protein hunting" by generating large maps of human cell lines, we aimed to find and identify HPs and provide an analytical tool thereof.

Cell lysates from human bronchial epithelial, fibroblast, amnion, lymphocyte, mesothelial and kidney cell lines were prepared and proteins run on two-dimensional gel-electrophoresis (2DE) with in-gel digestion and mass spectrometrical analysis using the MALDI-TOF principle.

16 HPs were found in these cell lines and some show cell-specific expressional patterns. HPs belong to several protein classes including structural, signaling, transcriptional/translational, chaperone-related and others. We furthermore provide analytical data i.e. pIs that were often different from predicted values in data banks.

A list of HPs has been shown to really exist in several human cell lines thus contributing to knowledge on protein machineries and cascades. Observed and predicted pI values are given representing an analytical tool along with unambiguous identification of protein spots by mass spectrometry independent of antibody availability and specificity thus complementing established methods.

**Keywords:** Hypothetical proteins – Structural proteins – Signaling proteins – Transcriptional/translational proteins – Chaperone related proteins

## Introduction

The large amount of genes of Humans leads to a multitude of gene products, the proteins. In contrast to genomic sequences, proteomic structures are presenting with a vast molecular diversity with maybe a million of proteins. Protein databanks are covering information on many thousands of proteins but the majority of gene products, its isoforms and posttranslational modifications have not been described at the protein level.

A large series of proteins have been theoretically "identified" or predicted from DNA or RNA sequences but have not been shown to really exist. It was the aim of this study, which forms a part of our "brainprot" – project (Lubec, 2003), to carry out "protein hunting" in order to show the existence of proteins, so-called hypothetical proteins (HPs) that have been predicted from nucleic acid sequences only and to provide the analytical basis for their characterisation, identification and determination. Although some protein databases provide some clues for predicting analytical parameters, these are misleading in a high percentage and analytical data have to be generated experimentally. Prediction of pIs (SWISSPROT, http://www.expasy.org) for instance, is not useful in many instances as the many isoforms and posttranslational modifications of the individual proteins and in particular of HPs, do not allow to determine the pI of a specific protein and not at all to help for searching a protein e.g. in a two-dimensional gel.

The advent of proteomic techniques as two-dimensional gel electrophoresis (2DE) with subsequent mass spectrometrical identification of proteins spots enables "protein hunting" for new or predicted proteins and their unambiguous identification (Fountoulakis, 2001) in a high-throughput mode (Myung et al., 2003; Peyrl et al., 2003a, b; Lubec et al., 2003; Engidawork et al., 2003; Cheon et al., 2001) warranting the generation of protein maps and expression patterns. In this approach we not

only intend to present new components for protein cascades and protein machineries of the cell but also show cell specific HPs expression in numerous cell lines that may help to identify cell specific marker candidate proteins.

We here report and classify predicted proteins based upon functional domain(s) and, when not possible, the principle of homology alignments with proteins of known functions is chosen.

## Materials and methods

### *Cell culture*

Bronchial epithelial cells (16HBE14o), fibroblast cells, mesothelial cells (Met-5A), lymphocytes, amniocytes and kidney cells (HK-2) were cultured and used for generation of maps. Cultures were kept at 37°C in the presence of 5% $CO_2$.

### *Bronchial epithelial cell line*

The human bronchial epithelial 16HBE14o- cell line is derived from surface epithelium of mainstream, second-generation bronchi (Cozens et al., 1994). Cells were grown on a collagen/fibronectin coating, in ecgonine methyl ester (EME)-medium supplemented with 10% fetal bovine serum (FBS), 1% penicillin (Gibco-BRL 15140-122), 1% streptomycin (Gibco-BRL 15140-122) and 1% L-glutamine.

### *Fibroblast cell line*

A human skin fibroblast cell line was obtained from Prof. M. Hengstschläger, Dpt of Obstetrics and Gynecology, Prenatal Diagnosis and Therapy, University of Vienna and cultured in Dulbecco's modified Eagle's medium (Life Technologies, Inc.) supplemented with 10% FBS (PAA Laboratories, Linz, Austria), 4 mM glutamine (Biochrom, Berlin, Germany) and 25 mM glucose.

### *Lymphocytic cell line*

Lymphocyte cell line 3610 is a spontaneously EBV transformed cell line from a patient with osteosarcoma and was obtained from the St. Anna Kinderspital-Forschungsinstitut (Vienna, Austria). The cell line was established from peripheral heparinised blood by density gradient centrifugation using Ficoll-Paque (Pharmacia, Uppsala, Sweden) and grown in RPMI 1640 with 10% FBS, 70 $\mu$M gentamicin sulfate and 2 mM glutamine at a density of $2 \times 106$ cells per ml in 96 well plates. The medium was changed every other day.

### *Amnion cell line*

Human amniotic fluid cell samples were obtained from amniocentesis performed for routine prenatal genetic diagnosis (Prof. M. Hengstschläger, Dpt of Obstetrics and Gynecology, Prenatal Diagnosis and Therapy, University of Vienna). Amniotic fluid cells were grown according to the standard culturing procedure in standard medium: Nutrient Mixture Ham's F10 (Gibco, Austria) medium supplemented with 10% FBS (Gibco, Austria), Ultroser G (BioSepra, France), 75 $\mu$M/L gentamicin (Biochrom, Germany) and L-Glutamine (Biochrom, Germany). Logarithmically growing amniocytes were harvested by trypsinisation.

### *Mesothelial cell line*

A mesothelial cell line, Met-5A (Arbeiter et al., 2001) was cultured in EME-medium supplemented with 10% FBS, 75 $\mu$g of streptomycin per ml, 75 units of penicillin per ml, 1% (v/v) dextrose, and 2 $\mu$g of fungizone® per ml (Gibco).

### *Kidney cell line*

Human immortalised epithelial HK-2 cells, derived from normal proximal convoluted tubules (Ryan et al., 1994), were grown according to the standard culturing procedure (ATCC, CRL-2190) in keratinocyte-serum free medium (Gibco-BRL 17005-042) with 5 ng/ml recombinant epidermal growth factor (positive for alkaline phosphatase, gamma glutamyltranspeptidase, leucine aminopeptidase, acid phosphatase, cytokeratin, alpha 3 beta 1 integrin, fibronectin; negative for factor VIII-related antigen, 6.19 antigen and CALLA endopeptidase) and 0.05 mg/ml bovine pituitary extract.

### *Sample preparation*

Harvested cells were washed three times in 10 mL phosphate buffered saline (Gibco BRL), centrifuged for 10 min at 800 g at room temperature and subsequently homogenised with 1.0 ml of sample buffer consisting of 7 M urea (Merck, Germany), 2 M thiourea (Sigma, St. Louis, MO), 4% CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propane-sulfonate) (Sigma), 65 mM 1,4-dithioerythritol (Merck), 1 mM EDTA (ethylenediaminetraacetic acid) (Merck), 1 mM PMSF, 0.5% carrier ampholytes and protease inhibitor complete (Roche, Switzerland). After homogenization samples were left at room temperature for 1 h and centrifuged at 14,000 rpm for 60 min and the supernatant was transferred into Ultrafree-4 centrifugal filter unit (Millipore, Bedford, MA), for desalting and concentrating proteins. Protein content of the supernatant was quantified by Bradford protein assay system (Bradford, 1976). The standard curve was generated using bovine serum albumin and absorbance was measured at 595 nm.

### *Two-dimensional gel electrophoresis (2-DE)*

Samples prepared from each cell line were subjected to 2-DE as described elsewhere (Langen et al., 1999; Weitzdoerfer et al., 2002). 1 mg protein was applied on immobilised pH 3–10 nonlinear gradient strips in sample cups at their basic and acidic ends. Focusing was started at 200 V and the voltage was gradually increased to 8000 V at 4 V/min and kept constant for a further 3 h (approximately 150000 Vh totally). After the first dimension, strips (13 cm) were equilibrated for 15 min in the buffer containing 6 M urea, 20% glycerol, 2% SDS, 2% DTT and then for 15 min in the same buffer containing 2.5% iodoacetamide instead of DDT. After equilibration, strips were loaded on 9–16% gradient sodium dodecylsulfate polyacrylamide gels for second-dimensional separation. Gels ($180 \times 200 \times 1.5$ mm) were run at 40 mA per gel. Immediately after the second dimension run, gels were fixed for 12 h in 50% methanol containing 10% acetic acid and stained with colloidal Coomassie blue (Novex, San Diego, CA) for 12 h on a rocking shaker. Molecular masses were determined by running standard protein markers (Bio-Rad) Laboratories, Hercules, CA), covering the range 10–250 kDa. pI values were used as given by the supplier of the immobilized pH gradient strips (Amersham Bioscience, Uppsala, Sweden). Excess of dye was washed out from the gels with distilled water and gels were scanned with Imagescanner (Amersham Bioscience, Uppsala, Sweden). Electronic images of the gels were recorded using Photoshop (Adobe) and PowerPoint (microsoft) software.

### *Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS)*

MALDI-MS analysis was performed as described (Berndt et al., 1999; Krapfenbauer et al., 2001) with some modifications. Spots were excised with a spot picker and gel pieces were washed twice in 100 $\mu$l of 30% acetonitrile in 50 mM ammonium bicarbonate and dried in a Speedvac

evaporator for 10 min. Proteins were rehydrated with 4 $\mu$l of 3 mM Tris-HCl, pH 9.0, containing 50 ng trypsin (Promega, Madison, WI) for 16 h or overnight at 37°C. Peptide extracts were vacuum-dried and resuspended in 7 $\mu$l of distilled water and shaken for 10 min. 4 $\mu$l of 50% acetonitrile; containing 0.3% trifluoroacetic acid and the standard peptides, des-Arg-bradykinin (Sigma, 2465.1989 Da); were added to each gel piece and shaken for 10 min. Sample application was performed using SymBiot I sample processor (PE Biosystems, Framingham, MA). 1.5 $\mu$l of digested peptide extracts were simultaneously spotted onto a MALDI target in 1 $\mu$l of matrix, consisting of a saturated solution of alpha-cyano-4-hydroxycinnamic acid (Sigma) in 50% acetonitrile containing 0.1% trifluoroacetic acid. MALDI-MS was performed using a Reflex III reflector time-of-flight mass spectrometer (Reflex 3, Bruker Analytics, Bremen, Germany). An accelerating voltage of 20 kV was used. Peptide matching and protein searches were performed automatically. The peptide masses were compared with the theoretical peptide masses of all available proteins from all species. Monoisotopic masses were used and a mass tolerance of 0.0025% was allowed. The algorithm used for determining the probability of a false positive match with a given MS-spectrum is described elsewhere (Morimote, 1994).

*Sequence alignments*

Most nucleic acid sequences of hypothetical proteins were directly submitted to the GenBank/EMBL/DDBJ database. Based on the assumption that sequence-domain similarities reflect functional relationship, it may be predicted how hypothetical proteins play a role in biological mechanism. A hypothetical protein showing one or more significant structural homologs, is predicted to have molecular properties similar to the homologs. For three hypothetical proteins (CAD23880, CAD34734 and CAD35032), poor information on predicted domains and function was available in data banks including SWISS-PROT database. Therefore, we performed BLAST search (NCBI, http://www.ncbi.nlm.nih.gov) to predict functional domains of these three proteins by their homologs.

## Results

*1) Structural proteins*

The nucleic acid sequence for Q8N473 was originally described in brain and submitted to EMBL/GenBank/DDBJ database (Strausberg R in 2002) and is a structural protein containing a fibrillar collagen domain at the C-terminus.

Q8WU19 protein contains a tubulin/FtsZ family domain involved in polymer formation. The nucleic acid sequence was observed in human eye tissue (Strausberg R, submission to database above, 2002). FtsZ is known as a polymer forming protein of bacterial cell division. We here show the presence of this HP in the mesothelial cell lineage where it may serve as a constituent of the cytoskeleton.

*2) Signaling proteins*

CAD35032 protein appears to contain the KOG3725, SH3 domain protein SH3GLB, which is involved in signal transduction mechanisms. For this protein, we could not find any data for functional domains, therefore we took the principle of homology alignments with proteins of known functions. KIAA 1848 protein, which contains KOG3725 domain, shows the 100% homology with CAD35032.

The nucleic acid sequence for Q9CYG9 protein was described in embryonic tissue (Kawai et al., 2001) and presents with an A-kinase (or PKA)-anchoring protein domain (AKAP95) reflecting a function for signaling and plays a role in mitotic chromosome condensation by acting as a targeting molecule for the condensin complex. The protein contains two zinc fingers which are thought to mediate the binding of AKAP95 to DNA (Eide et al., 2002).

*3) Transcriptional/translational proteins*

CAD34734 protein appears to contain KOG0460 as a functional domain which is observed in the Tu translation elongation factor protein, which shows 100% homology (Table 2) with CAD34734 thus identifying this structure as a component of the translation machinery.

*4) Chaperone related-protein*

cDNA sequence for BAC11617 protein was detected in placenta and submitted by Isogai T and Yamamoto J. It contains a KOG0713 domain which is representing DnaJ-class molecular chaperones in bacteria (Szyperski et al., 1994) and we here describe a possible human homolog.

*5) Others*

Q96CX2 protein contains two major functional domains, BTB/POZ and K+ channel tetramerisation. The BTB (for BR-C, ttk and bab) or POZ (for Pox virus and Zinc finger) domain is present near the N terminus of a fraction of zinc finger proteins and in proteins that contain the motif such as Kelch and a family of pox virus proteins. The BTB/POZ domain mediates homomeric dimerisation and in some instances heteromeric dimerisation. The tetramerization domain of voltage-gated K+ channels is found in a subset of a larger group of proteins that obtain BTB/POZ domain. The nucleic acid sequence was described in testis and embryonic carcinoma (Strausberg R, submission to database above, 2001).

Q8N1G4 protein includes the Leucine rich repeats domain present in a number of proteins with diverse functions, such as hormone-receptor interactions, enzyme

**Table 1.** Hypothetical proteins in bronchial, fibroblast, lymphocyte, amnion, mesothelial and kidney cell lines

| Name | Accession number (*Table 2) | Putative function | MW (kDa) | Domain | Theoretical pI (*predicted) | Observed pI | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | BR | FB | LY | AC | MES | KID |
| *1. Structural proteins* | | | | | | | | | | | |
| Hypothetical protein | Q8N473 | extracellular matrix constituent | 139.01 | Fibrillar collagen C-erminal domain, Collagen triple helix repeat | 5.7 | | 5.4 | | | | |
| Similar to tubulin alpha 2 | Q8WU19 | cytoskeleton | 37.22 | Tubulin/FtsZ family | 4.87 | | | | | 7.65 | |
| *2. Signaling proteins* | | | | | | | | | | | |
| Sequence 340 from patent wo022260 | *CAD35032 | signal transduction | 44.77 | KOG3725 | *5.42 | | | 5.65 | | | |
| 5730470h14rik protein | Q9CYG9 | mediate binding of AKAP (PKA or A-kinase anchoring protein)-95 to DNA | 65.17 | A-kinase anchoring protein 95 (AKAP95) | 5.12 | 5.35 | | | | | |
| *3. Transcription/translation realted proteins* | | | | | | | | | | | |
| Sequence 42 from patent wo022260 | *CAD34734 | translation, GTPase activity | 49.87 | KOG0460 | *7.26 | 8.2 | | | | 7.8 | |
| *4. Chaperone related proteins* | | | | | | | | | | | |
| cDNA psec0121 fis, clone place 1003085 | BAC11617 | protein turnover, posttranslational modification, chaperone-like activity | 40.51 | KOG0713 | *5.81 | 7.1 | | | | | |
| *5. Others* | | | | | | | | | | | |
| Hypothetical 35.7 kDa protein | Q96CX2 | voltage gated potassium channel activity, protein binding | 35.7 | BTB/POZ domain, K+ channel tetramerisation domain | 5.51 | | | | 5.45 | | |
| Hypothetical protein Kiaa1185 | Q8N1G4 | protein–protein interaction | 63.47 | Leucine rich repeats | 8.55 | | | | 8.6 | | |
| Sequence 1 from patent wo0208400 | *CAD23880 | amino acid transport | 57.93 | KOG0258 | *7.85 | | | | | | 7.0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| P 60 protein | O35814 | protein–protein interaction | 62.57 | tetratricopeptide repeat domain | 6.4 | | | 7.65<br>7.85 | |
| P 1.11659_4 | O60376 | integral membrane protein | 38.75 | SPFH domain/ Band 7 family | 6.4 | | 5.4 | | |
| Mitotic check point protein, MAD1 (mitotic arrest deficient, yeast homologue)-like 1 | Q9UNH0 | cell cycle regulation | 83.07 | MAD domain | 5.72 | | | | 5.73<br>5.7 |
| 16 days neonate thymus cDNA Riken full-length enriched library clone: a 130080m17 product | O9DCI1 (BAC29936) | mitochondrial inner membrane protein | 83.90 | KOG1854 | 6.18 | 6.65<br>6.8 | | | |
| Proliferation-associated protein 2g4 (cell cycle protein p38-2g4 homologue)-like 1 | Q9UQ80 | cell cycle arrest/ proliferation | 43.79 | metallopeptidase family M24 | 6.13 | 7.1<br>7.4 | 6.0<br>6.1<br>6.3 | 7<br>7.1<br>7.55 | 6.4 |
| Novel protein dj149a16.6 | CAB38260 | Unknown | 55.21 | KOG3833 | *6.77 | | | 8.2 | |
| Thioredoxin domain containing protein 5 (precursor) | Q8NBS9 | Anti-oxidant | 47.63 | Thioredoxin | 5.63 | | | 7.65 | |

inhibition, cell adhesion and cellular trafficking. The primary function of these motifs appears to be to provide a versatile structural framework for the formation of protein–protein interactions. The nucleic acid sequence was described in human uterus (Strausberg R, submission to database above, 2002).

CAD23880 protein seems to include KOG0258 domain as predicted from homology searches. This domain is found in alanine aminotransferase 2, which shows the 99% homology with CAD23880 (Table 2) and involved in amino acid transport and metabolism.

O35814 protein contains a tetratricopeptide repeat domain which mediates protein–protein interactions. The nucleic acid sequence was described in rat liver (Hoehfeld J, submission to database above, 1997). The tetratricopeptide repeat of typically 34 amino acids was first described in the yeast cell cycle regulator Cdc23p and later found to occur in a large number of proteins. Although common features in the interaction partners have not been defined. It has been proposed that TPR proteins preferably interact with WD-40 repeat proteins, but in many instances several TPR-proteins seem to aggregate to multi-protein complexes. Prominent examples of TPR-proteins include, Cdc16p, Cdc23p and Cdc27p components of the cyclosome/APC, the Pex5p/Pas10p receptor for peroxisomal targeting signals, the Tom70p co-receptor for mitochondrial targeting signals, Ser/Thr phosphatase 5C and the p110 subunit of O-GlcNAc transferase.

O60376 protein consists of SPFH domain/band 7 family, an integral membrane protein which is thought to regulate cation conductance. A variety of proteins belong to this family including prohibitins, cytoplasmic anti-proliferative proteins and stomatin, an erythrocyte membrane protein. The nucleic acid sequence of O60376 protein was described in fibroblast (Lamerdin JE et al., submission to database above, 1998).

Q9UNH0 protein contains a mitotic checkpoint protein family domain. This family consists of several eukaryotic mitotic checkpoint (Mitotic arrest deficient or MAD) proteins. The mitotic spindle checkpoint monitors proper attachment of the bipolar spindle to the kinetochores of aligned sister chromatids and causes cell cycle arrest in prometaphase when failures occur (Wassmann et al., 2003). The nucleic acid sequence of Q9UNH0 protein was described in testis (Seeley TW, submission to database above, 1999) and pancreas (Strausberg R, 2001).

The O9DCI1 protein includes a KOG1854 domain, which reflects the structure of a mitochondrial inner membrane protein. The nucleic acid sequence of this protein was described in the Mus musculus thymus (Carninci et al., 2000).

Q9UQ80 contains a metallopeptidase domain M24. Metallopeptidases are the most diverse of the four main types of protease, with more than 30 families identified to date (Rawlings et al., 1995). In these enzymes, a divalent cation, usually zinc, activates the water molecule. The metal ion is held in place by amino acid ligands, usually three in number. The known metal ligands are His, Glu, Asp or Lys and at least one other residue is required for catalysis, which may play an electrophillic role. Of the known metalloproteases, around half contain an HEXXH motif, which has been shown in crystallographic studies to form part of the metal-binding site.

The HEXXH motif is relatively common, but can be more stringently defined for metalloproteases as abXHEbbHbc, where 'a' is most often valine or threonine and forms part of the S1' subsite in thermolysin and neprilysin, 'b' is an uncharged residue, and 'c' a hydrophobic residue. The corresponding nucleic acid was described in lung and skin (Strausberg et al., 2002).

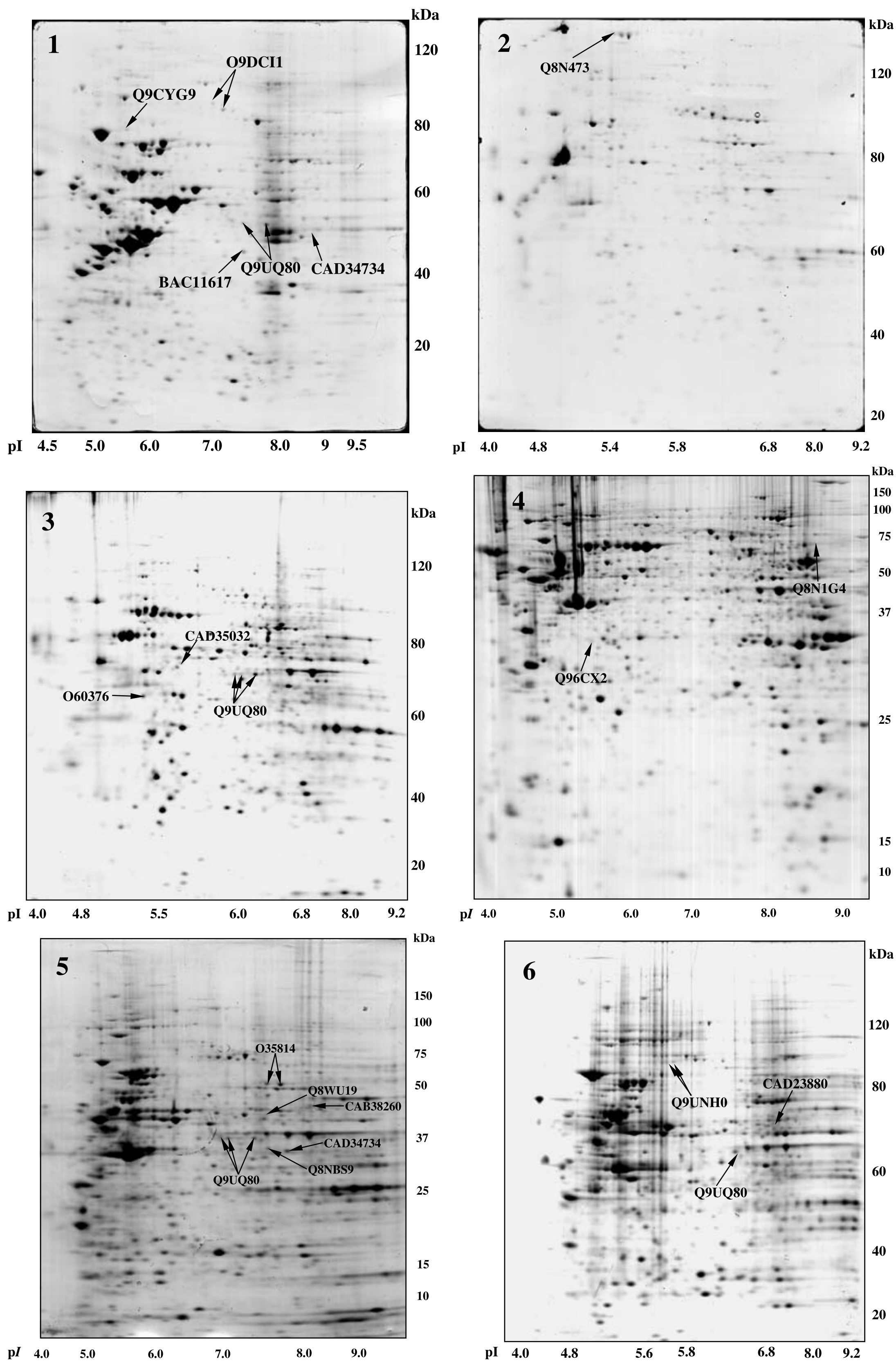
Expression at the nucleic acid level was reported for CAB38260 (Smith M, submission to Sanger centre, Hinxton, Cambridgeshire, UK) containing the domain, KOG3833 with so far unknown function. The gene was assigned to human chromosome 22. We here present this HP in a human mesothelial cell line.

Q8NBS9 protein contains 3 thioredoxin domains, which are small enzymes of a large enzyme family that participate in redox reactions via the reversible oxidation of an active centre disulfide bond. The nucleic acid sequence was originally detected in human eye (Strausberg et al., 2002). The HP is found in the endoplasmic reticulum and we detected expression in a mesothelial cell line.

## Discussion

The existence of so far predicted proteins is herein reported and thus contributes to the knowledge on individual protein machineries and cascades. This clearly shows that proteomics techniques using 2DE with subsequent mass spectrometrical identification of proteins by specific software is an appropriate method for ''protein hunting'' (Peyrl et al., 2003a, b). Our results also show that the number of known cellular structures is far from being fully identified and analysed.

One of the major findings of our study is that one protein may show several different pIs and the same protein can be detected in different cell lines. Furthermore,




**Figs. 1–6.** The position of individual formerly hypothetical proteins in 2DE gels in the individual cell lines assigned to accession numbers. **1** 16HBE 14o cell line; **2** Fibroblast cell; **3** Lymphocyte cell; **4** Amnion cell; **5** Met-5A; **6** Kidney cell HK-2

**Table 2.** Expected domains of some hypothetical proteins according to their homologues

| Accession numbers of query proteins | Matching proteins | Identities | positives | Domains |
|---|---|---|---|---|
| CAD23880 | alanine aminotransferase 2 [Homo sapiens] | 522/523 (99%) | 522/523 (99%) | 1 100 200 300 400 500 523 (aa) KOG0258 |
| CAD34734 | Tu translation elongation factor, mitochondrial [Homo sapiens] | 455/455 (100%) | 455/455 (100%) | 1 100 200 300 400 455 (aa) KOG0460 |
| CAD35032 | KIAA1848 protein [Homo sapiens] | 404/404 (100%) | 404/404 (100%) | 1 100 200 300 404 (aa) KOG3725 |

observed pIs can vary according to different cell lines and be totally different from pIs predicted by databases (SWISSPROT, http://www.expasy.org). Huge differences in pIs seem to be due to isoforms and posttranslational modifications, such as phosphorylation, nitrosylation, lipoylation and methylation generating the vast molecular diversity of proteins. Observed pIs of most HPs show basic shift compared to theoretical pIs (Table 1, Fig. 1–6). This might be possibly associated with level of phosphorylation/dephosphorylation, which is one of major processes of posttranslational modification. About the pI shifts, one group demonstrated that more acidic isoforms of Ah receptor nuclear translocator protein appear to have a higher level of phosphorylation (Tsai et al., 1997).

Predicting phosphorylation sites on proteins is an important step towards understanding the catalytic process itself and the resulting effects on signal transduction events. The prediction can be helpful to design a mutagenesis experiment on a newly found protein of possibly unknown function. The method may be integrated as a part of proteomics identification approach, where the whole protein repertoire from a specific cell type or organism is analysed for post-translational modifications and functionality (Blom et al., 1999). For example, one of our HPs, Q9UQ80 was observed in four different cell lines and each pI was different. In particular, in bronchial and mesothelial cell lines, all spots for this protein show shift-pattern toward basic. Furthermore, different isoforms of this one protein were observed even in one cell line. We have found 29 phosphorylation sites for this protein by using prediction system (Ser: 16, Thr: 9, Tyr: 4) (NetPhos 2.0 Server, http://www.cbs.dtu.dk/services/NetPhos/). This protein is likely to be phosphorylated/dephosphorylated these sites and thus causing pI shifts.

Therefore, we have learned that pI prediction systems for a HP cannot be reliably used in this context. A series of HPs were represented by spots with pI values very much different from predicted values, just to name similar to tubulin alpha 2 which showed pI of 7.65 in mesothelial cells rather than the predicted/calculated pI of 4.87 as an example. This finding of variable analytical data underscores the necessity of determining pIs experimentally and underlines the importance of the Rationale for carrying out this study. We also reconfirmed the existence of four HPs. Previously, our group observed one HP (Q8N1G4) in human brain and three (Q8WU19, O60376, CAD34734) were from rat tissues (Shin et al., unpublished).

The proteomic approach also provides evidence for the cell-specific protein expression of HPs studied herein as expressional patterns vary in the several cell lines used. This may be however, due to different culture conditions (Seow et al., 2001) used for cultivating the individual cell lines but we have taken that into account as we intended to examine protein expression under the conditions widely used by the scientific community, rather than inducing stress by using comparable media, antibiotics and chemicals. We also have to respect that the individual proteins may be presenting with different posttranslational modifications and thus escaped detection by 2DE.

The unambiguous identification of HPs is a great advantage of proteomics methods using the mass spectrometrical identification principle, independent of antibody specificity and availability. Specific software allows the

identification with high identification rates with up to 90 percent (own observations using the combination of MS and MS–MS data). The limitations of the proteomic method using 2DE are, however, enormous and include problems with determination of hydrophobic, highly insoluble hydrophilic, very low molecular weight proteins, just to name a few (Fountoulakis, 2001; Lubec et al., 2003).

Proteomics can show the existence of a protein sequence based upon a nucleic acid sequence and assignment of function is simply based upon the availability of known domains in nucleic acid or protein databases as e.g. SWISSPROT (http://www.expasy.org) or NCBI (http://www.ncbi.nlm.nih.gov/). We here show that sequence alignments may be used to provide a tentative function by homology searches (http://www.ncbi.nlm.nih.gov) for the proteins (CAD23880, CAD34734 and CAD35032) indicating the importance of bioinformatics for protein chemistry.

Several areas may benefit from the findings of differential expression of (now established) hypothetical proteins in individual cell lines as e.g. an approach to finding cell specific markers, identification of new components of structural, cytoskeleton, signaling, enzyme, chaperone and transcription cascades and we here provide an analytical tool for qualitative and quantitative determination of proteins and in particular, of hypothetical proteins, although the method of MALDI-TOF has to be complemented by additional proteomic instrumentation when specific questions as posttranslational modifications are arising.

## Acknowledgements

We are highly indebted to the Red Bull Company, Salzburg, Austria, for generous support of the study.

We appreciate provision of cell lines from Prof. Dr. M. Hengstschläger, Dr. Margit Rosner, Dr. Michaela Endemann, Prof. Szolt Szepfalusi, Prof. P. Ambros.

## References

Arbeiter K, Bidmon B, Endemann M, Bender TO, Eickelberg O, Ruffingshofer D, Mueller T, Regele H, Herkner K, Aufricht C (2001) Peritoneal dialysate fluid composition determines heat shock protein expression patterns in human mesothelial cells. Kidney Int 60: 1930–1937

Berndt P, Hobohm U, Langen H (1999) Reliable automatic identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. Electrophoresis 20: 3521–3526

Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294: 1351–1362

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254

Carninci P, Shibata Y, Hayatsu N, Sugahara Y, Shibata K, Itoh M, Konno H, Okazaki Y, Muramatsu M, Hayashizaki Y (2000) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res 10: 1617–1630

Cheon MS, Fountoulakis M, Dierssen M, Ferreres JC, Lubec G (2001) Expression profiles of proteins in fetal brain with Down syndrome. J Neural Transm [Suppl 61]: 311–319

Cozens AL, Yezzi MJ, Kunzelmann K, Ohrui T, Chin L, Eng K, Finkbeiner WE, Widdicombe JH, Gruenert DC (1994) CFTR expression and chloride secretion in polarized immortal human bronchial epithelial cells. Am J Respir Cell Mol Biol 10: 38–47

Eide T, Carlson C, Tasken KA, Hirano T, Tasken K, Collas P (2002) Distinct but overlapping domains of AKAP95 are implicated chromosome condensation and condensin targeting. EMBO Rep 3: 426–432

Engidawork E, Gulesserian T, Fountoulakis M, Lubec G (2003) Expression of hypothetical proteins in human fetal brain: increased expression of hypothetical protein 28.5 kDa in Down syndrome, a clue for its tentative role. Mol Genet Metab 78: 295–301

Fountoulakis M (2001) Proteomics: current technologies and applications in neurological disorders and toxicology. Amino Acids 21: 363–381

Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, et al (2001) Functional annotation of a full-length mouse cDNA collection. Nature 409: 685–690

Krapfenbauer K, Berger M, Friedlein A, Lubec G (2001) Changes in the levels of low-abundance brain proteins induced by kainic acid. Eur J Biochem 268: 3532–3537

Langen H, Berndt P, Roder D, Cairns N, Lubec G, Fountoulakis M (1999) Two-dimensional map of human brain proteins. Electrophoresis 20: 907–916

Lubec G (2003) Extending the scope of the journal. Amino Acids 24: I

Lubec G, Krapfenbauer K, Fountoulakis M (2003) Proteomics in brain research: potentials and limitations. Prog Neurobiol 69: 193–211

Morimote RI, Tissieres A, Georgopoulos C (eds) (1994) Cold Spring Harbor Laboratory, Cold Spring Harbor, NewYork

Myung JK, Gulesserian T, Fountoulakis M, Lubec G (2003) Deranged hypothetical proteins Rik protein, Nit protein 2 and mitochondrial inner membrane protein, Mitofilin, in fetal Down syndrome brain. Cell Mol Biol 49: 739–746

Peyrl A, Krapfenbauer K, Slavc I, Yang JW, Strobel T, Lubec G (2003a) Protein profiles of medulloblastoma cell lines DAOY and D283: identification of tumor-related proteins and principles. Proteomics 3: 1781–1800

Peyrl A, Krapfenbauer K, Slavc I, Strobel T, Lubec G (2003b) Proteomic characterization of the human cortical neuronal cell line HCN-2. J Chem Neuroanat 26: 171–178

Rawlings ND, Barrett AJ (1995) Evolutionary families of metallopeptidases. Methods Enzymol 248: 183–228

Ryan MJ, Johnson G, Kirk J, Fuerstenberg SM, Zager RA, Torok-Storb B (1994) HK-2: an immortalized proximal tubule epithelial cell line from normal adult human kidney. Kidney Int 45: 48–57

Seow TK, Korke R, Liang RC, Ong SE, Ou K, Wong K, Hu WS, Chung MC (2001) Proteomic investigation of metabolic shift in mammalian cell culture. Biotechnol Prog 17: 1137–1144

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, et al (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci USA 99: 16899–16903

Szyperski T, Pellecchia M, Wall D, Georgopoulos C, Wuthrich K (1994) NMR structure determination of the Escherichia coli DnaJ molecular

chaperone: secondary structure and backbone fold of the N-terminal region (residues 2–108) containing the highly conserved J domain. Proc Natl Acad Sci USA 91: 11343–11347

Tsai JC, Perdew GH (1997) Ah receptor nuclear translocator protein heterogeneity is altered after heterodimerization with the Ah receptor. Biochemistry 36: 9066–9072

Wassmann K, Liberal V, Benezra R (2003) Mad2 phosphorylation regulates its association with Mad1 and the APC/C. EMBO J 22: 797–806

Weitzdoerfer R, Fountoulakis M, Lubec G (2002) Reduction of actin-related protein complex 2/3 in fetal Down syndrome brain. Biochem Biophys Res Commun 293: 836–841

---

**Authors' address:** Prof. Dr. Gert Lubec, CChem, FRSC (UK), Department of Pediatrics, University of Vienna, Währinger Gürtel 18–20, A-1090 Vienna, Austria,
Fax: +43 40400 3194, E-mail: gert.lubec@akh-wien.ac.at