Amino Acids (2004) 27: 249–259
DOI 10.1007/s00726-004-0141-1

Amino Acids
Printed in Austria

Depletion of the high-abundance plasma proteins

M. Fountoulakis*, **J.-F. Juranville, L. Jiang, D. Avila, D. Röder, P. Jakob, P. Berndt, S. Evers,** and **H. Langen**

F. Hoffmann-La Roche Ltd., Center of Medical Genomics, Basel, Switzerland

Received August 25, 2004
Accepted October 4, 2004
Published online December 2, 2004;

Summary. Body fluids, like plasma and urine, are comparatively easy to obtain and are useful for the detection of novel diagnostic markers by applying new technologies, like proteomics. However, in plasma, several high-abundance proteins are dominant and repress the signals of the lower-abundance proteins, which then become undetectable either by two-dimensional gels or chromatography. Therefore, depletion of the abundant proteins is a prerequisite for the detection of the low-abundance components. We applied affinity chromatography on blue matrix and Protein G and removed the most abundant human plasma proteins, albumin and the immunoglobulin chains. The plasma proteins, prior to albumin and immunoglobulin depletion, as well the eluates from the two chromatography steps were analyzed by two-dimensional electrophoresis and the proteins were identified by MALDI-TOF-MS. The analysis resulted in the identification of 83 different gene products in the untreated plasma. Removal of the high-abundance proteins resulted in the visualization of new protein signals. In the eluate of the two affinity steps, mostly albumin and immunoglobulin spots were detected but also spots representing several other abundant plasma proteins. The methodology is easy to perform and is useful as a first step in the detection of diagnostic markers in body fluids by applying proteomics technologies.

Keywords: Affinity chromatography – Albumin depletion – Antibody chains depletion – Diagnostic markers – High-abundance proteins – Mass spectrometry – Plasma – Proteomics

Introduction

Proteomics studies the proteins of biological systems in a high-throughput mode and has as goal the detection of novel drug targets and diagnostic markers. A proteomic analysis usually employs two-dimensional electrophoresis or multi-dimensional chromatography for protein separation and mass spectrometry for protein identification. In general, in the small sample volume (about 10–300 μl) that is usually used in a proteomic analysis, a large percentage of the expressed proteins are not present in sufficient quantities to be detected, so that certain low-abundance proteins can not be readily detected during the analysis of total proteins (Fountoulakis, 2001). Nevertheless, the study of the low-copy-number gene products is of highest importance as such proteins are the most likely drug targets and diagnostic markers. Low-abundance proteins can be enriched from crude extracts by biochemical protein-enriching approaches, where the original protein mixture is separated into less complex fractions, each containing a lower number of total proteins in comparison with the starting material. Enrichment of proteins from larger volumes can be achieved by selective fractionation, chromatography, electrophoretic procedures or by a combination of the various methods (Fountoulakis and Takács, 2002; Fountoulakis et al., 2004).

Plasma is very interesting from the medical point of view as most cells communicate with it and many cells release at least part of their content into the plasma upon damage or death (Anderson et al., 2004). Whereas plasma can be easily obtained and the information it comprises is immense, it is one of the most difficult samples to analyze by applying proteomics tools (Anderson and Anderson, 2002). The reason is that it contains about ten high-abundance proteins which together represent about 97% of the total plasma proteins. Probably, less than 1% of all proteins are prime targets for the identification of novel markers (Zolg and Langen, 2004). The most abundant proteins are albumin and the immunoglobulin heavy and light chains, representing together about 80% of the plasma proteins. Therefore, depletion of at least these two high abundance proteins is a prerequisite of a successful search for disease markers in plasma.

* Current address: Foundation for Biomedical Research of the Academy of Athens, Greece

There exist several studies dealing with the proteomic analysis of plasma and with the depletion of abundant proteins, involving selective fractionation (Jiang et al., 2004), chromatography steps (Barroso et al., 2003; Govorukhina et al., 2003; Pieper et al., 2003a and 2003b) and various proteomics approaches (Veenstra and Conrads, 2003; Vlahou et al., 2003; Zhang et al., 2004). By using 2-D gels for protein visualization, about 60–80 different proteins are usually detected. Through combination of protein enriching steps and 2-D electrophoresis, 325 different proteins were identified in plasma (Pieper et al., 2003a). Here we present an updated 2-D protein map of untreated plasma and report a simple approach involving two chromatography steps for the depletion of the two major plasma components, albumin and IgG chains.

Materials and methods

Materials

Immobilized pH-gradient (IPG) strips and IPG buffers were purchased from Amersham Biosciences (Uppsala, Sweden). Acrylamide/piperazine-diacrylamide (PDA) solution (37.5:1, w/v) was purchased from Biosolve Ltd. (Valkenswaard, The Netherlands) and the other reagents for the polyacrylamide gel preparation were from Bio-Rad Laboratories (Hercules, CA, USA). CHAPS was obtained from Roche Diagnostics (Mannheim, Germany), urea from AppliChem (Darmstadt, Germany), thiourea from Fluka (Buchs, Switzerland), 1,4-dithioerythritol (DTE) and EDTA from Merck (Darmstadt, Germany) and tributylphosphine (TBP) from Pierce Biotechnology (Rockford, IL, USA). The reagents were kept at 4°C. HiTrap blue columns were obtained from Amersham Biosciences. Plasma samples were from controls, EDTA-treated and were stored at −80°C until use. The protein content was determined with the Coomassie blue method (Bradford, 1976). The average plasma protein concentration was 65 mg/ml.

Chromatography on Mimetic blue and Protein G

Protease inhibitors cocktail Complete™ (Roche Diagnostics) was added to the plasma (one tablet to 50 ml of volume) upon thawing. 40 mg of plasma (0.6 ml) was diluted ten-fold with 25 mM MES, pH 6.0, to reduce the salt concentration and to adjust the pH to about 6.0. Albumin was removed by chromatography on Mimetic blue SA P6XL (2 ml, ProMetic BioSciences Ltd.) and the IgG chains were removed by chromatography on HiTrap Protein G HP (1 ml, Amersham Biosciences). The two columns were connected in series and equilibrated with 25 mM MES, pH 6.0. The diluted plasma solution was filtered through a 0.22 μm filter and applied onto the Mimetic blue column at 1 ml/min. The flow through of this column was directly loaded onto the Protein G column and the flow-through fraction from the latter column was collected (about 6 mg totally, 16%). The two columns were washed with 10 ml of 25 mM MES, pH 6.0 and then separated. The Mimetic blue column was eluted with a step gradient of 2 M NaCl in 50 mM Tris-HCl, pH 7.5 and the Protein G was first eluted with 50 mM Tris-HCl, pH 7.5, containing 1 M NaCl and then with 100 mM glycine-HCl, pH 2.8 and the eluate was neutralized with 1 M Tris base. The two eluates were desalted by acetone precipitation (Jiang et al., 2004). The flow through fraction and the eluates were analyzed by two-dimensional gels and the proteins were identified by MALDI-TOF-MS.

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis was performed essentially as reported (Langen et al., 1997). Samples of 750 μg were applied on immobilized pH 3–7 or pH 3–10 nonlinear IPG strips. Focusing started at 200 V and the voltage was gradually increased to 5000 V at 3 V/min and kept constant for a further 6 h. The second-dimensional separation was performed in 12% SDS-polyacrylamide gels. The gels (180 × 200 × 1.5 mm) were run at 50 mA per gel, in an ETTAN DALT II apparatus (Amersham Biosciences). After protein fixation with 50% methanol, containing 5% phosphoric acid for 2 h, the gels were stained with colloidal Coomassie blue (Novex, San Diego, CA, USA) for 16 h. Excess of dye was washed out from the gels with H_2O and the gels were scanned in an Agfa DUOSCAN densitometer (resolution 400). Protein spots were quantified using the ImageMaster 2-D Elite software (Amersham Biosciences). The percentage of the volume of the spots representing a certain protein was determined in comparison with the total proteins present in the 2-D gel.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS)

MALDI-TOF-MS analysis was essentially performed as described (Fountoulakis and Gasser, 2003; Jiang et al., 2003; Fountoulakis, 2004). The spots were excised from 2-D gels, destained with 30% acetonitrile in 50 mM ammonium bicarbonate and dried in a Speedvac evaporator. Each dried gel piece was rehydrated with 5 μl of 1 mM ammonium bicarbonate, containing 50 ng trypsin (Roche Diagnostics). After 16 h at room temperature, 20 μl of 50% acetonitrile, containing 0.3% trifluoroacetic acid were added to each gel piece and incubated for 15 min with constant shaking. The sample application to the sample target was performed with a CyBiWell apparatus (Cybio AG, Jena, Germany). Peptide mixture (1.5 μl) was simultaneously applied with 1 μl of matrix solution, consisting of 0.025% α-cyano-4-hydroxycinnamic acid (Sigma) and the standard peptides des-Arg-bradykinin (Sigma, 904.4681 Da) and adrenocorticotropic hormone fragment 18–39 (Sigma, 2465.1989 Da) in 65% ethanol, 35% acetonitrile and 0.03% trifluoroacetic acid. Samples were analyzed in a time-of-flight mass spectrometer (Ultraflex, Bruker Daltonics, Bremen, Germany). Peptide matching and protein searches were performed automatically with the use of in-house developed software (Berndt et al., 1999). 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% or lower was allowed. The probability of a false positive match with a given MS-spectrum was determined for each analysis. Unmatched peptides or miscleavage were not considered.

Results

Two-dimensional electrophoretic analysis of plasma

Prior to chromatographic treatment, plasma was analyzed by 2-D gels, using narrow pH 3–7 and broad pH 3–10 non-linear IPG strips. Because albumin amounts up to about 50% of total proteins and because a minimal protein quantity should be applied per gel for the visualization and the identification of a large number of spots (0.5–1.0 mg of total protein per gel stained with Coomassie blue), various protein amounts were applied, 0.10, 0.25, 0.50, 0.75 and 1.00 mg, in order to find the optimal load quantity. When 0.10 and 0.25 mg were loaded and the gels

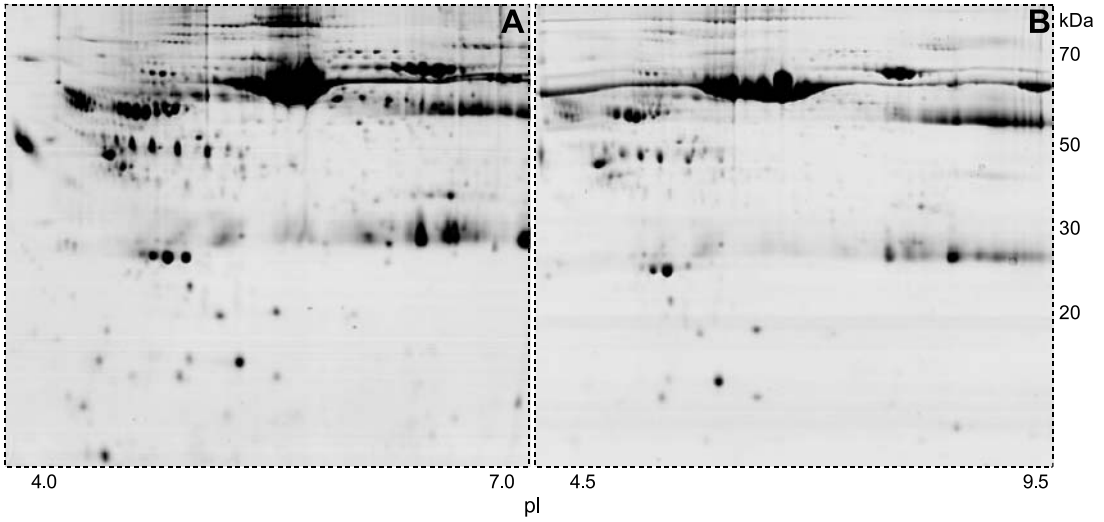

Fig. 1. Two-dimensional electrophoresis analysis of plasma proteins on pH 3–7 non-linear (**A**) and on pH 3–10 non-linear (**B**) IPG strips, followed by 12% SDS-gels. 0.75 mg protein was applied on each gel. The gels were stained with Coomassie blue. A higher spot resolution was achieved with the narrow pH range strip

were stained with Coomassie blue, albumin was the predominant spot and a relative small number of minor spots were detected. Upon loading of 1 mg, the albumin amount was large, usually caused artifacts and influenced the quality of the 2-D gels. Application of 0.50 to 0.75 mg appeared to represent the best choice for gels to be stained with Coomassie blue (data not shown). Figure 1A and 1B show representative images of the plasma proteins analyzed on narrow pH 3–7 and broad pH 3–10 IPG strips, respectively, on which 0.75 mg of protein was applied. In the narrow pH range gel, a more efficient spot resolution was usually obtained and a larger number of spots in the acidic region were resolved (Fig. 1A).

Protein identification

The proteins were identified by MALDI-TOF-MS on the basis of peptide mass matching (Henzel et al., 1993), following in-gel digestion with trypsin. About 400 spots were excised from one 2-D gel. Each excised spot was analyzed individually. The peptide masses were matched with the theoretical peptide masses of all known proteins from all species. 83 different gene products were identified in the unfractionated plasma. The proteins identified are listed in Table 1 together with their theoretical MW and pI values and data from the mass spectrometry analysis, i.e. the numbers of matching peptides and the probability that the protein identity assigned could be random. The spots representing the proteins identified are shown in Fig. 2 and are labeled with their abbreviated names. The large spots represent albumin (ALBU), IgG heavy and light chains (GC1, GC2, KAC), serotransferrin (TRFE), alpha-1 antitrypsin (A1AT), haptoglobin 1 (HPT1), macroglobulin (A2MG), fibrinogen (FIBB), and apolipoprotein A-1 (APA1) (Fig. 2).

The relatively low number of different gene products detected is due to the presence of the high-abundance proteins and to their high heterogeneity. Most of the proteins are represented by a large number of spots. The highest heterogeneity showed albumin which is represented by a very strong spot corresponding to the full-length protein and many weaker spots with higher and lower masses (Fig. 2). Excluding database redundancies and considering the various immunoglobulin chains as one protein, the number of the identified different gene products is reduced to 68.

Plasma fractionation

We studied various approaches to remove the high-abundance proteins, albumin and immunoglobulin chains and the reproducibility of the method. Various chromatography steps were tested, including affinity chromatography on immobilized monoclonal anti-albumin antibodies and matrix blue gels. The scheme chosen comprises two affinity steps, matrix blue and Protein G. For albumin removal, we tested two kinds of blue matrix gels, Mimetic blue and HiTrap blue. Several protein concentration and pH conditions were studied. Application of 20 mg protein per 1 ml of bed column at pH 6.0 was found to result in optimal

Table 1. Human plasma proteins. Proteins from the plasma of controls were analyzed by 2-D gel electrophoresis and identified by MALDI-TOF-MS, following in-gel digestion with trypsin as described under Materials and methods. The search in protein databases was performed with in house developed software. The number of matching peptides is listed in Table 1 (matches). The spots representing the identified proteins are indicated in Figs. 2, 4 and 5 and are designated with their abbreviated names of SWISS-PROT or of the other databases. The theoretical Mr and pI values are given

Abbr. name	Acc. number	Protein	Full name	pI	MW	Matches	Probability
74561	074561-17-0	HSUGP:074561-17-0	Alpha-2-macroglobulin	5.85	57172	7	2.41E – 05
74561	074561-19-0	HSUGP:074561-19-0	Alpha-2-macroglobulin	6.75	174362	10	7.83E – 10
151242	151242-19-0	HSUGP:151242-19-0	Serine (or cysteine) proteinase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	6.85	57709	5	1.16E – 04
151242	151242-21-0	HSUGP:151242-21-0	Serine (or cysteine) proteinase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	7.59	59409	5	1.12E – 04
170250	170250-1-0	HSUGP:170250-1-0	Complement component 4A	7.00	188945	6	1.64E – 06
272572	272572-6-0	HSUGP:272572-6-0	Hemoglobin, alpha 1	9.78	35093	5	1.18E – 04
278625	278625-10-0	HSUGP:278625-10-0	Complement component 4B	8.33	28988	6	2.49E – 06
284394	284394-20-0	HSUGP:284394-20-0	Complement component 3	5.13	33683	6	1.86E – 06
296634	296634-13-0	HSUGP:296634-13-0	Ceruloplasmin (ferroxidase)	5.30	120513	7	2.61E – 06
297681	297681-31-0	HSUGP:297681-31-0	Serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member	5.28	45835	5	2.17E – 04
Q13677	Q13677	SWTR_HUM:Q13677	Alpha-2-macroglobulin (fragment)	5.50	71321	6	1.70E – 06
Q96KX8	Q96KX8	SWTR_HUM:Q96KX8	Hypothetical 53.4 kDa protein	7.71	54327	5	2.71E – 05
A1AT	P01009	SW:A1AT_HUMAN	Alpha-1-antitrypsin precursor (alpha-1 protease inhibitor) (alpha-1-antiproteinase)	5.40	46878	9	3.59E – 12
A1BG	P04217	SW:A1BG_HUMAN	Alpha-1b-glycoprotein	5.90	52478	11	7.34E – 20
A2GL	P02750	SW:A2GL_HUMAN	Leucine-rich alpha-2-glycoprotein (lrg)	5.93	34553	6	1.79E – 08
A2MG	P01023	SW:A2MG_HUMAN	Alpha-2-macroglobulin precursor (alpha-2-m)	6.40	164600	17	2.44E – 24
AACT	P01011	SW:AACT_HUMAN	Alpha-1-antichymotrypsin precursor (act)	5.29	47791	6	4.15E – 07
AFAM	P43652	SW:AFAM_HUMAN	Afamin precursor (alpha-albumin) (alpha-alb)	5.77	70962	10	3.10E – 12
ALBU	P02768	SW:ALBU_HUMAN	Serum albumin precursor	6.23	71317	9	1.58E – 13
ALC1	P01876	SW:ALC1_HUMAN	Ig alpha-1 chain c region	6.48	38485	7	1.89E – 09
ALS	P35858	SW:ALS_HUMAN	Insulin-like growth factor binding protein complex acid labile chain precursor (als)	6.77	66735	11	1.95E – 18
ANGT	P01019	SW:ANGT_HUMAN	Angiotensinogen precursor	6.28	53405	7	1.01E – 09
ANT3	P01008	SW:ANT3_HUMAN	Antithrombin-iii precursor (atiii)	6.68	53025	7	2.17E – 07
APA1	P02647	SW:APA1_HUMAN	Apolipoprotein A-I precursor (apo-ai)	5.59	30758	7	2.87E – 09
APA4	P06727	SW:APA4_HUMAN	Apolipoprotein A-IV precursor (apo-aiv)	5.17	45343	5	8.89E – 05
APOH	P02749	SW:APOH_HUMAN	Beta-2-glycoprotein i precursor (apolipoprotein h) (apo-h) (activated protein c-binding protein) (apc inhibitor)	7.86	39584	8	4.19E – 12
C1QB	P02746	SW:C1QB_HUMAN	Complement C1q subcomponent, b chain precursor	8.86	26670	5	1.71E – 05
C1R	P00736	SW:C1R_HUMAN	Complement C1r component precursor (EC 3.4.21.41)	6.25	81661	12	3.67E – 17
C1S	P09871	SW:C1S_HUMAN	Complement C1s component precursor (EC 3.4.21.42) (c1 esterase)	4.69	78174	9	4.80E – 12
C4BP	P04003	SW:C4BP_HUMAN	C4b-binding protein alpha chain precursor (proline-rich protein) (prp)	7.22	69042	12	3.80E – 20

CAH1	P00915	SW:CAH1_HUMAN	Carbonic anhydrase I (EC 4.2.1.1) (carbonate dehydratase i)	7.14	28778	4	1.19E – 05
CBP8	P22792	SW:CBP8_HUMAN	Carboxypeptidase N 83 kDa chain (carboxypeptidase n regulatory subunit) (fragment)	6.23	59410	7	2.07E – 08
CBPN	P15169	SW:CBPN_HUMAN	Carboxypeptidase N catalytic chain precursor (EC 3.4.17.3) (arginine carboxypeptidase) (plasma carboxypeptidase b)	7.34	52538	6	4.32E – 06
CERU	P00450	SW:CERU_HUMAN	Ceruloplasmin precursor (EC 1.16.3.1) (ferroxidase)	5.57	122982	11	2.48E – 10
CFAB	P00751	SW:CFAB_HUMAN	Complement factor b (EC 3.4.21.47) (c3/c5 convertase) (properdin factor b) (glycine-rich beta glycoprotein) (gbg)	7.03	86847	11	1.27E – 17
CFAH	P08603	SW:CFAH_HUMAN	Complement factor h precursor	6.64	143709	8	3.38E – 07
CFAI	P05156	SW:CFAI_HUMAN	Complement factor i precursor (EC 3.4.21.45) (c3b/c4b inactivator)	7.39	68071	7	5.51E – 07
CLUS	P10909	SW:CLUS_HUMAN	Clusterin (complement-associated protein sp-40,40) (complement cytolysis inhibitor) (apolipoprotein j) (apo-j)	6.23	53031	7	1.01E – 08
CO3	P01024	SW:CO3_HUMAN	Complement C3 precursor [contains: c3a anaphylatoxin]	6.35	188585	16	1.53E – 24
CO4	P01028	SW:CO4_HUMAN	Complement C4 precursor [contains: c4a anaphylatoxin]	7.17	193811	9	3.16E – 08
CO6	P13671	SW:CO6_HUMAN	Complement component C6 precursor	6.64	108425	9	4.10E – 12
CO7	P10643	SW:CO7_HUMAN	Complement component C7 precursor	6.43	96646	5	1.23E – 05
CO8B	P07358	SW:CO8B_HUMAN	Complement component C8 beta chain precursor	8.01	68714	12	1.21E – 19
CO9	P02748	SW:CO9_HUMAN	Complement component C9 precursor	5.37	64615	5	4.21E – 05
F13B	P05160	SW:F13B_HUMAN	Coagulation factor xiii b chain precursor (EC 2.3.2.13) (protein-glutamine gamma-glutamyltransferase b chain)	6.33	77722	7	3.63E – 07
FBL1	P23142	SW:FBL1_HUMAN	Fibulin-1 precursor	5.01	81328	6	1.47E – 08
FCN3	O75636	SW:FCN3_HUMAN	Ficolin 3 (collagen/fibrinogen domain-containing protein 3) (collagen/fibrinogen domain-containing lectin 3 p35)	6.66	33381	7	4.46E – 10
FIBB	P02675	SW:FIBB_HUMAN	Fibrinogen beta chain precursor	8.26	56576	7	4.36E – 07
GC1	P01857	SW:GC1_HUMAN	Ig gamma-1 chain c region	8.17	36596	7	2.53E – 10
HBB	P02023	SW:HBB_HUMAN	Hemoglobin beta chain	7.31	15971	7	1.16E – 10
HC	P02760	SW:HC_HUMAN	Alpha-1-microglobulin/inter-alpha-trypsin inhibitor light chain precursor (iti-lc) (protein hc) (hi-30) (bikunin)	6.18	39886	4	1.03E – 05
HEMO	P02790	SW:HEMO_HUMAN	Hemopexin precursor (beta-1b-glycoprotein)	7.00	52384	9	2.30E – 14
HPT1	P00737	SW:HPT1_HUMAN	Haptoglobin-1 precursor	6.58	38940	6	2.71E – 07
HPT2	P00738	SW:HPT2_HUMAN	Haptoglobin-2 precursor	6.57	45860	5	7.37E – 05
HPTR	P00739	SW:HPTR_HUMAN	Haptoglobin-related protein precursor	6.88	39495	6	1.38E – 08
HRG	P04196	SW:HRG_HUMAN	Histidine-rich glycoprotein precursor (histidine-proline rich glycoprotein) (hprg)	7.48	60510	6	3.01E – 07
IC1	P05155	SW:IC1_HUMAN	Plasma protease C1 inhibitor precursor (c1 inh)	6.52	55347	11	6.45E – 18
ITH2	P19823	SW:ITH2_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H2 precursor (iti heavy chain h2) (serum-derived hyaluronan-associated protein)	6.86	106825	10	7.02E – 13
ITH3	Q06033	SW:ITH3_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H3 precursor (iti heavy chain h3) (serum-derived hyaluronan-associated protein)	5.79	99401	8	4.34E – 10

(*continued*)

Table 1 (*continued*)

Abbr. name	Acc. number	Protein	Full name	pI	MW	Matches	Probability
ITH4	Q14624	SW:ITH4_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (iti heavy chain h4) (plasma kallikrein sensitive glycoprotein 120)	6.97	103549	6	2.88E – 06
K1CI	P35527	SW:K1CI_HUMAN	Keratin, type i cytoskeletal 9 (cytokeratin 9) (k9) (ck 9)	5.00	62177	5	9.03E – 05
KAL	P03952	SW:KAL_HUMAN	Plasma kallikrein precursor (EC 3.4.21.34) (plasma prekallikrein) (kininogenin) (fletcher factor)	8.09	73432	7	1.58E – 07
KNH	P01042	SW:KNH_HUMAN	Kininogen, hmw precursor (alpha-2-thiol proteinase inhibitor) (contains: bradykinin)	6.80	72983	7	1.82E – 07
LUM	P51884	SW:LUM_HUMAN	Lumican precursor (lum) (keratan sulfate proteoglycan)	6.60	38716	8	2.23E – 12
MUC	P01871	SW:MUC_HUMAN	Ig mu chain c region	6.75	50209	7	4.50E – 08
MUCB	P04220	SW:MUCB_HUMAN	Ig mu heavy chain disease protein (bot)	5.00	43543	6	3.05E – 07
PLMN	P00747	SW:PLMN_HUMAN	Plasminogen precursor (EC 3.4.21.7)	7.17	93247	9	2.84E – 11
RETB	P02753	SW:RETB_HUMAN	Plasma retinol-binding protein precursor (prbp) (rbp)	5.43	23195	5	1.30E – 05
SAMP	P02743	SW:SAMP_HUMAN	Serum amyloid p-component precursor (sap) (9.5s alpha-1-glycoprotein)	6.53	25485	5	3.45E – 05
THRB	P00734	SW:THRB_HUMAN	Prothrombin precursor (EC 3.4.21.5) (coagulation factor ii)	5.76	71474	14	9.21E – 25
TRFE	P02787	SW:TRFE_HUMAN	Serotransferrin precursor (siderophilin) (beta-1-metal binding globulin)	7.06	79280	10	1.03E – 09
TTHY	P02766	SW:TTHY_HUMAN	Transthyretin precursor (prealbumin) (tbpa) (ttr) (attr)	5.64	15991	4	1.32E – 05
VTDB	P02774	SW:VTDB_HUMAN	Vitamin D-binding protein precursor (dbp) (group-specific component) (gc-globulin) (vdb)	5.32	54525	5	9.85E – 06
VTNC	P04004	SW:VTNC_HUMAN	Vitronectin precursor (serum spreading factor) (s-protein) [contains: somatomedin b]	5.61	55069	7	1.02E – 07
AAG00911	AAG00911	SWTR_HUM:AAG00911	Recombinant IgG3 heavy chain (Fragment){E1}	8.00	38699	6	5.61E – 07
BAA82105	BAA82105	SWTR_HUM:BAA82105	Anti-Entamoeba histolytica immunoglobulin kappa light chain (Fragment){E1}	8.05	23575	5	1.14E – 04
CAA01533	CAA01533	SWTR_HUM:CAA01533	Alpha 2-macroglobulin 690–740	6.45	165177	6	2.09E – 07
CAA67886	CAA67886	SWTR_HUM:CAA67886	Gamma 3 immunoglobulin constant heavy chain (fragment)	8.00	38658	7	6.88E – 09
CAC10234	CAC10234	SWTR_HUM:CAC10234	Immunoglobulin heavy chain (Fragment)	7.88	40491	7	2.06E – 06
CD5L	O43866	SW:CD5L_HUMAN	Cd5 antigen-like precursor (sp-alpha) (ct-2) (igm-associated peptide)	5.24	39602	10	1.49E – 16
Q08380	Q08380	HUMANGP:CHR17-Q08380	MAC-2 binding protein precursor (lectin, galactoside-binding, soluble, 3 binding protein) (galectin 6 binding protein)	5.00	66201	6	7.15E – 07
Q96PD5	Q96PD5	HUMANGP:CHR19-Q96PD5	Peptidoglycan recognition protein L precursor was used to identify this gene	7.51	62748	4	4.13E – 05
Q9Y220	Q9Y220	HUMANGP:CHR17-Q9Y220	Grb7V protein was used to identify this gene	8.74	50190	5	1.31E – 04

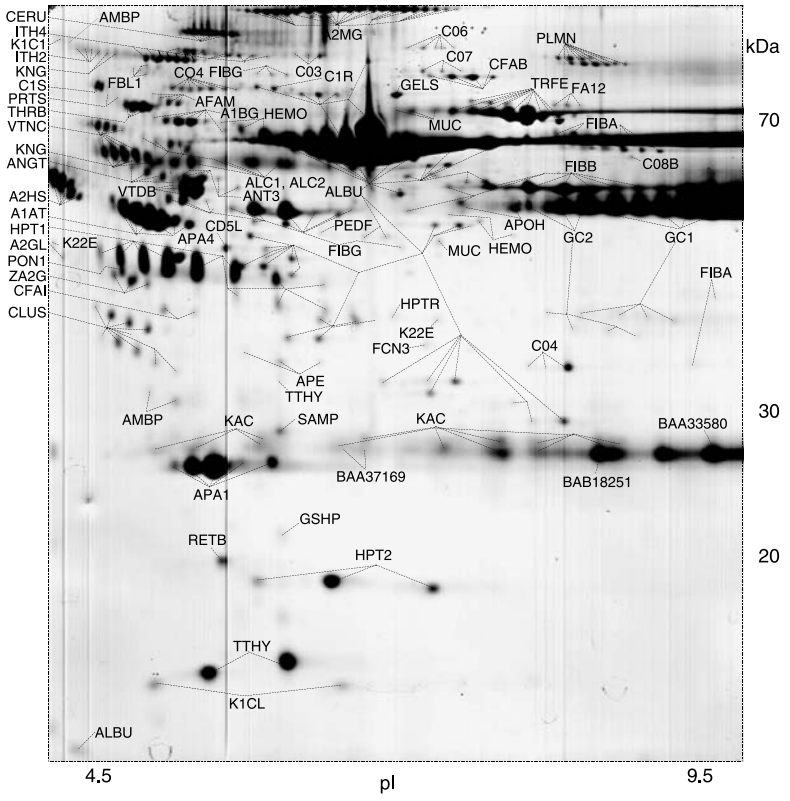

Fig. 2. Two-dimensional map of the human plasma proteins. The proteins were separated on a pH 3–10 non-linear IPG strip, followed by a 12% SDS-polyacrylamide gel, as stated under Materials and methods. The gel was stained with Coomassie blue. The spots were analyzed by MALDI-TOF-MS. The proteins identified are designated with their abbreviated names and are listed in Table 1 (not all proteins of Table 1 are shown)

albumin depletion and recovery of the other proteins. The two gel types delivered comparable results. The further work was continued with Mimetic blue as albumin depletion at pH 6.0 was slightly more efficient with this gel matrix (data not shown). The procedure was repeated three times to control the reproducibility and to establish standard operation procedures. A high reproducibility was achieved and the protein elution profile, the 1-D and the 2-D gel analyses were practically identical.

The columns were connected in series so that the flow-through fraction of the Mimetic blue step was directly applied onto the Protein G column. After washing with the application buffer, the two columns were separated and

Table 2. Albumin and IgG chains depletion. Protein recovery during albumin and immunoglobulin chains depletion by chromatography on Mimetic blue and Protein G, respectively. The chromatography was performed on a 2-ml Mimetic blue and a 1-ml Protein G column as described in Materials and methods

Purification step	Total protein (mg)	Recovery (%)
Starting material	40.0	100
Mimetic blue + Protein G flow-through	5.9	15
Mimetic elution	28.6	72
Protein G elution	4.8	12

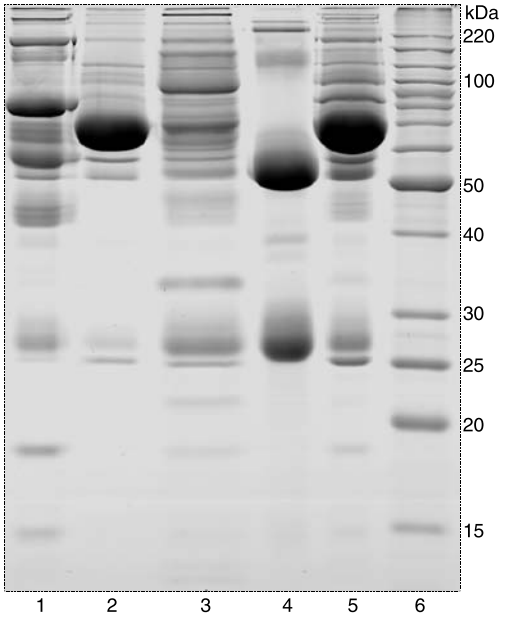

Fig. 3. SDS-PAGE analysis of the plasma proteins eluted from the Mimetic blue and protein G columns. 50 μg of protein was applied per lane and the gel (12.5%) was stained with Coomassie blue. Lane *1*, flow through from the Mimetic blue and Protein G columns (the material applied onto the ion exchange column). Lane *2*, eluate from the Mimetic blue column. Lane *3*, eluate from the Protein G with 1 M NaCl. Lane *4*, eluate from the Protein G with a buffer of pH 2.8. Lane *5*, unfractionated plasma (starting material). Lane *6*, size markers

eluted with a high salt and a low-pH buffer. Approximately 85% of the applied protein amount was retained on the two columns and 15% was recovered in the flow-through fraction (Table 2). Figure 3 shows an SDS-PAGE analysis of the eluate from the two columns (lane 1) and of the starting material (lane 5). A significant number of bands are seen in the eluate that were hidden under the signals of albumin (lane 2) and IgG chains (lane 4). The Protein G column was first eluted with a high-salt buffer which did not result in the elution of the heavy and light chains (lane 3). The chains were eluted with a low-pH buffer (lane 4).

In Fig. 4, the 2-D gel analysis of the untreated plasma and of the flow-through fractions from the Mimetic blue and the Protein G columns are shown. The Mimetic blue step resulted in an almost complete removal of albumin. After this step, albumin represented about 1% of total proteins (Fig. 4B). Albumin could be completely depleted

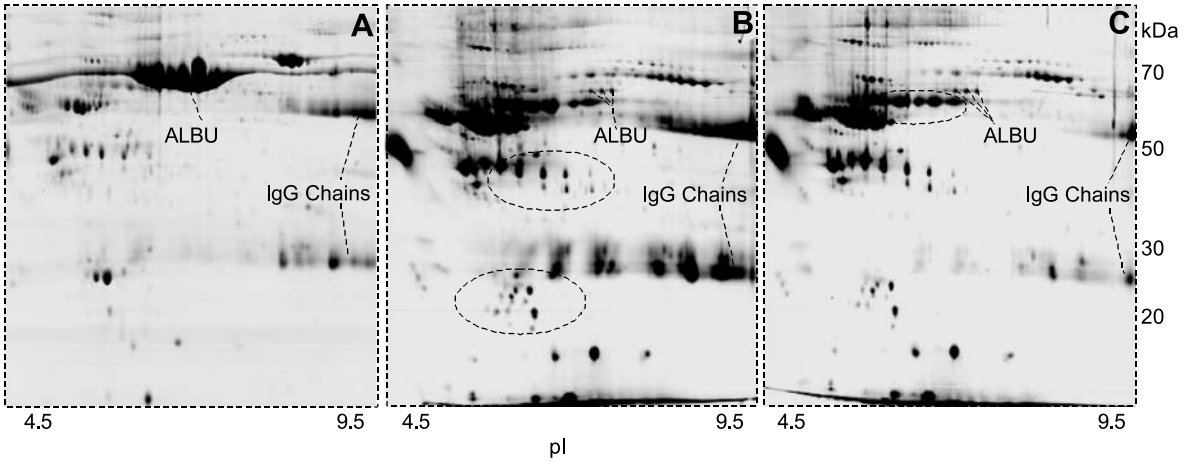

Fig. 4. Two-dimensional electrophoresis analysis of untreated plasma (**A**), after albumin-depletion (**B**) and after albumin- and IgG chain-depletion (**C**). The proteins were separated on pH 3–10 non-linear IPG strips, followed by 12% SDS-polyacrylamide gels, which were stained with Coomassie blue. B, C, the elliptic regions show increased number of signals after the removal of the abundant proteins

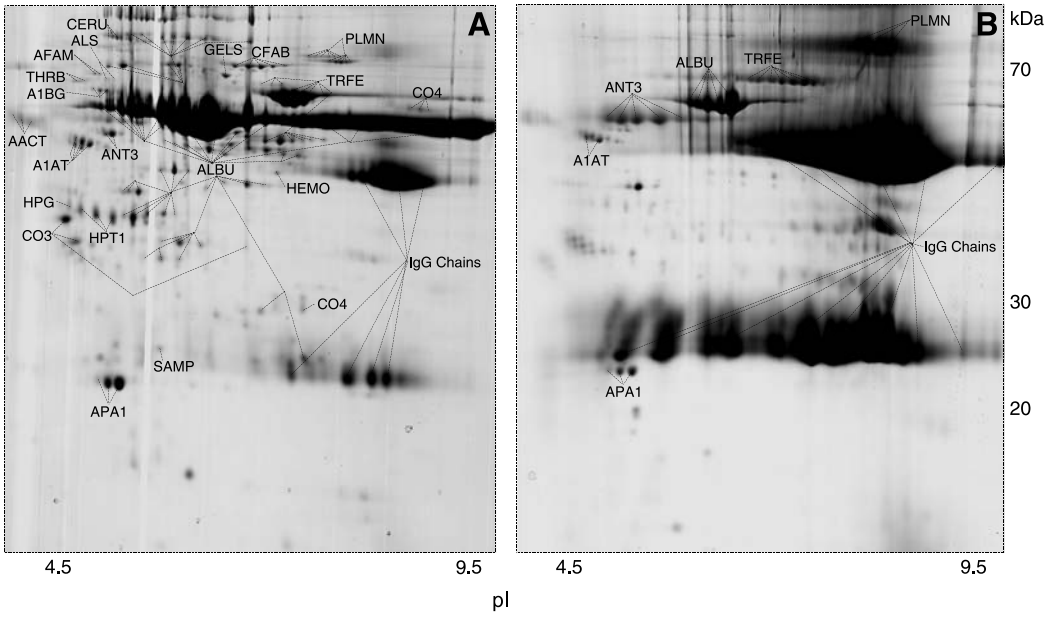

Fig. 5. Two-dimensional electrophoresis analysis of proteins eluted from the matrix blue column (**A**) and the Protein G column (**B**). Plasma was applied onto the Mimetic blue and Protein G columns and the proteins bound were eluted as described in Materials and methods. The eluate was analyzed as described in the legend of Figure 4. The proteins are designated with their abbreviated names and are listed in Table 1

when 10 mg of plasma were applied per ml of gel, instead of 20 mg, but this had as a consequence that also other proteins were retained by the matrix gel (data not shown). After albumin depletion, a more efficient detection of weak spots became possible (Fig. 4B, the circled areas). Passage through the Protein G column resulted in an efficient removal of approx. 95% of the immunoglobulins and an additional enrichment of weak spots (Fig. 4C). Protein G retained many other proteins as well which were eluted with high salt (Fig. 3, lane 3).

We further analyzed the eluates from the two columns by 2-D gels to identify which proteins were retained in addition to albumin and immunoglobulins. In the eluate of the Mimetic blue, mainly spots representing full-length albumin and protein forms of larger and smaller molecular mass were detected (Fig. 5A). The larger albumin forms are most likely artifacts of the 2-D gels and the shorter forms are fragments, which show a reproducible distribution profile. The predominant spot represents intact albumin. Strong spots representing other proteins are also visible, like serotrasferrin (TRFE), IgG chains, plasminogen (PLMN), complement proteins (CFAB, CO3, CO4), alpha-1 antitrypsin (A1AT), antithrombin (ANT3), haptoglobin (HPT1), apolipoprotein A-1 (APA1) and others (Fig. 5A). In the eluate from the Protein G column, mainly heavy and light IgG chains were detected but also spots representing albumin not bound to Mimetic blue (ALBU), serotrasferrin (TRFE), plasminogen (PLMN), alpha-1 antitrypsin (A1AT), antithrombin (ANT3), apolipoprotein A-1 (APA1) and others (Fig. 5B). Most of the other plasma proteins were recovered in the flow through fraction.

Discussion

Our goal was the establishment of standard operation procedures for the removal of the most abundant plasma proteins, albumin and immunoglobulin heavy and light chains, which together amount to about 80% of total plasma proteins. Plasma contains a large number of proteins with existing and potential therapeutic value. Application of proteomics methods may lead to the development of multi-protein, disease-specific biomarkers to improve the reliability and specificity of diagnostics (Lathrop et al., 2003). The major drawback of a plasma proteomic analysis is that the gene product copy number shows a wide distribution of about 12 orders of magnitude (Lescuyer et al., 2004), which makes the detection of the low-abundance proteins an extremely difficult undertaking.

The developed purification scheme involves chromatography on Mimetic blue and Protein G columns. The two columns mainly bound albumin and IgG chains, respectively; however, the binding was not specific. Both columns bound several other high-abundance plasma proteins as well and probably also low-abundance proteins which were not visible in the 2-D gels because of the small amount. The unspecific binding is an unavoidable compromise in the process of depletion of the high-abundance proteins to increase the probability of detection of the low-abundance counterparts. The other alternative is the direct analysis of the untreated plasma. Such an analysis can result in the detection of differences between the about 80 abundant proteins in control and diseased samples but will not result in the detection of differences in the levels of the low-abundance proteins.

The depletion of albumin and immunoglobulins resulted in the detection of additional weak spots or hidden spots which were co-migrating with those of abundant proteins but not many new spots could be detected, in particular enrichment of the low-molecular-mass proteins was inefficient. This could be due to the fact that the low-abundance proteins are present in very small amounts so that their detection is not possible in the 2-D gels stained with Coomassie blue and to the technical limitation of inefficient detection of small proteins in 2-D gels. This indicates that additional enrichment steps are necessary for their enrichment and visualization.

Instead of matrix blue columns, albumin can be retained by immobilized specific antibodies. Comparison between affinity chromatography on monoclonal antibody and matrix blue showed that the latter step is more efficient as it has a high capacity (approximately 20 mg/ml of gel) whereas about 3 mg of antibody are required for the binding of 1 mg of albumin. Antibodies are more difficult to produce and the binding specificity is only slightly higher than that of the matrix blue gels. Thus, in addition to the albumin forms, several other proteins were detected in the eluate from antibody columns, i.e. serotransferrin, alpha-1 anti-trypsin, vitamin D-binding protein, apolipoprotein A-I, haptoglobin and others (data not shown). Use of multi-component immunoaffinity chromatography has resulted in the detection of many lower abundance proteins (Pieper et al., 2003b); however, this approach may not be practical for the analysis of a large number of samples.

Govorukhina et al. (2004) have compared several matrix blue gels for their efficiency in albumin and immunoglobulin depletion in serum. They found that with the Poros anti-albumin/Protein G and the combination of

HiTrap Blue/HiTrap Protein G columns the protein depletion was the highest. The depletion of those proteins from serum or plasma has facilitated the detection of disease markers. Removal of the two most abundant proteins is indeed essential for the detection of lower abundant proteins; however, the complexity is much larger on account of the dynamic range of the protein distribution. For example, albumin is present in plasma at about 50 mg/ml, the tumor marker prostate-specific antigen at 1 ng/ml and troponin at 1 pm/ml. If albumin could be removed to 99.9%, the remaining protein would be present at 50 μg/ml, i.e. at a 50000-fold higher concentration in comparison with the prostate-specific antigen (Zolg and Langen, 2004).

Therefore, for the detection of the low-copy-number gene products, in addition to the depletion of the abundant proteins, technologies for the efficient separation of the low- from the high-abundant and for the efficient enrichment of the low-abundant proteins to detectable levels have to be developed. The latter has to include chromatography and electrophoresis steps. We have applied preparative electrophoresis for the fractionation of plasma proteins on cylindrical LDS-gels. The method did not result in a significant enrichment because albumin could not be efficiently separated from the other proteins (unpublished results). The same method was very successful in the fractionation of brain and liver proteins (Fountoulakis and Juranville, 2003; Fountoulakis et al., 2004). A simple method for albumin depletion comprises ammonium sulfate precipitation steps. In the pellet of the 50% ammonium sulfate precipitation, the majority of the plasma proteins and only a minor amount of albumin were found and in the pellet of the 70% precipitation mainly albumin, serotransferrin, anti-trypsin and haptoglobin-1 were detected. This fraction did not include antibody chains (Jiang et al., 2004). Following albumin and immunoglobulin depletion with the approach described in this study, the flow-through fraction was subjected to chromatography over an ion exchanger. The step resulted in the identification of more than 380 proteins in comparison with the 83 identified in the 2-D gels of the untreated plasma (manuscript in preparation).

Although approx. 400 different gene products have been identified in plasma until now using 2-D gels/MALDI-TOF-MS and a few hundred more may be detected if additional protein fractionation and enrichment steps will be employed, still this classical proteomic approach may not be the right tool in analyzing body fluids which include dominating proteins like albumin. Complete removal of such proteins, even if it can be achieved, may not be desirable because such proteins will trap many of the low-abundance proteins, their fragments and peptides, which will thus be lost and not detected. In parallel to the described method, alternative proteomic approaches may be useful, such as protein digestion, multidimensional chromatography for peptide separation and enrichment and identification by tandem MS or MALDI-TOF-MS. The presence of the high-abundance proteins will still be an issue for these methods whereas their removal will be associated with the drawbacks discussed above. Moreover, the multidimensional-tandem MS method is still under development, whereas the 2-D gel/MALDI-TOF-MS is a well documented, robust method, efficiently applied in most proteomics studies. The enormous potential of detection of novel drug targets and diagnostic markers in plasma and other body fluids will further drive the development of the analytical techniques.

In conclusion, we constructed a 2-D map for plasma proteins, including 83 gene products, and developed a simple and reproducible method for the depletion of the abundant proteins albumin and immunoglobulin heavy and light chains. The scheme comprises chromatography on Mimetic blue and Protein G columns at pH 6.0. The method is convenient to perform, results in the removal of about 99% of albumin and 95% of antibody chains and is useful in the search of molecules of medical interest in body fluids.

References

Anderson NL, Anderson NG (2002) The human plasma proteome: History, character and diagnostic prospects. Mol Cell Proteomics 1: 845–876

Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, Veenstra TD, Adkins JN, Pounds JG, Fagan R, Lobley A (2004) The human plasma proteome. Mol Cell Proteomics 3: 311–326

Barroso B, Lubda D, Bischoff R (2003) Applications of monolithic silica capillary columns in proteomics. J Proteome Res 2: 633–642

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

Bradford M (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

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

Fountoulakis M (2004) Application of proteomics technologies in the investigation of the brain. Mass Spectrom Rev 23: 231–258

Fountoulakis M, Gasser R (2003) Proteomic analysis of the cell envelope fraction of *Escherichia coli*. Amino Acids 24: 19–41

Fountoulakis M, Juranville J-F (2003) Enrichment of low-abundance brain proteins by preparative electrophoresis. Anal Biochem 313: 267–282

Fountoulakis M, Takács B (2002) Enrichment and proteomic analysis of low-abundance bacterial proteins. Methods Enzymol 358: 288–306

Fountoulakis M, Juranville J-F, Tsangaris G, Suter L (2004) Fractionation of liver proteins by preparative electrophoresis. Amino Acids 26: 27–36

Govorukhina NI, Keizer-Gunnink A, van der Zee AG, de Jong S, de Bruijin HW, Bischoff R (2003) Sample preparation of human serum for the analysis of tumor markers. Comparison of different approaches for albumin and gamma-globulin depletion. J Chromatogr A 1009: 171–178

Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci USA 90: 5011–5015

Jiang L, He L, Fountoulakis M (2004) Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. J Chromatogr A 1023: 317–320

Jiang L, Lindpaintner K, Li H-F, Gu N-F, Langen H, He L, Fountoulakis M (2003) Proteomic analysis of the cerebrospinal fluid of patients with schizophrenia. Amino Acids 25: 49–57

Langen H, Roeder D, Juranville J-F, Fountoulakis M (1997) Effect of the protein application mode and the acrylamide concentration on the resolution of protein spots separated by two-dimensional gel electrophoresis. Electrophoresis 18: 2085–2090

Lathrop JT, Anderson NL, Anderson NG, Hammond DJ (2003) Therapeutic potential of the plasma proteome. Curr Opin Mol Ther 5: 250–257

Lescuyer P, Hochstrasser DF, Sanchez JC (2004) Comprehensive proteome analysis by chromatographic protein prefractionation. Electrophoresis 25: 1125–1135

Pieper R, Gatlin CL, Makusky AJ, Russo PS, Schatz CR, Miller SS, Su Q, McGrath AM, Estock MA, Parmar PP, Zhao M, Huang ST, Zhou J, Wang F, Esquer-Blasco R, Anderson NL, Taylor J, Steiner S (2003a) The human serum proteome: Display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. Proteomics 3: 1345–1364

Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S (2003b) Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome. Proteomics 3: 422–432

Veenstra TD, Conrads TP (2003) Serum protein fingerprinting. Curr Opin Mol Ther 5: 584–593

Vlahou A, Schorge JO, Gregory BW, Coleman RL (2003) Diagnosis of ovarian cancer using decision tree classification of mass spectral data. J Biomed Biotechnol 2003: 308–314

Zhang R, Barker L, Pinchev D, Marshall J, Rasamoelisolo M, Smith C, Kupchak P, Kireeva I, Ingratta L, Jackowski G (2004) Mining biomarkers in human sera using proteomic tools. Proteomics 4: 244–256

Zolg JW, Langen H (2004) How industry is approaching the search for new diagnostic markers and biomarkers. Mol Cell Proteomics 3: 345–354

Authors' address: Michael Fountoulakis, Foundation for Biomedical Research of the Academy of Athens, Soranou Ephessius 4, 11527 Athens, Greece, E-mail: mfountoulakis@bioacademy.gr; Hanno Langen, F. Hoffmann-La Roche Ltd., Center of Medical Genomics, Building 93-832, 4070 Basel, Switzerland, E-mail: hanno.langen@roche.com