Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 808 (2004) 185–208
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Expressional pattern of known and predicted signaling proteins in seven human cell lines

Daniela Pollak [a], Kurt Krapfenbauer [b], Michael Fountoulakis [b], Andreas Peyrl [a], Gert Lubec [a,*]

[a] *Department of Pediatrics, Division Basic Science, University of Vienna, Währinger Gürtel 18, A-1090 Vienna, Austria*
[b] *Center for Medical Genomics, F. Hoffmann-La Roche, Basel, Switzerland*

Received 17 December 2003; received in revised form 29 March 2004; accepted 13 May 2004

Available online 15 June 2004

Abstract

Although a variety of signaling systems and signaling proteins have been described, cell specific expression of these structures has not yet been systematically studied. Human amnion, bronchial epithelial, fibroblast, glial, kidney, lymphocyte and mesothelial cells were subjected to two-dimensional-gel electrophoresis followed by analysis of protein spots by MALDI-TOF and subsequent identification by specific software. A series of well-documented signaling proteins showed cell specific expressional patterns. Five hypothetical proteins—hypothetical 37.5 kDa protein, similar to calsyntenin 1, hypothetical armadillo repeat/plakoglobulin ARM-repeat profile containing protein, 11 days embryo cDNA clone 2700084k13, hypothetical protein flj22171—so far predicted from their nucleic acid sequence only, were identified, complementing already reported signaling cascades. An analytical tool for the concomitant determination of a large series of signaling structures by an antibody independent protein-chemical method is provided.

Keywords: Proteomics; Signaling proteins; Hypothetical proteins

1. Introduction

All aspects of cellular life and function depend on the cells ability to receive extrinsic and intrinsic stimuli, to properly process the obtained information and finally react in an adequate and defined manner.

The execution of the developmental program responsible for directing the spatially and temporally correct differentiation of the vast spectrum of highly specialised cell types relies upon the proper integration and interpretation of a multiplicity of environmental and internal cues [1]. Not only differentiation but also growth, motility and apoptosis, which are important events during embryonic development as well as in adult life [2] require coordinate and organized reaction and adaptation to constantly changing conditions.

Several primary classes of signaling systems, operating at different time courses, provide great flexibility for intracellular communication. One class comprises ligand-gated ion channels, a second class consists of receptor tyrosine kinases, which typically respond to growth and trophic factors while a third class utilizes G-protein-linked signals and constitutes the largest number of receptors. The Rationale for carrying out the present study was to reveal distinct expressional patterns of signaling proteins (sp) in various human cell lines by a proteomic approach, as unlike the genome, which is essentially the same in all the somatic cells of an organism, the proteome is a dynamic entity different in each cell type. With regards to posttranslational mechanisms and the ensuing non-predictive correlation between mRNA and protein in terms of quality and quantity [3] it can be assumed that biological processes and systems can be described upon the comparison of protein expression patterns from cells or tissues. Taking into account that although a vast variety of signaling systems and signaling proteins have been studied so far, cell specific expression of these structures has not yet been systematically evaluated, we decided to carry out the present study.

We aimed to provide an analytical tool for the concomitant determination of a large series of signaling structures by a protein-chemical method independent of antibody

* Corresponding author. Tel.: +43 1 40400 3215; fax: +43 1 40400 3194.
E-mail address: gert.lubec@akh-wien.ac.at (G. Lubec).

1570-0232/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2004.05.008

availability and specificity, which forms the basis for respective studies at the protein level.

Secondly, we selected a series of well-documented and widely applied cell lines including some cell types used for medical diagnosis as, e.g. amnion cells, fibroblasts and lymphocytes. Defects of signaling proteins were linked to disease as 14–3–3 epsilon in patients with Miller–Dieker lissencephaly [4], adenine phosphoribosyltransferase in 2,8-dihydroxyadenine urolithiasis [5], mln50 in breast cancer [6], nucleophosmin in myelodysplastic syndrome [7] and nucleoside phosphate kinase a in neuroblastoma [8]. Determination of known or predicted structures in the cell lines used herein may be of importance and forming the basis for future diagnostic screens.

Using two-dimensional gel electrophoresis (2-DE) followed by matrix-assisted laser desorption ionization-mass spectroscopy to unambiguously identify proteins and determine their analytical characteristics (Mw, p*I*) this experiment was also set up to confirm the existence of hypothetical proteins, so far predicted from their nucleic acid structure only.

By these means we targeted the verification of "new" elements contributing to already well-described signaling cascades in the human organism.

A further objective was to show cell type specific expression of these additional signaling structures, not only accounting to functional characterisation of novel proteins as genomic data alone are not sufficient to determine the operational mode of a given gene product, but also to offer supplementary candidates in the hunt for cellular marker proteins.

2. Materials and methods

2.1. Cell culture

Amniocytes, bronchial epithelial cells (16HBE14o), fibroblasts (Hs 545 SK), glial cells (SVG p12), kidney cells (HK-2), lymphocytic cells and mesothelial cells Met-5A (HK-2) were cultured as given below, harvested and used for generation of maps. Cells were not synchronised and grown under conditions warranting optimal growth as high amounts of protein is needed for the generation of protein profiles. Cells were not available at comparable passage numbers from American Type Culture Collection (ATCC, Menassas, VA), the main source for cell cultures.

2.1.1. Amniocytes

Human amniocytes were obtained from amniocentesis performed for routine prenatal genetic diagnosis. They were grown according to the standard in vitro culturing procedure (two to six passages) in standard medium: Nutrient Mixture Ham's F10 (Gibco, Austria) supplemented with 10% fetal bovine serum (Gibco, Austria), Ultroser G (Biosepra, France), 75 μM/L gentamicin (Biochrom, Germany), L-glutamine (Biochrom, Germany) in a 95% humidified, 5% CO_2 chamber at 37 °C.

2.1.2. Bronchial epithelial cell line

The human bronchial epithelial 16HBE14o-cell line is derived from surface epithelium of mainstream, second-generation bronchi [9]. Cells (35–40 passages) were grown on a collagen/fibronectin coating, in ecgonine methyl ester (EME)-M supplemented with 10% fetal bovine serum (FBS), 1% penicillin (Gibco-BRL 15140-122), 1% streptomycin (Gibco-BRL 15140-122) and 1% L-glutamine.

2.1.3. Fibroblast cell line

The fibroblast cell line Hs 545 SK (three to five passages) obtained from ATCC is derived from human primary skin (trunk). Cells were cultured in DMEM-medium supplemented with 10% fetal calf serum (PAA: Lot A-1128-539), 4 mM glutamine and 4.5 g/l glucose.

2.1.4. Glial cells

SVG p12, a glial cell line from the third to fifth passage [10] was grown and maintained on Eagle's Minimal Essential Medium (EME, Biowhittaker) supplemented with 10% FBS, 75 μg of streptomycine/ml, 75 U of penicillin/ml, 1% (v/v) dextrose and 2 μg of fungizone®/ml (Gibco, Austria).

2.1.5. Kidney cell line

Human immortalised epithelial HK-2 cells (5–10 passages), derived from normal proximal convoluted tubule [11], 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 at 37 °C.

2.1.6. Lymphocytic cell line

Lymphocyte cell line 3610 is a spontaneously EBV transformed cell line (20–25 passages) 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 a density gradient centrifugation using Ficoll-Paque (Pharmacia, Uppsala, Sweden) and grown in RPMI 1640 with 10% FBS, 70 μM gentamicin sulfate and 2 mM glutamine at a density of 2×10^6 cells/ml in 96 well plates at 37 °C in a humidified atmosphere with 5% CO_2. The medium was changed every other day.

2.1.7. Mesothelial cell line

A mesothelial cell line (15–20 passages), Met-5A [12] was cultured in EME-medium supplemented with 10% FBS, 75 μg of streptomycin/ml, 75 units of penicillin/ml, 1% (v/v) dextrose, and 2 μg of Fungizone®/ml (Gibco).

2.2. 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 (Merck), 1 mM PMSF, 0.5% carrier ampholytes and protease inhibitor complete (Roche, Switzerland). After homogenisation 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 [13]. The standard curve was generated using bovine serum albumin and absorbance was measured at 595 nm.

2.3. Two-dimensional gel electrophoresis (2-DE)

Samples prepared from each cell line were subjected to 2-DE as described elsewhere [14,15]. Each sample was run in duplicate representing reference gels. One milligram of protein was applied on immobilized 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 1,50,000 Vh totally). After the first dimension, strips (13 cm) were equilibrated for 15 min in a 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 (SDS) polyacrylamide gels for second-dimensional separation. Gels (180 mm × 200 mm × 1.5 mm) were run at 40 mA per gel. Immediately after the second dimension, 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. p*I* 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.

2.4. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS)

MALDI-MS analysis was performed as described [16,17] with minor modifications. Briefly, spots were excised with a spot picker and gel pieces were washed twice in 100 μl of 30% acetonitrile in 50 mM ammonium bicarbonate and dried in a Speedvac evaporator for 10 min. Proteins were rehydrated with 4 μl 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 μl of distilled water and shaken for 10 min. Four microliters of 50% acetonitrile, containing 0.3% trifluoroacetic acid and the standard peptide, 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 Biopsystem, Framingham, MA). Digested peptide extracts (1.5 μl) were stimultaneously spotted onto a MALDI target in 1 μl 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 by the use of in-house developed software [16]. Peptide masses were compared to 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 probability of a false positive match with a given MS spectrum was determined for each analysis. Four matching peptides was the minimal requirement for an identity assignment. Unmatched peptides or miscleavage sites were not considered. The automatically identified proteins were checked individually and only human proteins or highly homologous counterparts from other species were considered.

3. Results

3.1. Protein identification

Proteins were identified by MALDI-MS on the basis of peptide mass matching [18], following in-gel digestion with trypsin. High-abundance spots, i.e. staining with Coomassie-blue were considered for analysis. Approximately 200 spots were excised from each of seven 2D gels (reference gels; Fig. 1A–G). The spots of each gel were selected randomly with the goal to detect as many gene products as possible; only spots identified as sp were included in this study. Each excised spot was analyzed individually.

In Table 1B, the theoretical *M*r and theoretical and observed p*I* values of identified proteins are listed together with data from the MS analysis, i.e. the numbers of matching peptides, peptide masses and peptide sequences. Identification was usually based on five or more matching peptides. In some cases, mainly for proteins of low molecular masses delivering few peptides only, the identification was based on four matching peptides.

Table 1
Expressional pattern of signaling proteins in human amniocytes, kidney cells, fibroblast cells, lymphocytes, mesothelial cells, bronchial epithelial cells and glial cells

Accession number	Protein name	Amnion	Kidney	Fibroblast	Lymphocyte	Mesothelial	Bronchial	Glial
A(a)								
P42655	14–3–3 protein epsilon (mitochondrial import stimulation factor l subunit) (protein kinase c inhibitor protein-1) (kcip-1) (14–3–3e)	X						
P29312	14–3–3 protein zeta/delta (protein kinase c inhibitor protein-1) (kcip-1) (factor activating exoenzyme s) (fas)	X				X		
P07741	Adenine phosphoribosyltransferase (ec 2.4.2.7) (aprt)	X						
P55263	Adenosine kinase (ec 2.7.1.20) (ak) (adenosine 5′-phosphotransferase)	X						
P27144	Adenylate kinase isoenzyme 4, mitochondrial (ec 2.7.4.3) (atp-amp transphosphorylase)	X						
Q01518	Adenylyl cyclase-associated protein 1 (cap 1)	X						X
Q9BX86	ALG-2 interacting protein 1			X	X			X
P35221	Alpha-1 catenin (cadherin-associated protein) (alpha e-catenin)	X						
P50995	Annexin a11 (annexin xi) (calcyclin-associated annexin 50) (cap-50) (56 kDa autoantigen)	X						X
P09525	Annexin a4 (annexin iv) (lipocortin iv) (endonexin i) (chromobindin 4) (protein ii) (p32.5) (placental anticoagulant protein ii) (pap-ii) (pp4-x) (35-beta calcimedin) (carbohydrate-binding protein p33/p41) (p33/41)	X			X			
P20073	Annexin a7				X			
P04083	Annexin I (lipocortin i) (calpactin ii) (chromobindin 9) (p35) (phospholipase a2 inhibitory protein)	X		X		X	X	
P07355	Annexin ii (lipocortin ii) (calpactin i heavy chain) (chromobindin 8) (p36) (protein i) (placental anticoagulant protein iv) (pap-iv)	X	X	X		X	X	X
P12429	Annexin iii (lipocortin iii) (placental anticoagulant protein iii) (pap-iii) (35-alpha calcimedin) (inositol 1,2-cyclic phosphate 2-phosphohydrolase)	X	X					

P08758	Annexin v (lipocortin v) (endonexin ii) (calphobindin i) (cbp-i) (placental anticoagulant protein i) (pap-i) (pp4) (thromboplastin inhibitor)	X	X	X		X	X	X
P08133	Annexin vi				X			
Q9P129	Calcium-binding transporter (fragment)	X						
Q05682	Caldesmon (cdm)	X		X		X		
P13861	Camp-dependent protein kinase type ii-alpha regulatory chain	X						
O00299	Chloride intracellular channel protein 1 (nuclear chloride ion channel 27) (ncc27)		X			X		X
O95833	Chloride intracellular channel protein 3						X	
Q9Z0W7	Chloride intracellular channel protein 4			X			X	
O88544	Cop9 subunit 4				X			
Q99829	Copine I	X						
P24387	Corticotropin-releasing factor binding protein precursor (crf-binding protein) (crf-bp) (corticotropin-releasing hormone-binding protein) (crh-bp)	X						
O95302	Fk-506 binding protein 9 (63 kDa)	X						
O14908	Gaip c-terminus interacting protein gipc (rgs-gaip interacting protein) (tax interaction protein 2) (tip-2)	X						
P29354	Growth factor receptor-bound protein 2 (grb2 adapter protein) (sh2/sh3 adapter grb2) (ash protein)	X						
P17080	gtp-binding nuclear protein ran (tc4) (ran gtpase) (androgen receptor-associated protein 24	X			X			X
P25388	Guanine nucleotide-binding protein beta subunit-like protein 12.3 (p205) (receptor of activated protein kinase c 1) (rack1) (receptor for activated c kinase)	X			X	X		
P11016	Guanine nucleotide-binding protein g(i)/g(s)/g(t) beta subunit 2 (transducin beta chain 2)	X						
P12268	Inosine-5′-monophosphate dehydrogenase 2 (ec 1.1.1.205) (imp dehydrogenase 2) (impdh-ii) (impd 2)	X						
Q8C1X9	Kidney ccl-142 rag cDNA, Riken full-length enriched library Clone: g430081d22 product: annexin a3, full insert sequence						X	
Q14847	lim and sh3 domain protein 1 (lasp-1) (mln 50)	X						

Table 1 (*Continued*)

Accession number	Protein name	Amnion	Kidney	Fibroblast	Lymphocyte	Mesothelial	Bronchial	Glial
Q99PF5	Map2 rna *trans*-acting protein marta1					X		
P28482	Mitogen-activated protein kinase 1 (ec 2.7.1) (extracellular signal-regulated kinase 2) (erk-2) (mitogen-activated protein kinase 2) (map kinase 2) (mapk 2) (p42-mapk) (ert1)	X						
Q9UFN0	Nipsnap4 protein (mgc:14553) (dkfzp564d177) (flj13953) (hspc299)	X						
P06748	Nucleophosmin (NPM) (Nucleolar phosphoprotein B23) (Numatrin)	X						
P15531	Nucleoside diphosphate kinase a (ec 2.7.4.6) (ndk a) (ndp kinase a) (tumor metastatic process-associated protein) (metastasis inhibition factor nm23) (nm23-h1)	X						
P22392	Nucleoside diphosphate kinase b (ec 2.7.4.6) (ndk b) (ndp kinase b) (nm23-h2) (c-myc purine-binding transcription factor puf)	X						
Q9Y5Y2	Nucleotide binding protein 2 (nbp 2)	X						
P30086	Phosphatidylethanolamine-binding protein (pebp) (neuropolypeptide h3) (hippocampal cholinergic neurostimulating peptide) (hcnp) (raf kinase inhibitor protein) (rkip)	X				X		X
Q9P1B1	pkcq-interacting protein picot	X	X					
P50395	rab gdp dissociation inhibitor beta (rab gdi beta) (gdi-2)	X		X	X			
P43487	ran-specific gtpase-activating protein (ran binding protein 1) (ranbp1)	X				X		
Q13283	ras-gtpase-activating protein binding protein 1 (gap sh3-domain binding protein 1) (g3bp-1)	X					X	X
Q969Q5	ras-related protein Rab-24 (Hsrbc unknown protein for mgc: 20400)	X						
Q15293	Reticulocalbin 1 precursor	X						
AAC36349	Retinoblastoma binding protein						X	
Q07960	rho-gtpase-activating protein 1 (gtpase-activating protein rhoogap) (rho-related small gtpase protein activator) (cdc42 gtpase-activating protein) (p50-rhogap)	X		X				

Q00007	Serine/threonine protein phosphatase 2a, 55 kDa regulatory subunit b, alpha isoform (pp2a, subunit b, b-alpha isoform) (pp2a, subunit b, b55-alpha isoform) (pp2a, subunit b, pr55-alpha isoform) (pp2a, subunit b, r2-alpha isoform)	X				X	
P08129	Serine/threonine protein phosphatase pp1-alpha 1 catalytic subunit	X			X		
P37140	Serine/threonine protein phosphatase pp1-beta catalytic subunit (ec 3.1.3.16) (pp-1b)	X					
Q13177	Serine/threonine-protein kinase pak 2 (ec 2.7.1) (p21-activated kinase 2) (pak-2) (pak65) (gamma-pak) (s6/h4 kinase)	X					
P30085	ump-cmp kinase (ec 2.7.4.14) (cytidylate kinase) (deoxycytidylate kinase) (cytidine monophosphate kinase)	X					
P21796	Voltage-dependent anion-selective channel protein 1 (vdac-1)		X	X	X	X	X
P45880	Voltage-dependent anion-selective channel protein 2 (vdac-2)		X	X	X	X	X
A(b)	Hypothetical proteins						
Q9CZI7	11 days embryo cDNA, Riken full-length enriched library Clone: 2700084k13, full insert sequence					X	
Q96D15	Hypothetical 37.5 kDa protein (reticuloalbin 3 [precursor]) (EF-hand calcium binding protein RLP 49)	X					
Q8BNU0	Hypothetical Armadillo repeat/plakoglobulin ARM-repeat profile Containing protein (Riken cDNA 2410153k17 gene)	X					
Q9H6K9	Hypothetical protein FLJ22171 (epidermal growth factor receptor pathway substrat 8 related protein 2)	X					
Q8N4K9	Similar to calsyntenin 1			X			

Table 1 (*Continued*)

Accession number	MW (theoretical)[a]	p*I* (theoretical)[b]	p*I* (experimental)[c]	Peptides matched	Start	Stop	Peptide mass	Peptide sequence
B(a)								
P42655	29173.90	4.63	AC[d]:4.5	8	94	117	2581,3319	LICCDILDVLDKHLIPAANTGESK
					130	140	1256,5889	YLAEFATGNDR
					130	141	1384,6836	YLAEFATGNDRK
					153	169	1819,9342	AASDIAMTELPPTHPIR
					170	189	2331,1982	LGLALNFSVFYYEILNSPDR
					170	192	2661,345	LGLALNFSVFYYEILNSPDRACR
					196	214	2087,9624	AAFDDAIAELDTLSEESYK
					215	224	1189,659	DSTLIMQLLR
P29312	28171.40	4.8	AC:4.6–4.7 (four spots)	7	27	40	1643,7847	NVTELNEPLSNEER
			MC[e]:5.0		41	54	1503,8617	NLLSVAYKNVVGAR
					90	108	2178,1325	ELEAVCQDVLSLLDNYLIK
					131	141	1236,6564	YLAEVATGEKR
					161	170	1245,614	EHMQPTHPIR
					197	225	3301,6088	TAFDDAIAELDTLNEDSYKDSTLIMQLL
					216	225	1189,659	DSTLIMQLLR
07741	27745.10	4.73	AC:5.9	5	13	25	1465,7453	SFPDFPTPGVVFR
					26	38	1444,7772	DISPVLKDPASFR
					56	65	1122,5774	IDYIAGLDSR
					90	105	1781,9081	LPGPTLWASYSLEYGK
					106	120	1696,8837	AELEIQKDALEPGQR
P55263	40545.43	6.24	AC:6.9	11	47	60	1633,8406	YSLKPNDQILAEDK
					63	70	992,5285	ELFDELVK
					99	116	1930,0111	AATFFGCIGIDKFGEILK
					99	110	1242,6169	AATFFGCIGIDK
					149	161	1351,7018	SLIANLAAANCYK
					180	197	1973,042	VCYIAGFFLTVSPESVLK
					198	207	1134,5385	VAHHASENNR
					208	223	1889,0177	IFTLNLSAPFISQFYK
					307	337	3265,6353	EIIDTNGAGDAFVGGFLSQLVSDKPLTECI
					338	348	1171,6565	AGHYAASIIIR
					349	361	1534,7086	RTGCTFPEKPDFH
P27144	25268.02	8.47	AC:8.9	8	7	17	995,5869	AVILGPPGSGK
					24	40	1925,011	IAQNFGLQHLSSGHFLR
					60	70	1249,7243	SLLVPDHVITR
					81	91	1325,6731	GQHWLLDGFPR
					125	133	1105,5999	RWIHPPSGR
					126	133	949,4991	WIHPPSGR
					175	185	1274,7334	DVAKPVIELYK
					188	200	1417,705	GVLHQFSGTETNK
Q01518	51541.84	8.12	GC[f]:7.7	6	36	58	2351,2243	AGAAPYVQAFDSLLAGPVAEYLK
					83	98	1700,8609	ALLVTASQCQQPAENK
					99	117	2172,2236	LSDLLAPISEQIKEVITFR
					154	165	1462,607	EMNDAAMFYTNR
					197	207	1276,6303	EFHTTGLAWSK
					329	346	2073,0313	VENQENVSNLVIEDTELK
Q9BX86	96023.17	6.13	GC:6.7	21	23	40	2104,9362	FIQQTYPSGGEEQAQYCR
			FC[g]:7.7,		60	69	1138,6198	HEGALETLLR
			LC[h]:7.8		120	146	2837,3398	SCVLFNCAALASQIAAEQNLDNDEGLK
					151	163	1518,783	HYQFASGAFLHIK

					164	172	975,5455	ETVLSALSR
					215	228	1530,7202	LANQAADYFGDAFK
					268	275	949,5089	KFGEEIAR
					312	321	1322,6107	KDNDFIYHDR
					313	326	1746,842	DNDFIYHDRVPDLK
					357	372	1790,9189	MVPVSVQQSLAAYNQR
					422	437	1727,9509	SVIEQGGIQTVDQLIK
					438	445	997,5662	ELPELLQR
					446	455	1244,6575	NREILDESLR
					448	455	974,5139	EILDESLR
					478	488	1317,7141	TPSNELYKPLR
					583	605	2504,2836	FLTALAQDGVINEEALSVTELDR
					640	653	1643,8322	QSNNEANLREEVLK
					675	685	1396,7449	FYNELTEILVR
					690	697	910,444	CSDIVFAR
					707	714	930,4991	DLQQSIAR
					715	744	2936,4489	EPSAPSIPTPAYQSSPAGGHAPTPPTPAPR
P35221	100071.27	5.95	AC:6.9; 7.0	12	81	97	1962,0146	ESQFLKEELVAAVEDVR
					287	299	1532,7931	QIIVDPLSFSEER
					300	307	1033,5411	FRPSLEER
					332	340	974,5074	IVAECNAVR
					417	423	912,4563	EYAQVFR
					496	524	3200,5791	VLTDAVDDITSIDDFLAVSENHILEDVNK
					525	539	1673,85	CVIALQEKDVDGLDR
					653	669	1818,88	TSVQTEDDQLIAGQSAR
					683	694	1378,6829	IAEQVASFQEEK
					773	781	1119,614	QDLLAYLQR
					782	794	1505,758	IALYCHQLNICSK
					889	900	1368,725	HVNPVQALSEFK
P50995	54389.71	7.53	AC:8.4	5	191	209	2043,0473	GTITDAPGFDPLRDAEVLR
			GC:7.3		286	301	1646,8392	GVGTDEACLIEILASR
					336	345	1100,6405	LLISLSQGNR
					388	399	1446,7467	AHLVAVFNEYQR
					430	438	1052,5147	NTPAFFAER
P09525	35751.52	5.85	AC:6.4	9	27	42	1692,8776	GLGTDEDAIISVLAYR
			LC:6.8		99	114	1604,7923	GAGTDEGCLIEILASR
					122	132	1371,6633	ISQTYQQQYGR
					149	158	958,5665	VLVSLSAGGR
					159	170	1379,6418	DEGNYLDDALVR
					191	198	938,5114	FLTVLCSR
					201	211	1414,7094	NHLLHVFDEYK
					224	239	1666,8508	SETSGSFEDALLAIVK
					258	268	1174,6045	GLGTDDNTLIR
P20073	50315.67	6.25	LC:7.6	6	177	192	1690,8369	GFGTDEQAIVDVVANR
					256	264	1059,6213	VLIEILCTR
					275	282	989,4135	CYQSEFGR
					326	333	893,4465	LYQAGEGR
					386	401	1849,9712	TILQCALNRPAFFAER
					408	418	1091,5313	GAGTDDSTLVR
P04083	38583.05	6.64	AC:7.0; 7.4; 7.8	16	8	24	2141,0153	QAWFIENEEQEYVQTVK
			MC:7.8, FC:7.7		28	51	2356,1531	GGPGSAVSPYPTFNPSSDVAALHK
			BC[i]:7.8		97	111	1605,9546	ALTGHLEEVVLALLK
					112	122	1262,5994	TPAQFDADELR

Table 1 (*Continued*)

Accession number	MW (theoretical)[a]	p*I* (theoretical)[b]	p*I* (experimental)[c]	Peptides matched	Start	Stop	Peptide mass	Peptide sequence
					127	142	1702,883	GLGTDEDTLIEILASR
					153	159	936,5136	VYREELK
					165	175	1213,5316	DITSDTSGDFR
					165	183	2024,0262	DITSDTSGDFRNALLSLAK
					184	202	2067,8821	GDRSEDFGVNEDLADSDAR
					187	202	1739,7331	SEDFGVNEDLADSDAR
					203	210	908,4461	ALYEAGER
					203	211	1064,5469	ALYEAGERR
					212	226	1678,9095	KGTDVNVFNTILTTR
					213	226	1550,8149	GTDVNVFNTILTTR
					268	279	1299,6383	CATSKPAFFAEK
					302	310	1064,4914	SEIDMNDIK
P07355	38472.85	7.56	AC:8.3, MC:8.6	16	8	24	2141,0153	QAWFIENEEQEYVQTVK
			FC:8; 8.5; 8.9; 9;		28	51	2356,1531	GGPGSAVSPYPTFNPSSDVAALHK
			9.1; 9.2		97	111	1605,9546	ALTGHLEEVVLALLK
			BC:7.4–7.6 (four spots)		112	122	1262,5994	TPAQFDADELR
			KC[j]:7.4		127	142	1702,883	GLGTDEDTLIEILASR
			GC:7.3; 7.5		153	159	936,5136	VYREELK
					165	175	1213,5316	DITSDTSGDFR
					165	183	2024,0262	DITSDTSGDFRNALLSLAK
					184	202	2067,8821	GDRSEDFGVNEDLADSDAR
					187	202	1739,7331	SEDFGVNEDLADSDAR
					203	210	908,4461	ALYEAGER
					203	211	1064,5469	ALYEAGERR
					212	226	1678,9095	KGTDVNVFNTILTTR
					213	226	1550,8149	GTDVNVFNTILTTR
					268	279	1299,6383	CATSKPAFFAEK
					302	310	1064,4914	SEIDMNDIK
P12429	36375.27	5.63	AC:5.7	16	9	26	1844,8997	LSLEGDHSTPPSAYGSVK
			KC:7.4		27	45	2155,0632	AYTNFDAERDALNIETAIK
					27	35	1086,4838	AYTNFDAER
					48	61	1542,8461	GVDEVTIVNILTNR
					67	75	1111,5516	QDIAFAYQR
					87	102	1650,976	SALSGHLETVILGLLK
					118	133	1720,8395	GLGTDEDSLIEIICSR
					134	143	1244,6212	TNQELQEINR
					151	166	1811,863	TDLEKDIISDTSGDFR
					156	166	1225,5679	DIISDTSGDFR
					177	194	2064,9801	RAEDGSVIDYELIDQDAR
					178	194	1908,8793	AEDGSVIDYELIDQDAR
					211	218	1035,5277	WISIMTER
					232	243	1460,6705	SYSPYDMLESIR
					248	271	2781,3983	GDLENAFLNLVQCIQNKPLYFADR
					312	322	1421,6927	SLYYYIQQDTK
P08758	35808.58	4.94	AC:4; 5; 5.5, FC:5.3	13	13	28	1781,8202	DYPDFSPSVDAEAIQK

			BC:4.9, KC:5.1		104	119	1673,8678	GAGTNEDALIEILTTR
			GC:5.0, MC:5.2		126	136	1350,6557	DISQAYYTVYK
					137	152	1713,7901	KSLGDDISSETSGDFR
					138	152	1585,6954	SLGDDISSETSGDFR
					154	162	929,5401	ALLTLADGR
					196	203	994,5376	FTEILCLR
					210	216	943,4508	LTFDEYR
					229	247	2085,0763	GELSGHFEDLLLAIVNCVR
					248	256	1018,5303	NTPAFLAER
					263	273	1222,6045	GIGTDEFTLNR
					279	287	1073,5821	SEIDLLDIR
					293	303	1301,6507	HYGYSLYSAIK
P08133	75742.08	5.42	LC:6.4; 6.5; 6.6	8	13	28	1781,8202	DYPDFSPSVDAEAIQK
					104	119	1673,8678	GAGTNEDALIEILTTR
					154	162	929,5401	ALLTLADGR
					210	216	943,4508	LTFDEYR
					229	247	2085,0763	GELSGHFEDLLLAIVNCVR
					248	256	1018,5303	NTPAFLAER
					263	273	1222,6045	GIGTDEFTLNR
					279	287	1073,5821	SEIDLLDIR
Q9P129	45819.43	5.31	AC:6.3	10	0	15	1746,8453	FVLPTAACQDAEQPTR
					16	25	1255,63	YETLFQALDR
					26	41	1670,8318	NGDGVVDIGELQEGLR
					141	156	1984,0031	DYFLFNPVTDIEEIIR
					160	182	2547,2056	HSTGIDIGDSLTIPDEFTEDEKK
					160	181	2419,111	HSTGIDIGDSLTIPDEFTEDEK
					189	201	1212,704	QLLAGGIAGAVSR
					224	231	941,465	MNIFGGFR
					261	268	1134,5241	FWAYEQYK
					362	370	1143,5455	SYWLDNFAK
P17080	24423.11	7.01	AC:8.5, LC:8.7	7	37	55	2180,1938	KYVATLGVEVHPLVFHTNR
			GC:7.3		38	55	2052,0992	YVATLGVEVHPLVFHTNR
					56	70	1689,8569	GPIKFNVWDTAGQEK
					99	105	922,4632	NVPNWHR
					110	122	1401,7206	VCENIPIVLCGNK
					142	151	1214,6035	NLQYYDISAK
					152	165	1784,9092	SNYNFEKPFLWLAR
P25388	350763.73	7.6	AC:8.5; 8,7 MC:8.4; 8.8 LC:9.1	8	44	56	1562,7898	LTRDETNYGIPQR
					47	56	1192,5577	DETNYGIPQR
					88	98	1264,6513	LWDLTTGTTTR
					106	117	1309,6364	DVLSVAFSSDNR
					139	154	1924,8467	YTVQDESHSEWVSCVR
					225	244	2219,0879	HLYTLDGGDIINALCFSPNR
					280	307	3011,4156	AEPPQCTSLAWSADGQTLFAGYTDNLVR
					308	316	1059,593	VWQVTIGTR
P11016	37331.03	5.6	AC:5.5	7	23	41	1958,9569	ACADATLSQITNNIDPVGR
					57	67	1336,6085	IYAMHWGTDSR
					78	88	1353,7029	LIIWDSYTTNK
					137	149	1409,6281	ELAGHTGYLSCCR
					197	208	1168,565	LFVSGACDASAK
					283	303	2412,1615	LLLAGYDDFNCNVWDALKADR
					304	313	1009,5161	AGVLAGHDNR

Table 1 (*Continued*)

Accession number	MW (theoretical)[a]	p*I* (theoretical)[b]	p*I* (experimental)[c]	Peptides matched	Start	Stop	Peptide mass	Peptide sequence
P12268	55804.98	6.44	AC:8.2	7	108	123	1820,9763	KYEQGFITDPVVLSPK
					136	148	1373,6611	HGFCGIPITDTGR
					181	194	1481,8661	REDLVVAPAGITLK
					208	223	1779,9821	LPIVNEDDELVAIIAR
					291	310	2086,1256	DKYPNLQVIGGNVVTAAQAK
					311	321	1156,6303	NLIDAGVDALR
					356	374	1892,0357	FGVPVIADGGIQNVGHIAK
Q8C1X9	36356.04	5.5	BC:6.2	13	13	28	1781,8202	DYPDFSPSVDAEAIQK
					104	119	1673,8678	GAGTNEDALIEILTTR
					126	136	1350,6557	DISQAYYTVYK
					137	152	1713,7901	KSLGDDISSETSGDFR
					138	152	1585,6954	SLGDDISSETSGDFR
					154	162	929,5401	ALLTLADGR
					196	203	994,5376	FTEILCLR
					210	216	943,4508	LTFDEYR
					229	247	2085,0763	GELSGHFEDLLLAIVNCVR
					248	256	1018,5303	NTPAFLAER
					263	273	1222,6045	GIGTDEFTLNR
					279	287	1073,5821	SEIDLLDIR
					293	303	1301,6507	HYGYSLYSAIK
Q14847	29717.16	6.61	AC:7.8	7	59	72	1608,7662	QSFTMVADTPENLR
					75	84	1202,6107	QQSELQSQVR
					85	91	972,4661	YKEEFEK
					96	108	1418,7253	GFSVVADTPELQR
					121	127	981,4302	YHEEFEK
					144	152	1014,4112	DSQDGSSYR
					187	196	1067,5828	EPAAPVSIQR
Q99PF5	74226.45	6.38	MC:8.3	9	72	80	992,4784	DAFADAVQR
					151	162	1226,6906	VPDGMVGLIIGR
					178	190	1354,6941	VQISPDSGGLPER
					267	284	2042,0665	MILIQDGSQNTNVDKPLR
					307	320	1557,6657	DQGGFGDRNEYGSR
					385	394	1184,6978	IINDLLQSLR
					449	462	1533,7996	AINQQTGAFVEISR
					629	646	1980,9743	IGQQPQQPGAPPQQDYTK
					655	683	2952,3502	QAQVATGGGPGAPPGSQPDYSAAWAEYYR
P28482	41389.71	6.50	AC:8.3	8	15	23	974,5041	GQVFDVGPR
					55	66	1508,693	ISPFEHQTYCQR
					77	90	1709,9418	FRHENIIGINDIIR
					79	90	1406,7728	HENIIGINDIIR
					138	147	1209,647	YIHSANVLHR
					164	171	894,4491	ICDFGLAR
					172	190	2144,0011	VADPDHDHTGFLTEYVATR
					342	352	1348,7449	LKELIFEETAR
Q9UFN0	28466.62	9.21	AC:9.2	6	30	39	1337,6144	QYDGIFYEFR
					63	79	1894,9055	TAHSELVGYWSVEFGGR
					89	95	922,4156	YDNFAHR
					167	178	1309,6992	AVHAHVNLGYTK
					179	192	1575,8254	LVGVFHTEYGALNR
					193	205	1598,7688	VHVLWWNESADSR
P06748	32575.02	4.64	AC:4.7	5	24	31	1023,4882	ADKDYHFK
					32	44	1568,7277	VDNDENEHQLSLR

					80	100	2227,2116	MSVQPTVSLGGFEITPPVVLR
					212	220	974,4889	DSKPSSTPR
					277	290	1819,8406	MTDQEAIQDLWQWR
P15531	17148.73	5.83	AC:6.6	4	56	65	1149,6398	DRPFFAGLVK
					88	104	1785,9135	VMLGETNPADSKPGTIR
					105	113	994,4762	GDFCIQVGR
					114	127	1485,7158	NIIHGSDSVESAEK
P22392	17298.04	8.52	AC:8.8	4	88	104	1785,9135	VMLGETNPADSKPGTIR
					105	113	994,4762	GDFCIQVGR
					114	123	1069,5621	NIIHGSDSVK
					128	142	1895,9759	EISLWFKPEELVDYK
Q9Y5Y2	28825.33	5.55	AC:5.8	5	28	39	1274,7293	STISTELALALR
					44	59	1681,9277	KVGILDVDLCGPSIPR
					45	59	1553,833	VGILDVDLCGPSIPR
					74	82	1060,556	GWAPVFLDR
					234	260	2904,4367	TLEEGHDFIQEFPGSPAFAALTSIAQK
P30086	20925.59	7.43	AC:8.7	6	46	60	1632,7952	NRPTSISWDGLDSGK
					61	74	1560,8243	LYTLVLTDPDAPSR
					92	111	1949,9421	GNDISSGTVLSDYVGSGPPK
					118	130	1708,903	YVWLVYEQDRPLK
					131	139	1046,4921	CDEPILSNR
					160	177	2012,9029	APVAGTCYQAEWDDYVPK
Q9P1B1	37432.03	5.31	AC:5.4, KC:5.5	10	67	91	2845,4498	LEAEGVPEVSEKYEISSVPTFLFFK
					79	91	1577,8226	YEISSVPTFLFFK
					96	109	1535,8152	IDRLDGAHAPELTK
					114	135	2422,2071	HASSGSFLPSANEHLKEDLNLR
					114	129	1681,8268	HASSGSFLPSANEHLK
					171	187	2069,9533	HNIQFSSFDIFSDEEVR
					265	289	2954,4468	QILEILNSTGVEYETFDILEDEEVR
					294	307	1729,8559	AYSNWPTYPQLYVK
					319	331	1523,8766	ELKENGELLPILR
					322	331	1153,6557	ENGELLPILR
P50395	50663.25	6.11	AC:7.5, FC:7.7 LC:7.6	10	35	53	2117,0041	NPYYGGESASITPLEDLYK
					35	54	2273,1049	NPYYGGESASITPLEDLYKR
					68	78	1312,6989	GRDWNVDLIPK
					142	155	1712,8617	FLVYVANFDEKDPR
					173	192	2279,1782	KFDLGQDVIDFTGHALALYR
					174	192	2151,0835	FDLGQDVIDFTGHALALYR
					193	207	1845,7934	TDDYLDQPCYETINR
					221	239	2141,1032	SPYLYPLYGLGELPQGFAR
					348	359	1324,7337	YIAIVSTTVETK
					364	378	1778,0028	EIRPALELLEPIEQK
P43487	23310.12	5.19	AC:5.2, MC:5.3	5	39	49	1381,635	TLEEDEEELFK
					57	67	1335,6198	FASENDLPEWK
					118	140	2682,309	AWVWNTHADFADECPKPELLAIR
					141	149	1034,5252	FLNAENAQK
					152	158	912,4233	TKFEECR
Q13283	52164.24	5.36	AC:5.7, BC:5.9 GC:5.4	5	123	131	1210,5988	FYVHNDIFR
					335	352	2083,0324	HPDSHQLFIGNLPHEVDK
					353	369	2031,015	SELKDFFQSYGNVVELR
					357	369	1573,7623	DFFQSYGNVVELR
					376	392	1937,9614	LPNFGFVVFDDSEPVQK

Table 1 (*Continued*)

Accession number	MW (theoretical)[a]	p*I* (theoretical)[b]	p*I* (experimental)[c]	Peptides matched	Start	Stop	Peptide mass	Peptide sequence
Q969Q5	23124.12	5.85	AC:8.7; 8.8	6	59	68	1001,5723	IQSGLGALSR
					69	83	1599,8311	SHDTTSNTLAQLLAK
					87	97	1169,5642	VSSHANAAQER
					108	118	1192,6778	LEANHGLLVAR
					196	210	1509,8875	KGPAAPPPTPVKPPR
					197	210	1381,7928	GPAAPPPTPVKPPR
Q15293	38890.00	4.86	AC:4; 4.6; 5	6	37	65	3359,5874	VVRPDSELGERPPEDNQSFQYDHEAFLG
					70	82	1565,7419	TFDQLTPDESKER
					86	104	2136,0421	IVDRIDNDGDGFVTTEELK
					188	203	2020,908	EEFTAFLHPEEFEHMK
					249	255	969,4414	EQFNEFR
					270	285	1949,9225	HWILPQDYDHAQAEAR
AAC36349	47820.08	4.89	BC:4.9	5	115	123	1067,5215	INHEGEVNR
					138	166	3201,5105	TPSSDVLVFDYTKHPSKPDPSGECNPDLR
					291	298	973,5452	TVALWDLR
					335	343	1130,63	RLNVWDLSK
					344	370	2873,3543	IGEEQSPEDAEDGPPELLFIHGGHTAK
Q07960	50435.76	5.85	AC:7.1	12	59	68	1313,5781	WDDPYYDIAR
			FC:7.2		69	82	1586,7898	HQIVEVAGDDKYGR
					84	91	908,5009	IIVFSACR
					168	180	1532,9786	TLLILFKPLISFK
					185	199	1840,9451	IFYVNYLSELSEHVK
					200	207	925,5452	LEQLGIPR
					252	263	1404,7823	NPEQEPIPIVLR
					264	281	2020,0466	ETVAYLQAHALTTEGIFR
					323	348	3041,5567	ELPEPLLTFDLYPHVVGFLNIDESQR
					349	368	2311,2616	VPATLQVLQTLPEENYQVLR
					369	385	1918,9991	FLTAFLVQISAHSDQNK
					421	438	1968,9671	FLLDHQGELFPSPDPSGL
Q00007	51692.08	5.82	AC:7.4	7	18	47	3088,4543	GAVDDDVAEADIISTVEFNHSGELLATGD
			BC:5.5; 5.8		51	61	1361,7039	VVIFQQEQENK
					127	136	1219,6412	DKRPEGYNLK
					142	152	1322,7043	YRDPTTVTTLR
					170	198	3343,5812	IFANAHTYHINSISINSDYETYLSADDLR
					199	209	1409,7514	INLWHLEITDR
					267	277	1332,6049	LFEEPEDPSNR
P08129	37512.08	5.94	AC:8.7	15	6	14	1000,577	LNLDSIIGR
					15	25	1183,6775	LLEVQGSRPGK
					26	35	1215,631	NVQLTENEIR
					43	59	1953,1272	EIFLSQPILLELEAPLK
					60	73	1665,8029	ICGDIHGQYYDLLR
					74	95	2598,2148	LFEYGGFPPESNYLFLGDYVDR
					98	110	1494,8211	QSLETICLLLAYK
					111	121	1439,8023	IKYPENFFLLR
					113	121	1198,6238	YPENFFLLR
					122	131	1100,4888	GNHECASINR
					132	140	1137,4907	IYGFYDECK
					150	167	1999,9472	TFTDCFNCLPIAAIVDEK
					221	233	1311,6924	GVSFTFGAEVVAK
					234	245	1509,7971	FLHKHDLDLICR
					246	259	1639,7728	AHQVVEDGYEFFAK

P37140	37186.83	5.84	AC:6.8	12	42	58	1953,1272	EIFLSQPILLELEAPLK
					59	72	1603,7873	ICGDIHGQYTDLLR
					73	94	2582,2199	LFEYGGFPPEANYLFLGDYVDR
					97	109	1494,8211	QSLETICLLLAYK
					110	120	1439,8023	IKYPENFFLLR
					112	120	1198,6238	YPENFFLLR
					131	139	1137,4907	IYGFYDECK
					149	166	1999,9472	TFTDCFNCLPIAAIVDEK
					220	232	1313,6717	GVSFTFGADVVSK
					237	244	984,4918	HDLDLICR
					245	258	1639,7728	AHQVVEDGYEFFAK
					303	318	1761,9004	YQYGGLNSGRPVTPPR
Q13177	58004.60	5.69	AC:6.4	7	160	190	3315,5446	GTEAPAVVTEEEDDDEETAPPVIAPRPDH
					350	366	2021,0353	ECLQALEFLHANQVIHR
					383	398	1784,8497	LTDFGFCAQITPEQSK
					400	416	1924,9266	STMVGTPYWMAPEVVTR
					450	467	1972,0353	ALYLIATNGTPELQNPEK
					468	478	1377,7616	LSPIFRDFLNR
					492	500	1124,6445	ELLQHPFLK
P30085	22222.34	5.44	AC:6.0	7	26	38	1479,7569	YGYTHLSAGELLR
					42	54	1520,7568	KNPDSQYGELIEK
					61	72	1324,8426	IVPVEITISLLK
					88	95	964,5238	FLIDGFPR
					96	105	1216,569	NQDNLQGWNK
					111	129	2234,0222	ADVSFVLFFDCNNEICIER
					179	195	1954,9251	SVDEVFDEVVQIFDKEG
P21796	30641.40	8.6	KC:9.1; 9.2; 9.3	9	62	72	1374,6557	WTEYGLTFTEK
			LC:9.0; 9.4; 9.45; 9.5		73	91	2176,0483	WNTDNTLGTEITVEDQLAR
			MC:9.5;		95	107	1400,6673	LTFDSSFSPNTGK
			BC:7.6; GC:8.0		162	172	1213,6194	VTQSNFAVGYK
					173	195	2600,186	TDEFQLHTNVNDGTEFGGSIYQK
					199	216	1946,0059	KLETAVNLAWTAGNSNTR
					200	216	1817,9113	LETAVNLAWTAGNSNTR
					223	234	1357,6074	YQIDPDACFSAK
					235	254	2103,1771	VNNSSLIGLGYTQTLKPGIK
P45880	38092.73	6.32	KC:6.7; 7.5; 8.1	9	38	45	934,5092	AARDIFNK
			LC:8.6, MC:8.6		89	99	1376,6173	WCEYGLTFTEK
			BC:8.7,		122	134	1428,6985	LTFDTTFSPNTGK
			GC:7.2; 7.4		192	199	940,4624	NNFAVGYR
					200	222	2528,165	TGDFQLHTNVNDGTEFGGSIYQK
					223	243	2285,0138	VCEDLDTSVNLAWTSGTNCTR
					250	261	1293,6666	YQLDPTASISAK
					262	281	2103,1521	VNNSSLIGVGYTQTLRPGVK
					300	307	974,4831	VGSPWSWR

Table 1 (*Continued*)

Accession number	MW (theoretical)[a]	p*I* (theoretical)[b]	p*I* (experimental)[c]	Peptides matched	Start	Stop	Peptide mass	Peptide sequence
B(b)								
Q9CZI7	38609.13	7.53	BC:7.8	5	10	36	2938,3807	LSLEGDHSTPPSAYGSVKPYTNFDAER
					49	62	1542,8461	GVDEVTIVNILTNR
					68	76	1111,5516	QDIAFAYQR
					178	195	2064,9801	RAEDGSVIDYELIDQDAR
					313	323	1421,6927	SLYYYIQQDTK
Q96D15	37482.98	4.74	AC:4.5	7	33	61	3215,463	VHQAAPLSDAPHDDAHGNFQYDHEAFL
					62	78	1976,9529	EVAKEFDQLTPEESQAR
					66	78	1549,7107	EFDQLTPEESQAR
					89	102	1445,6999	AGDGDGWVSLAELR
					103	111	1110,5788	AWIAHTQQR
					201	215	1743,9095	DIVIAETLEDLDRNK
					201	213	1501,7721	DIVIAETLEDLDR
Q8BNU0	50683.25	5.65	AC:8.5	6	5	15	1220,6252	IAQETFDAAVR
					31	48	1991,9889	EAVEQFESQGVDLSNIVK
					231	241	1214,6162	VPFGHAHNHAK
					259	276	1905,9708	AFLDNPGILSELCGTLSR
					327	340	1456,7731	AIAGNDDVKDAIVR
					455	463	1041,5132	DLGCHVELR
Q9H6K9	80620.61	6.39	AC:8.1	6	66	76	1329,7141	VEHLTTFVLDR
					214	229	1575,8464	APAPAPPGTVTQVDVR
					308	318	1126,6561	KGPGEGVLTLR
					309	318	998,5615	GPGEGVLTLR
					319	335	2007,0013	AKPPPPDEFLDCFQKFK
					422	431	1245,6608	EQFIPPYVPR
Q8N4K9	108669.51	4.85	LC:7.5	6	254	270	1957,9671	ISIKPTCTPGWQGWNNR
					536	544	914,5404	GNLAGLTLR
					562	574	1430,71	EGLDLQVLEDSGR
					611	618	915,5035	QFPTPGIR
					654	664	1223,6626	ISLSGVHHFAR
					665	682	2014, 9726	AASEFESSEGVFLFPELR

[a] Molecular weight.
[b] Theoretical isoelectric point.
[c] Observed isoelectric point.
[d] Amnion cells.
[e] Mesothelial cells.
[f] Glial cells.
[g] Fibroblast cells.
[h] Lymhocyte cells.
[i] Bronchial cells.
[j] Kidney cells.

3.2. Hypothetical proteins

Five hypothetical proteins with putative signaling function have been identified.

The 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 homologues is predicted to have molecular properties similar to the homologues.

In the following domains obtained from Swiss-Prot/Trembl database (http://www.us.expasy.org/prot/), using PROSITE and PFAM program, are given in brackets.

3.2.1. Hypothetical 37.5 kDa protein (reticulocalbin 3 [precursor])

The sequence encoding this protein belongs to the HBG003824 gene family (CREC family) including 21 sequences of eight taxons, six of which have been identified in human. CREC family members are involved in Vitamin D signaling pathways and signaling in malignant transformation in papillomavirus infection [19].

Rat gene LOC308580 and mouse gene D7Ertd671e show approximately 87% sequence similarity to the human gene located on 19q13.33. Electronic Northern analysis shows high clone frequency in spleen and brain tissue (GeneCards: http://www.bioinfo.weizmann.ac.il/cards-bin).

Containing an endoplasmic reticulum targeting sequence (PS00014) which seems to distinguished proteins that permanently reside in the lumen of the endoplasmic reticulum

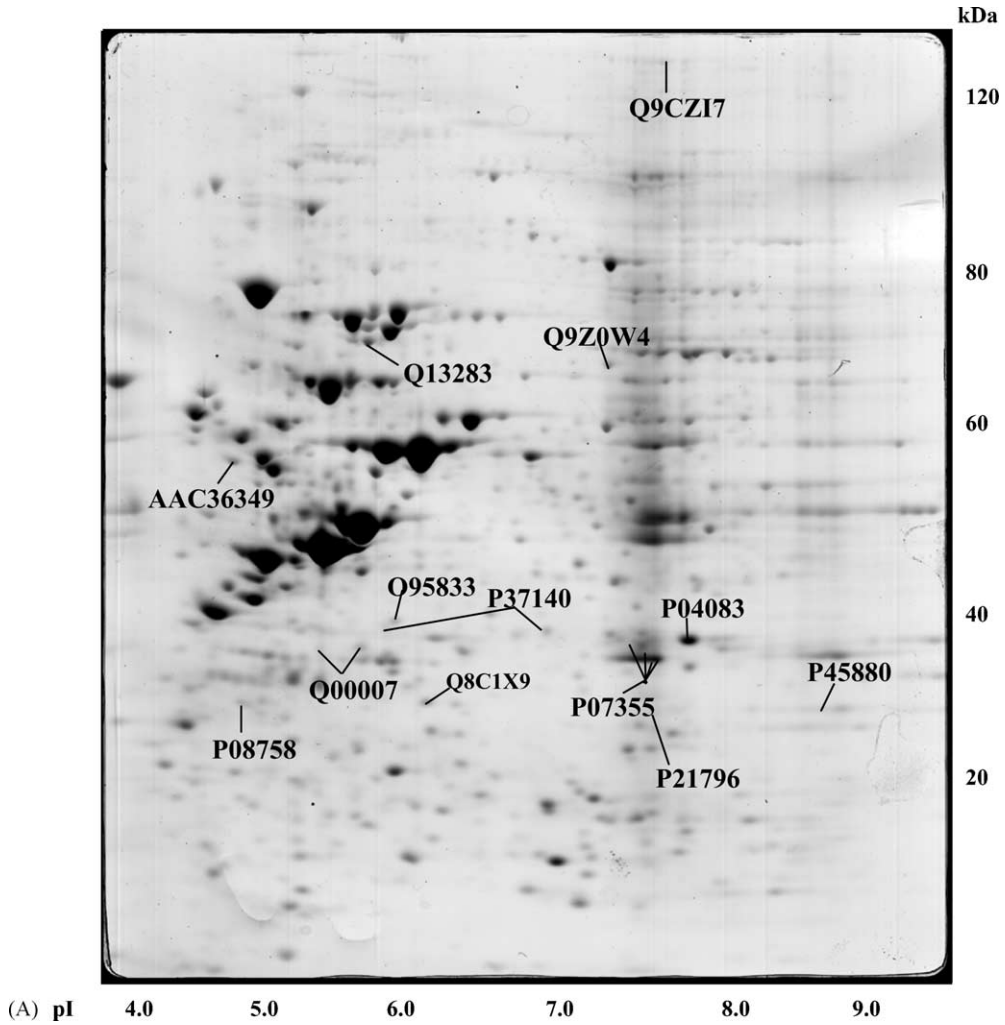

Fig. 1. (A) 2-DE gel image of bronchial epithelial cell line proteins depiciting identified proteins. Accession numbers are given. Proteins were extracted and separated on an immobilized pH 3–10 non-linear gradient strip followed by separation on a 9–16% gradient polyacrylamide gel. The gel was stained with Coomassie blue and spots were analyzed by MALDI-MS. (B) 2-DE gel image of amnion cells proteins depiciting identified proteins. (C) 2-DE gel image of mesothelial cell line proteins depicting identified proteins observed in this cell line. (D) 2-DE gel image of kidney cell line proteins depicting identified proteins. (E) 2-DE gel image of lymphocyte cell line proteins depicting identified proteins. (F) 2-DE gel image of glial cell line proteins depicting identified proteins. (G) 2-DE gel image of fibroblast cell line proteins depicting identified proteins.

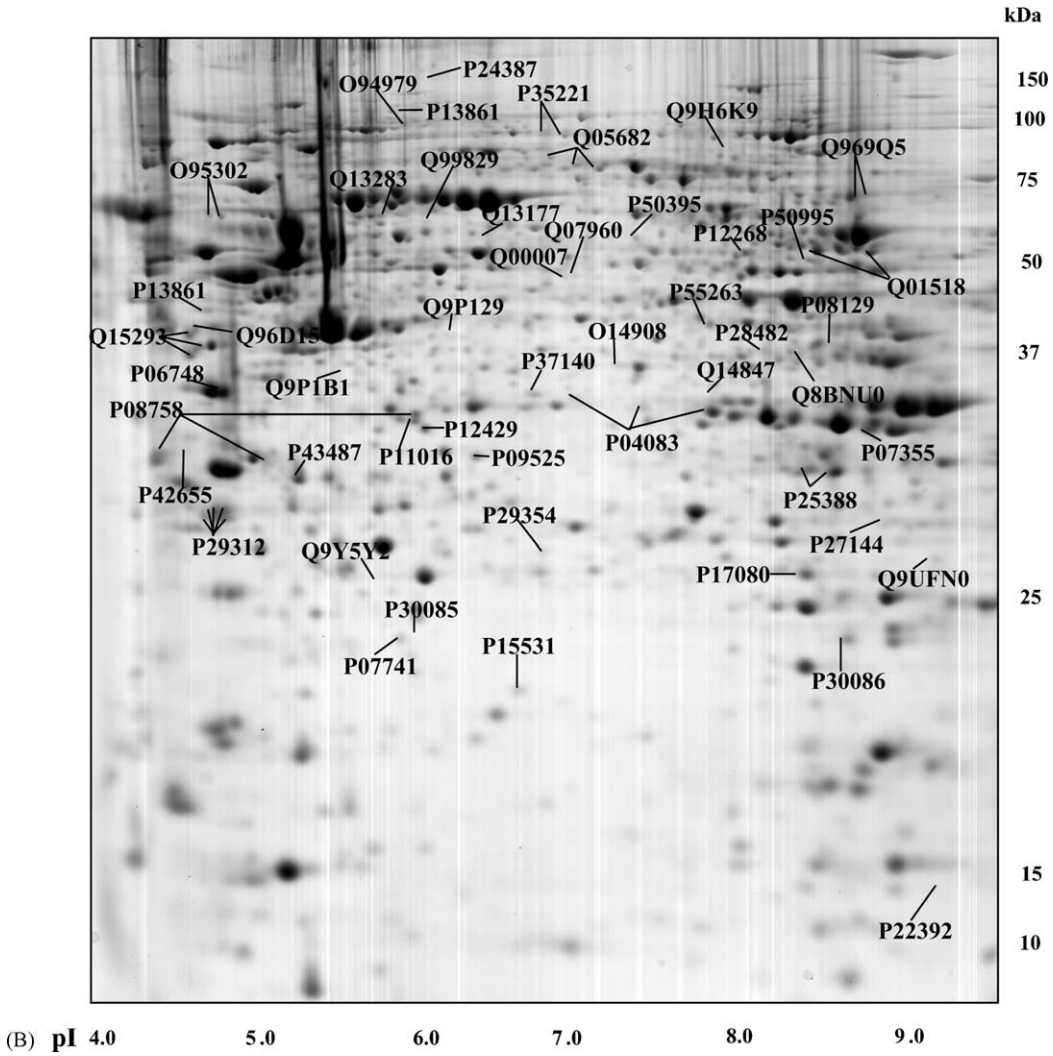

Fig. 1. (*Continued*)

(ER) from newly synthesized secretory proteins by the presence of the C-terminal sequence Lys–Asp–Glu–Leu (KDEL) the intracellular localisation of this protein can be deduced.

The presence of six EF-hand calcium binding domains (PS00018) leads to the assumption that hypothetical 37.5 kDa protein is involved in signal transduction processes.

Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand. This type of domain consists of a 12-residue loop flanked on both side by a twelve residue alpha-helical domain.

Calcium sensor proteins enable the cell to detect a stimulatory calcium influx and thereby transduce this signal into a variety of cellular processes. The mechanism of this molecular switch lies in the conformational changes induced by calcium binding. At low calcium concentrations EF-hands proteins are inactive and become active as the calcium concentration increases.

3.2.2. Similar to calsyntenin 1

The sequence encoding this protein belongs to the HBG025961 gene family comprising 16 sequences of four taxons, five of which have been identified in human.

Like 650 other proteins this sequence includes a cadherin domain (IPR002126) clearly assigning this protein to calcium signaling pathways. Cadherins show a rather low evolutionary rate providing 96% identical sequences of mouse and human N-cadherins.

Structurally, cadherins comprise a number of domains: these include a signal sequence, a propeptide of around 130 residues, an extracellular domain of around 600 residues, a single transmembrane domain and a well-conserved C-terminal cytoplasmic domain of about 150 residues. The calcium-binding region of cadherins is thought to be located in the extracellular domain. Cadherins are glycoproteins involved in Ca^{2+}-mediated cell-cell adhesion. They preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand.

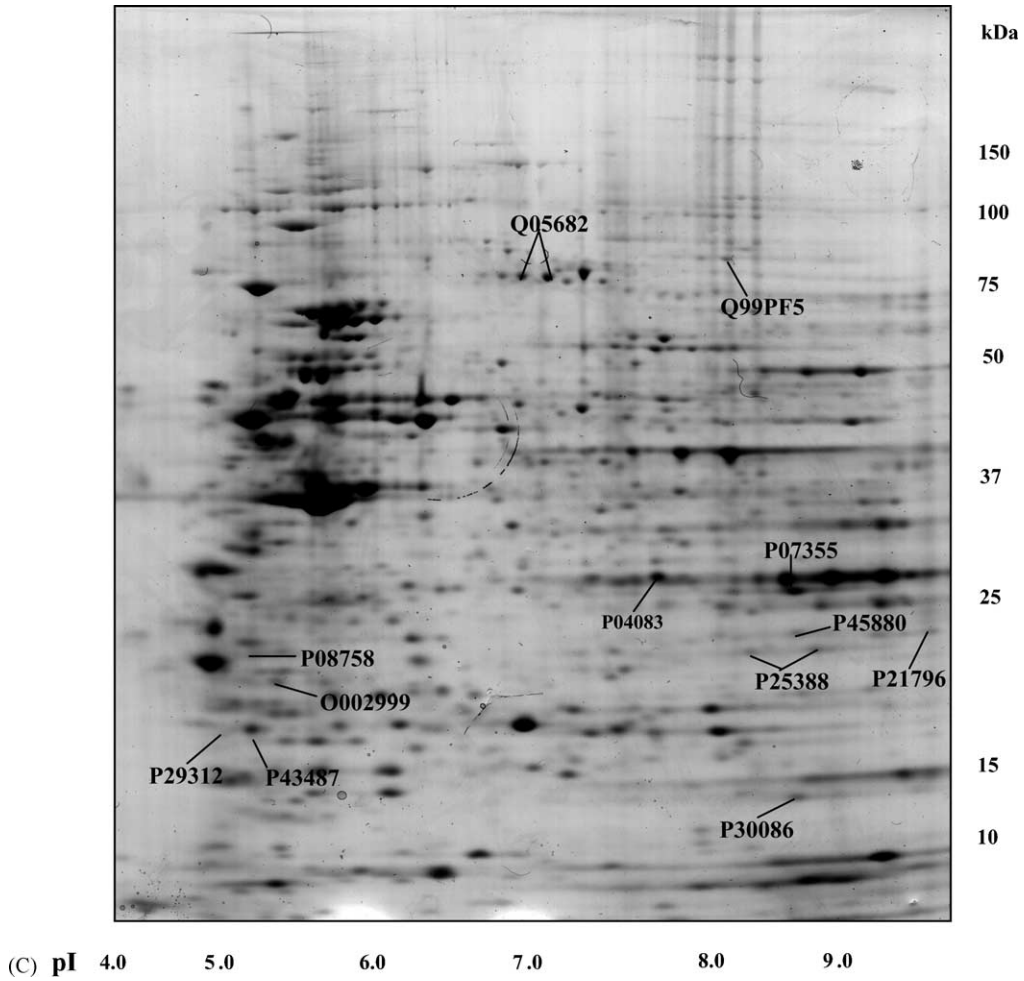

Fig. 1. (*Continued*)

There are a number of different isoforms distributed in a tissue-specific manner in a wide variety of organisms.

Similar to calsyntenin 1 nucleic acid sequence has first been identified from brain (http://www.us.expasy.org/prot/) and also calsyntenin 1 has been found released from synapse forming neurons [20] being localized in the postsynaptic membrane of both excitatory and inhibitory synapses. By binding calcium calsyntenin1 has been proposed to modulate calcium mediated postsynaptic signals [20].

3.2.3. Hypothetical armadillo repeat/plakoglobulin ARM-repeat profile containing protein

The sequence encoding this protein belongs to the HBG036800 gene family comprising eight sequences, two from human and six from mouse. The gene has been mapped to mouse chromosome 8 and shows approximately 83% identity to human gene MGC19595 located 19p13 (http://www.informatics.jax.org). DNA sources include neuronal tissue as well as samples derived from immunoendocrine organs (UNIGENE: http://www.us.expasy.org/prot/).

The armadillo repeat fold (IPR008938) contained in this sequence is found in 3213 proteins. This multihelical protein fold may contain any of a number of repeats or domains, including the armadillo domain, and the phosphoinositide 3-kinase accessory domain. The fold forms a right-handed superhelix. The armadillo repeat is an approximately 40 amino acid long tandemly repeated sequence motif first identified in the drosophila melanogaster segment polarity gene armadillo. Similar repeats were later found in the mammalian armadillo homolog beta-catenin, the junctional plaque protein plakoglobin, the adenomatous polyposis coli (APC) tumor suppressor protein, and a number of other proteins The three-dimensional fold of an armadillo repeat is known from the crystal structure of beta-catenin. There, the 12 repeats form a superhelix of alpha-helices, with three helices per unit. The cylindrical structure features a positively charged grove, which presumably interacts with the acidic surfaces of the known interaction partners of beta-catenin.

ARM repeat proteins function in various processes, including intracellular wnt-(winglesstype) signaling, required in embryonic development, and cytoskeletal regulation. Beta-catenin and its homologs are thought to act as

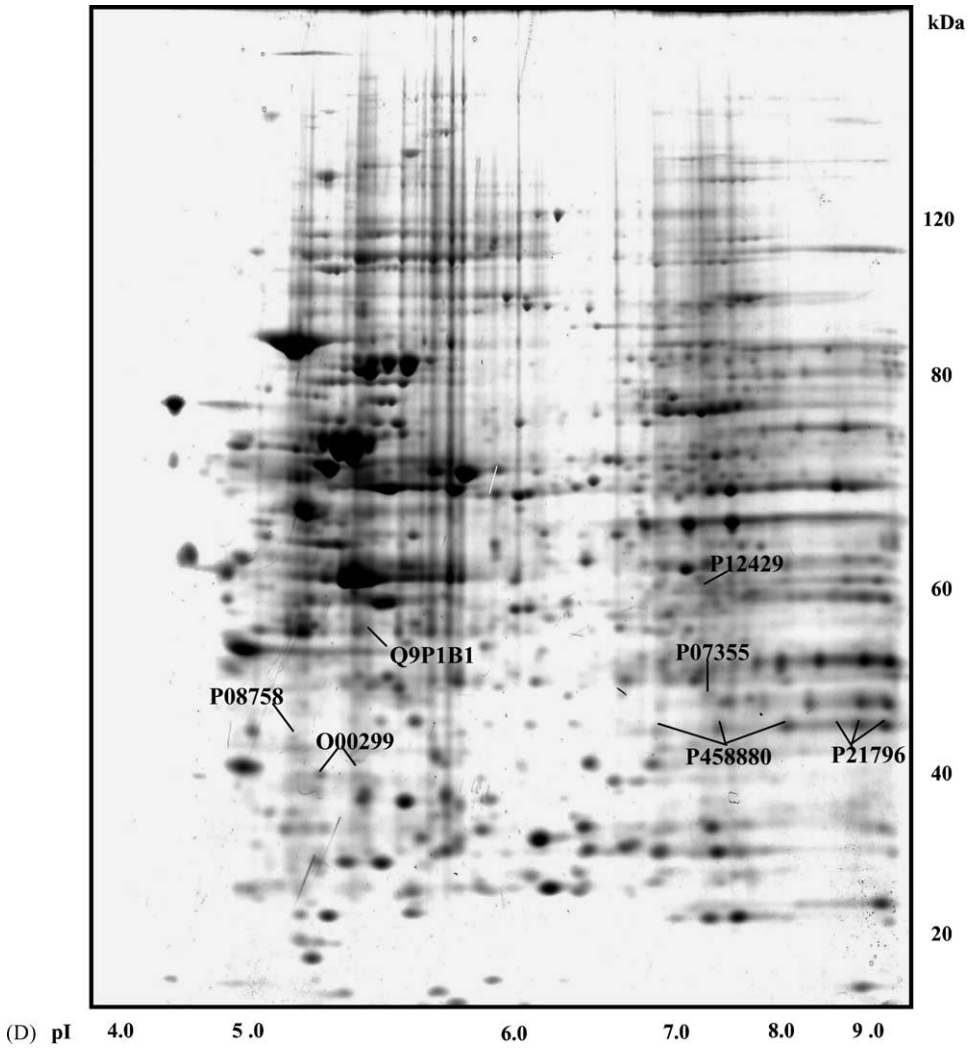

Fig. 1. (*Continued*)

regulators of gene expression, both during development and throughout adult life. These proteins can enter the nucleus in response to extracellular signals and bind to DNA in a complex with T-cell factor (TCF) transcription factors, thus altering gene expression.

3.2.4. 11 Days embryo cDNA clone 2700084k13

The gene encoding this protein is located on mouse chromosome 9 and shows 97% similarity to human annexin II gene mapped 15q21-q22 (http://www.informatics.jax.org).

Containing four annexin repeats (IPR001464) it belongs to the annexin family, which comprises 134 sequences of 18 taxons.

The annexins are a family of proteins representing calcium signaling, i.e. binding to phospholipids in a calcium-dependent manner. The binding is specific for calcium and for acidic phospholipids. Annexins have been claimed to be involved in cytoskeletal interactions, phospholipase inhibition, intracellular signaling, anticoagulation, and membrane fusion. There are eleven distinct classes of annexins, each of which has an amino acid sequence consisting of an N-terminal 'arm' followed by either four or eight copies of a conserved domain of 61 residues (only one of these residues, an arginine, is conserved among all copies). The calcium binding sites are found within the repeated domains. Individual repeats (sometimes known as endonexin folds) consist of five alpha helices wound into a right-handed superhelix. Each annexin class is thought to have a specific function, although for some the precise role is unclear. It has been suggested that the N-terminal residues confer the functional specificity that differentiates each class.

Annexin II a cell membrane organizing compound has been implicated in membrane trafficking and suggested as a regulator of cellular differentiation. It seems to inhibit PKC activity, possibly by regulating the various PKC isoforms.

3.2.5. Hypothetical protein flj22171 (Epidermal growth factor receptor pathway substrate 8 related protein 2)

The sequence encoding this protein belongs to the HBG003090 gene family comprising 17 sequences of three taxons, 11 of which have been found in human.

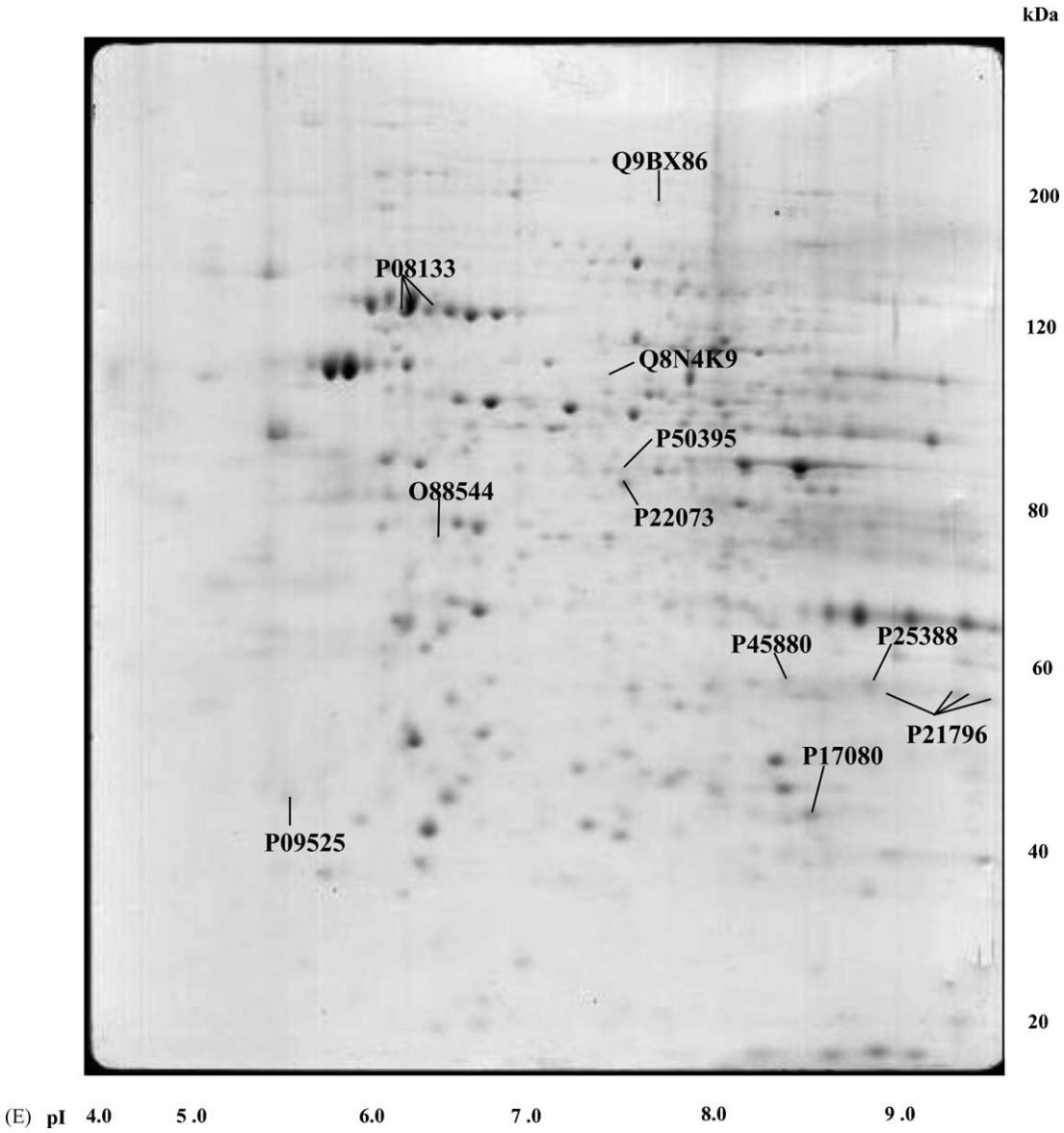

Fig. 1. (*Continued*)

This protein contains two domains, the phosphotyrosine interaction domain (IPR006020) and the SH3 domain (IPR001452) leading to activation of Rac, belonging to the Rho-family of GTPases.

Besides SH2, the phosphotyrosine interaction domain (PID or PI domain) is the second phosphotyrosine-binding domain found in the transforming protein Shc. Shc couples activated growth factor receptors to a signaling pathway that regulates the proliferation of mammalian cells and it might participate in the transforming activity of oncogenic tyrosine kinases. The PID domain of Shc specifically binds to the Asn–Pro–Xaa–Tyr (P) motif found in many tyrosine-phosphorylated proteins including growth factor receptors.

PID has an average length of about 160 amino acids. It is probably a globular domain with an antiparallel beta sheet.

src Homology-3 (SH3) domains are small protein modules containing approximately 50 amino acid residues. They are found in a great variety of intracellular or membrane-associated proteins for example, in a variety of proteins with enzymatic activity, in adaptor proteins that lack catalytic sequences and in cytoskeletal proteins, such as fodrin and yeast actin binding protein ABP-1.

The SH3 domain has a characteristic fold, which consists of five or six β-strands arranged as two tightly packed anti-parallel β sheets. The ligand binds with low affinity but this may be enhanced by multiple interactions. The region bound by the SH3 domain is in all cases proline-rich and contains PXXP as a core-conserved binding motif. The SH3 domain is perhaps the best-characterized member of the growing family of protein interaction molecules, which plays a critical role in a wide variety of biological

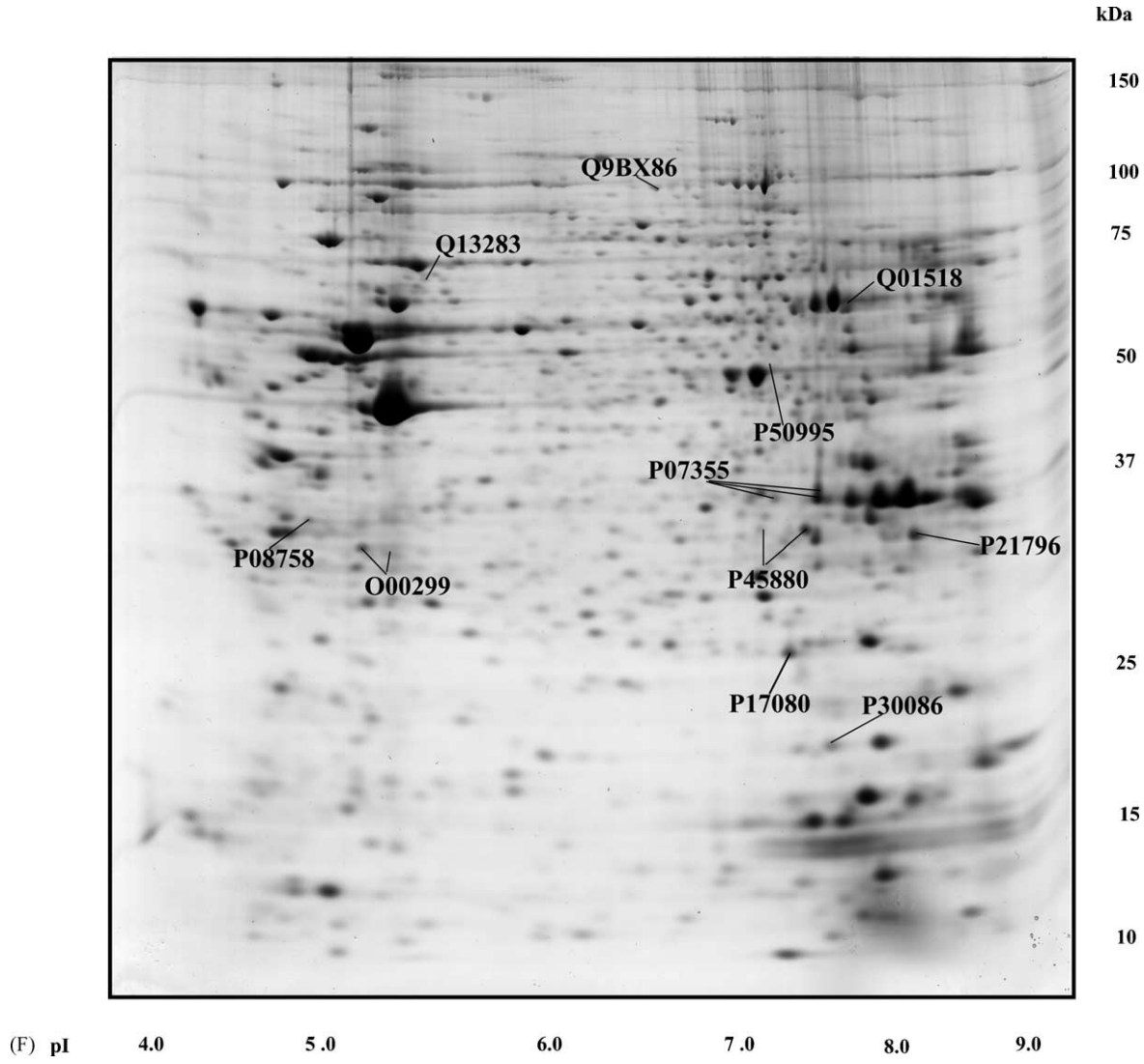

Fig. 1. (*Continued*)

processes ranging from regulation of enzymes by intramolecular interactions to altering the subcellular concentration and localization of components of signaling pathways.

SH3 domains and their binding site have cropped up in many hundreds of proteins from yeast to human, which suggest that they provide the cell with an especially handy and adaptable means of bringing proteins together.

4. Discussion

The major findings of this study are experimental evidence for the existence of five sp that have been predicted so far based upon nucleic acid sequences only.

Secondly we have generated an expressional pattern of signaling proteins in several human cell lines (see Table 1A) and thirdly, we provide an analytical tool for determination and characterisation of signaling structures.

Experimental data on expression of sp contribute to knowledge on calcium signaling, wnt-signaling and the src-signaling cascade by providing new members. Expression of these five structures was cell line specific and observed in amnion, lymphocyte and bronchial epithelial cell lineages. No individual isoforms were observed, i.e. for one protein only one spot was assigned. As shown in Table 1B observed p*I*s were different from the p*I*s predicted from databanks (Swiss-Prot/Trembl: http://www.us.expasy.org/prot/) as, e.g. “similar to calsyntenin and hypothetical Armadillo repeat containing protein” showed a higher observed pI value than the predicted p*I* value. This maybe due to post-translational modifications and clearly indicates that prediction programs are of limited use when a protein is to be searched for in a two-dimensional gel. Methodologically, we were able to unambiguously identify the so far only predicted/hypothetical proteins by MALDI-MS and corresponding software. The detection of new sp by this method, however, is determined by the fact that this proteomic

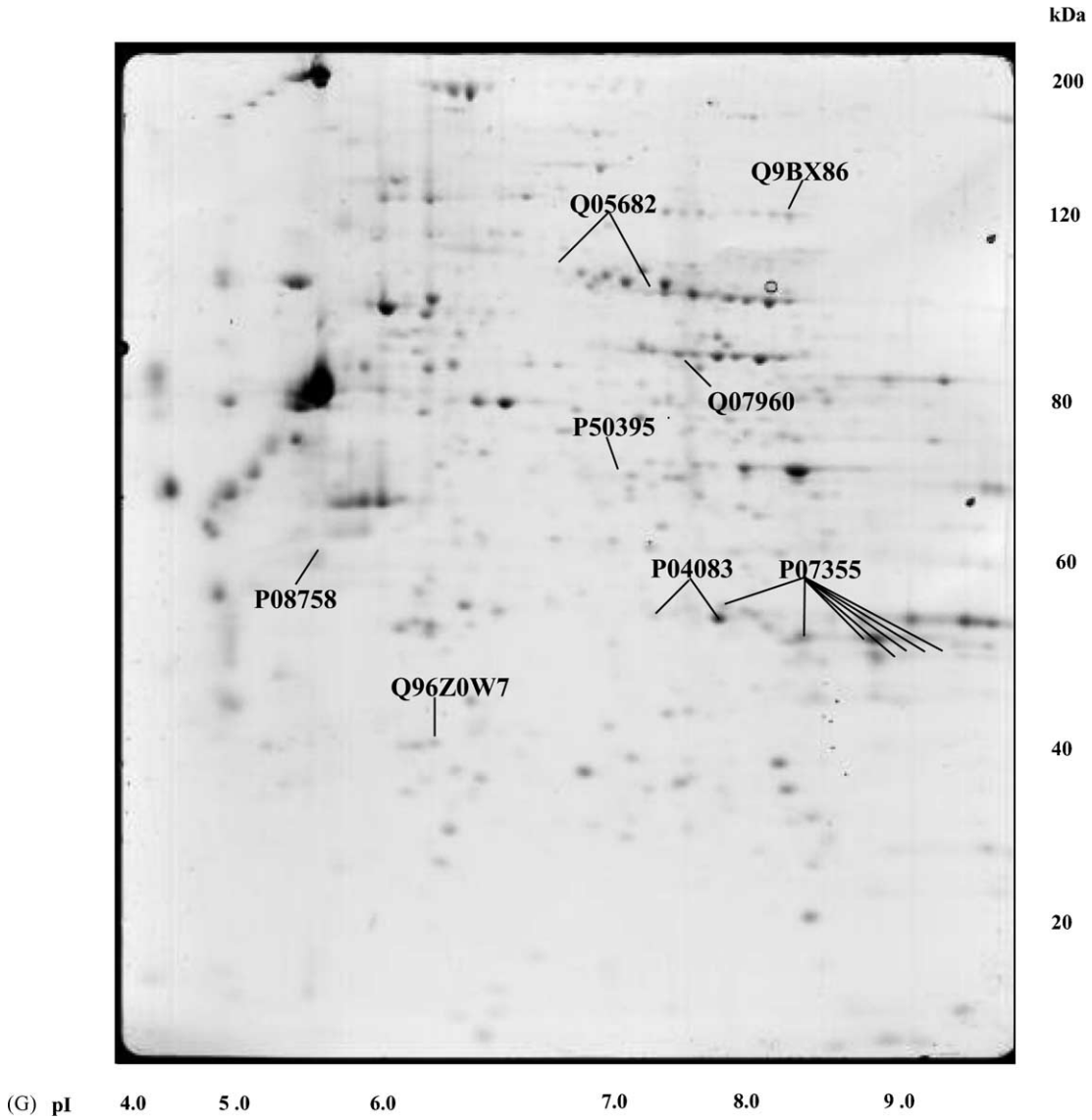

Fig. 1. (*Continued*).

technique fails to identify hydrophobic, highly acidic or basic proteins or low molecular weight proteins [23] and indeed the protein with the lowest molecular weight detected in this study had a molecular weight of 17,148 Da. Despite of these restrictions the use of this proteomic approach for protein hunting and in particular for the identification of predicted/hypothetical proteins is most useful and widely accepted [21–25].

Already known signaling structures shown in Table 1A were belonging to several major signaling pathways and cell-line specific expression was shown (Table 1A). Here it must be stated that only high abundance proteins, i.e. those that were stained by Coomassie blue, were picked from the gels and analysed. Therefore we do not exclude the possibility that the individual components of signaling cascades were expressed at lower levels in the corresponding cell lineages and thus remained undetected. The majority of sp was observed in one cell line exclusively probably pointing to different expressional levels thus possibly reflecting specific functions in individual cells.

Different culture conditions used for the cell lineages have to be taken into account as physiological changes are probably accompanied by gene expression reprogramming in response to the environment at the transcription, transcript processing, translation or posttranslational levels [26]. We decided to cultivate each cell line under corresponding optimal conditions widely used by the scientific community, to avoid artefacts by the use of uniform media resulting in suboptimal culturing. Seow et al. tested the effect of metabolic shifts on protein expression in mammalian cell culture by a comparable technique and reported aberrant structures remained unaffected [26].

The addition of recombinant epidermal growth to the kidney cell line (Materials and methods section) may well have led to phosphorylation/posttranslational modification [27]

of several proteins and this in turn may modify analytical determinants.

Likewise, no synchronisation of cell lines was carried out, as synchronisation would have been preventing to reach a sufficient amount of cells for analysis. Furthermore, cell lines at comparable passage numbers were not available from major sources as, e.g. ATCC. There is so far no published evidence for passage- or growth phase-dependent signaling protein expression.

We have proven the existence of five so far hypothetical proteins that were assigned to calcium, WNT and SH3 signaling cascades. During the revision process of this manuscript Offenhäuser et al. [28] provided immunochemical evidence for the existence "of epidermal growth factor receptor pathway substrate 8 related protein 2" using a monoclonal antibody and we are herewith confirming their findings by a protein chemical technique. Moreover, an analytical tool for sp expressional studies was established. Cell-line specific expressional patterns for high abundance sp were revealed fitting the concept that highly differentiated individual cell types require specific expression of signaling elements for executing specific functions.

Acknowledgements

We are highly indebted to the Red Bull Company, Salzburg, Austria for generous financial support. We also thank Dr. P. Ambros, Children Cancer Research Institute (CCRI, St. Anna Kinderspital, Vienna) for providing the lymphocyte cell line, Dr. M. Endemann (University of Vienna, Department of Pediatrics), Prof. Dr. M. Hengstschläger and Dr. M. Rosner (University of Vienna, Department of Obstetrics and Gynecology, Prenatal Diagnosis and Therapy), Prof. Dr. Z. Szepfalusi (University of Vienna, Department of Pediatrics) for providing cell lines. We appreciate the skillful technical assistance of Kiseok Lee, M.Sc. (University of Vienna, Department of Pediatrics).

References

[1] P.J. Kennely, J. Biochem. 370 (2003) 373.
[2] C.H. Heldin, Stem Cells 19 (2001) 295.
[3] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, PNAS 97 (2000) 9390.
[4] K. Toyo-oka, A. Shionoya, M.J. Gambello, C. Cardoso, R. Leventer, H. L Ward, R. Ayala, L.H. Tsai, W. Dobyns, D. Ledbetter, S. Hirotsune, A. Wynshaw-Boris, Nat. Genet. 34 (2003) 274.
[5] L. Deng, M. Yang, S. Frund, T. Wessel, R.A. De Abreu, J.A. Tischfield, A. Sahota, Mol. Genet. Metab. 72 (2001) 260.
[6] C. Tomasetto, C. Moog-Lutz, C.H. Regnier, V. Schreiber, P. Basset, M.C. Rio, FEBS Lett. 373 (1995) 245.
[7] D.A. Arber, K.L. Chang, M.H. Lyda, V. Bedell, R. Spielberger, M.L. Slovak, Hum. Pathol. 34 (2003) 809.
[8] C.L. Chang, X.X. Zhu, D.H. Thoraval, D. Ungar, J. Rawwas, N. Hora, J.R. Strahler, S.M. Hanash, E. Radany, Nature 370 (1994) 335.
[9] A.L. Cozens, M.J. Yezzi, K. Kunzelmann, T. Ohrui, L. Chin, K. Eng, W.E. Finkbeiner, J.H. Widdicombe, D.C. Gruenerts, Am. J. Respir. Cell. Mol. Biol. 10 (1994) 38.
[10] E.O. Major, A.E. Miller, P. Mourrain, R.G. Traub, E. de Widt, J. Sever, Proc. Natl. Acad. Sci. U.S.A. 83 (1985) 1257.
[11] M.J. Ryan, G. Johnson, J. Kirk, S.M. Fuerstenberg, R.A. Zager, B. Torok-Storb, Kidney Int. 45 (2001) 48.
[12] K. Arbeiter, B. Bidmon, M. Endemann, T.O. Bender, O. Eickelberg, D. Ruffingshofer, T. Mueller, H. Regele, K. Herkner, C. Aufricht, Kidney Int. 60 (2001) 1930.
[13] M.M. Bradford, Anal. Biochem. 72 (1976) 248.
[14] M. Fountoulakis, H. Langen, Anal. Biochem. 250 (1997) 153.
[15] R. Weitzdoerfer, M. Fountoulakis, G. Lubec, Biochem. Biophys. Res. Commun. 293 (2002) 836.
[16] P. Berndt, U. Hobohm, H. Langen, Electrophoresis 20 (1999) 3521.
[17] K. Krapfenbauer, M. Berger, A. Friedlein, G. Lubec, Eur. J. Biochem. 268 (2001) 3532.
[18] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 5011.
[19] T. Imai, K. Matsuda, T. Shimojima, T. Hashimoto, Y. Masuhiro, T. Kitamoto, A. Sugita, K. Suzuki, H. Matsumoto, H. Masushige, et al., Biochem. Biophys. Res. Commun. 233 (1997) 765.
[20] L. Vogt, S.P. Schrimpf, V. Meskenaite, R. Frischknecht, J. Kinter, D.P. Leone, U. Ziegler, P. Sonderegger, Mol. Cell. Neurosci. 17 (1994) 151.
[21] G. Lubec, K. Krapfenbauer, M. Fountoulakis, Prog. Neurobiol. 69 (2003) 193.
[22] E. Engidawork, T. Gulesserian, M. Fountoulakis, G. Lubec, Mol. Genet. Met. 78 (2003) 295.
[23] J.K. Myung, T. Gulesserian, M. Fountoulakis, G. Lubec, Cell. Mol. Biol. 49 (2003) 739.
[24] M.S. Cheon, M. Fountoulakis, M. Dierssen, J.C. Ferreres, G. Lubec, J. Neural Transm. Suppl. 61 (2001) 311.
[25] A. Peyrl, K. Krapenbauer, I. Slavc, J.W. Yang, T. Strobel T, G. Lubec, Proteomics 3 (2003) 1481.
[26] T.K. Seow, R. Korke, R.C.M.Y. Liang, S.E. Ong, K. Ou, K. Wong, W.S. Hu, M.C.M. Chung, Biotechnol. Prog. 17 (2001) 1137.
[27] B.E. Peace, K.J. Hill, S.J. Degen, S.E. Waltz, Exp. Cell. Res. 289 (2003) 317.
[28] N. Offenhäuser, A. Borgonovo, A. Disanza, P. Romano, I. Ponzanelli, G. Iannolo, P.P. Di Fiore, S. Scita, Mol. Biol. Cell 15 (2004) 91.