Short thesis for the degree of Doctor of Philosophy (PhD)

GENERATION AND CHARACTERIZATION OF AN ANALYTE LIBRARY REPRESENTING THE HUMAN PLASMA PROTEOME AND ITS APPLICATION FOR THE IDENTIFICATION OF ANTIGENS CORRESPONDING TO MONOCLONAL ANTIBODIES

by András Kovács

Supervisor:
András Guttman MHAS

UNIVERSITY OF DEBRECEN
Doctoral School of Molecular Medicine
Debrecen, 2014

Generation and characterization of an Analyte Library representing the human plasma proteome and its application for the identification of antigens corresponding to monoclonal antibodies

By András Kovács MSc

Supervisor: András Guttman PhD, DSc, MHAS

Doctoral School of Molecular Medicine, University of Debrecen

Head of the Examination Committee:	József Tőzsér PhD, DSc
Members of the Examination Committee:	Miklós Idei PhD, DSc
	Attila Bácsi PhD

The Examination takes place at Office No. 3.106, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen
11 AM, 15 May, 2014.

Head of the Defense Committee:	József Tőzsér PhD, DSc
Reviewers:	Huba Kalász PhD, DSc
	Péter Antal-Szalmás MD, PhD
Members of the Defense Committee:	Miklós Idei PhD, DSc
	Attila Bácsi PhD

The PhD Defense takes place at the Lecture Hall of Bldg. A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen
1 PM, 15 May, 2014.

1. INTRODUCTION

In biomarker research, the study of the human plasma proteome is of special significance. Plasma accounts for approximately 55% of the total blood volume, and contains a wide concentration range of proteins with diverse biological functions. Due to its high complexity, the characterization and identification of medium or low abundant proteins often represents a difficult and time-consuming task. The lower the protein concentration in the plasma, the higher the analyte volumes it requires for the analysis. In addition, labour-intensive and time-consuming sample preparation methods are needed before individual proteins can be adequately purified by various bioanalytical techniques, which methods sometimes alter the native state of the proteins. In addition to the difficulties related to the extremely wide concentration range, plasma and biomarker research deal with the identification of protein isoforms and post-translational modifications, which may cause slight but considerable changes in the function of a protein.

The majority of the approaches addressing the study of the plasma proteome use mass spectrometry for the identification of proteins of biomarker potential. However, these technologies are not always suitable to cover the dynamic range necessary for global proteome analysis, due to sensitivity, reproducibility, and throughput limitations.

Monoclonal antibody proteomics uses specific antibodies, which recognize most proteins at low concentration. The numerous produced antibodies make the comprehensive screening of the plasma possible and may detect different protein variants (isoforms, post-translational modifications, etc.), followed by mass spectrometric protein identification.

My doctoral thesis reports on the development of a method that aims to maintain the native state of proteins during the sample preparation phase, and then uses the workflow of monoclonal antibody proteomics–mass spectrometry for protein identification.

2. OBJECTIVES

1. Generation of a comprehensive Analyte Library representing the human plasma proteome, using specific depletion, protein precipitation and chromatographic techniques, expecting hundreds of fractions, each containing of some 10-100 intact proteins, balancing the existing 11-12 order of magnitude differences, which make isolation and identification of proteins occurring at low concentrations extremely difficult. The goal of the work was to generate a human plasma proteome library serving as a depository of potential biomarkers, providing sufficient amount of analyte for immunoseparation-linked liquid chromatography-mass spectrometry-based identification of medium- and low-abundant plasma proteins and their variants.

2. Characterization of the Analyte Library and the determination of its applicability for monoclonal antibody proteomics-based antigen identification. Following the separation route of vitamin D-binding protein through the different levels of fractionation to study the possible – even discrete – accumulation of the protein in the less complex final fractions compared to the upstream (interim), more complex ones. Comparison of the distribution of seven plasma proteins with the expected distributions predicted based on physical, chemical and biochemical properties and data in the literature comprehensive screening of the Library with monoclonal antibodies specific for known antigens.

3. The application of the Library fractions for monoclonal antibody proteomics-based antigen identification with mAb-s of a complex antibody library developed for the global profiling of the human plasma proteome.

3. MATERIALS AND METHODS

3.1. Analyte Library generation

For the generation of the Analyte Library, 500 mL commercially available normal human plasma, collected from single healthy donors between the ages of 18 and 65, pooled by FDA regulations and tested for blood-borne viruses was used.

First, the most abundant plasma component, human serum albumin (HSA) was depleted from 500 mL normal pooled human plasma by *Blue Sepharose affinity chromatography*. For the removal of HSA, twenty runs (each with 25 mL plasma) were applied, and the flowthroughs were collected for downstream fractionation. The albumin (with the associated proteins) was eluted from the column by 1 M KCl and after each run, and the column was regenerated by 8 M guanidine hydrochloride. The fractions (flowthrough, elution, and regeneration) obtained at this step were analyzed by SDS-PAGE and kept at -70°C until further processing.

For the depletion of the immunoglobulins, *thiophilic interaction chromatography* was used. After each individual Blue Sepharose chromatography step, the albumin depleted flowthrough was mixed with Pierce Thiophilic Adsorbent and potassium sulphate (concentration: 0.5 M), and after an overnight stirring it was loaded into a preparative chromatography column and the flowthrough and the first two wash fractions were collected. The bound immunoglobulins were eluted by 50 mM sodium phosphate (pH 8.0) and the column was regenerated by 8 M guanidine hydrochloride. The fractions (flowthrough, elution, and regeneration) obtained at this step were analyzed by SDS-PAGE and kept at -70°C until the next separation step.

The albumin and immunoglobulin-depleted plasma solution was *concentrated* by centrifugal filter devices (cut-off: 10 kDa) and buffer exchange to PBS (pH: 7.4) was performed by a 24 h *dialysis* at 4°C. The volume of one dialysed sample was 96 mL (290-mL plasma solution was divided into three parts), while that of the external PBS solution was 2 L.

For the pre-fractionation of the concentrated and dialyzed plasma *ammonium sulphate precipitation* was applied at four consecutive saturation levels of 35, 45,

65, and 75%. One salting out step was performed for a minimum of 1 h at 4°C. The protein precipitates were collected by centrifugation and redissolved in PBS (pH 7.4). The fractions were analyzed by SDS-PAGE and kept at -70°C until further processing.

Each PBS-solubilised precipitate was further fractionated by *size exclusion chromatography (SEC, a.k.a. gel filtration)* using Sephacryl S-200 media. One mL fractions were collected and analyzed by SDS-PAGE, the fractions with similar electrophoretic profiles were pooled and kept at -70°C.

Prior to the subsequent step, the *cation exchange chromatography*, the PBS buffer of every pooled size exclusion chromatography fraction was changed to 50 mM phosphate buffer (pH 5.5) by Amicon Ultra centrifugal filter units. Each SEC fraction was separated using a strong cation exchange resin-filled column (HiTrap HP SP) and the flowthoughs were collected. The bound proteins were eluted by a linear (0-0.5 M) KCl gradient in 50 mM phosphate buffer (pH 5.5), 0.5-1 mL fractions were collected depending on the protein amount in the loaded samples. The fractions were analyzed by SDS-PAGE and kept at -70°C.

Prior to *anion exchange chromatography*, the buffer of each cation exchange chromatography flowthrough was changed to 20 mM TRIS-HCl (pH 8.5) using Amicon Ultra centrifugal filter devices and were fractionated applying a strong anion exchange column (HiTrap HP Q). The bound proteins were eluted by a linear (0-0.5 M) KCl gradient in 20 mM Tris-HCl (pH 8.5) and 0.5-1 mL fractions were collected depending on the protein amount in the loaded samples. The collected fractions were analyzed by SDS-PAGE and kept at -70°C.

The 20 most concentrated ion exchange chromatography fractions (>8.0 mg/mL) were further separated by *hydrophobic interaction chromatography* using a phenyl sepharose resin-filled column (HiTrap Phenyl Sepharose HP). For precipitates with 45% ammonium sulphate saturation level, 1 M ammonium sulphate containing 20 mM Tris-HCl buffer (pH 7.1) was applied, while for precipitates with 65 and 75% ammonium sulphate saturation level, 2 M ammonium sulphate containing 20 mM Tris-HCl buffer (pH 7.1) sample/starting buffer, respectively. The elution was performed by a descending ammonium sulphate gradient (100-0%) in 20 mM

Tris-HCl buffer (pH 7.1), and 1 mL fractions were collected. The fractions were analyzed by SDS-PAGE, and kept at -70°C.

3.2. Fraction coding

The reference codes for the chromatography fractions comprise the mark for the % ammonium sulphate saturation level (35, 45, 65, or 75), the number of the pooled gel filtration fraction (G1-G4), specify the cation (C) or anion exchange chromatography fractions (A) and, designate the hydrophobic interaction chromatography fractions (H). CF stands for cation-exchange chromatography flowthrough. Numbers after C, CF, A, and H represent the position of fractions in the collection tray. For example, 65%-G2-A19-H9 codes a sample obtained by 65% ammonium sulphate precipitation followed by gel filtration (fraction #2), anion exchange chromatography (fraction #19), and hydrophobic interaction chromatography (fraction #9).

3.3. Characterization of the Analyte Library: dot blot and Western blot

Depending on the concentration of the actual fractions of interest, a total amount of 3 or 1 µg protein per fraction was spotted onto Protran BA 85 nitrocellulose membrane placed in a *dot blot* vacuum manifold designed for 96 samples and 10 kPa vacuum was applied. The dried membrane was treated for 30 min with the blocking buffer of 0.5% polyvinylpyrrolidone in 0.05% Tween 20 containing PBS, following 1 h incubation with the relevant monoclonal antibody diluted to 1 µg/mL in the blocking buffer. The membrane was washed for 5*1 min in water, 5*1 min in the blocking buffer, and 5*1 min in PBS-T (0.05% Tween 20 containing PBS). Goat anti-mouse IgG-HRP, diluted 8000-fold in the blocking buffer, was used as secondary antibody. The membrane was incubated with the secondary antibody for 1 h and washed for 5*1 min with PBS-T. Chemiluminescent signals were developed by the application of Pierce ECL Western Blotting Substrate and detected with Gel Logic 2200 PRO gel imaging system. An aliquot of 0.05 µg of the primary antibody was spotted on the membrane as a positive control for the reaction and the blocking buffer was used as a negative control. The increasing signal intensities were

evaluated by scores from 0.5 to 5.0; 0.5 was given to fractions with signal intensities barely distinguishable from the background, while score 5 meant a signal intensity that was comparable with that of the positive control (identical or higher). Fractions with scores above 3 were considered as useful for further examinations. Fractions with scores 1 or 2 were considered as unsuitable, although their signal intensities were distinguishable from the background, indicating antigen-antibody reaction, but it was presumed that these fractions may not contain enough proteins for the further analysis.

In the case of the *Western blot* analyses, 2 µg protein per fraction was separated by SDS-PAGE and transferred onto nitrocellulose membrane and blotting was carried out for 1 hour by Mini Trans-Blot Cell blotting equipment. The same buffer was applied for the blocking (1 h), and the membrane was incubated at 4 ºC with 5000-fold diluted primary monoclonal antibody containing blocking buffer. The membrane was washed as described earlier and incubated with 5000-fold diluted HRP conjugated goat anti-mouse IgG containing blocking buffer. The membrane was washed in PBS-T and the chemiluminescent signals were developed by the application of Pierce ECL Western Blotting Substrate and detected with Gel Logic 2200 PRO gel imaging system. The membrane was stained with 0.05% Amido black (in 10% acetic acid) for 1 h at room temperature, washed 2*10 min with 5% acetic acid, and visualized with UVIPro Gold Gel Documentation System.

3.4. Immunoprecipitation, SDS-PAGE and Western blot analysis of the eluates

During the Analyte Library project, I was in cooperation with BioSystems International Kft. a biotechnology company located at the campus of the University of Debrecen. The company provided the monoclonal antibodies for the investigations and colleagues contributed to the preparation of Analyte Library fractions for the mass spectrometric analyses by immunopecipitation experiments (Invitrogen Dynabeads Protein G kit), SDS-PAGE and Western blot studies. The brief outline of the method is as follows:

Immunoaffinity magnetic beads were prepared for each antibody selected during the dot blot screening, by mixing the antibody with protein G-coupled Dynabeads

followed by covalent cross-linking. The antigen was immunoprecipitated by mixing the beads with the fractions. The beads with bound antigens were washed and the antigen was eluted. The proteins in the eluates were separated by SDS-PAGE and the gels were subjected to Coomassie Blue staining.

3.5. Antigen identification by mass spectrometry

The in-gel trypsin digestion and follow-up mass spectrometric analysis were performed at the Laboratory of Proteomics Research at the Biological Research Centre of the Hungarian Academy of Sciences (Szeged, Hungary), using a Thermo LCQ Fleet three-dimensional ion trap coupled with Eldex nanoHPLC system. The "in-solution" mass spectrometric analyses of the specifically immunoprecipitated proteins were performed at the Institute of Structural Chemistry at the Chemical Research Center of the Hungarian Academy of Sciences (Budapest, Hungary). The samples were processed for mass spectrometric analysis using a Q-TOF Premier mass spectrometer coupled with a nanoflow Ultra HPLC system.

3.6. Composition of the applied buffers and solutions

SDS-PAGE running buffer: 25 mM TRIS, 192 mM glycine, 0.1% SDS, pH 8.6

SDS-PAGE staining solution: 0.2% Brilliant Blue R, 7.5% acetic acid, 50% ethanol

SDS-PAGE destain 1 solution: 10% acetic acid, 20% ethanol

SDS-PAGE destain 2 solution: 10% acetic acid, 5% ethanol

PBS: 3.8 mM sodium dihydrogen phosphate, 16.2 mM disodium hydrogen phosphate, 150 mM sodium chloride, pH 7.4

Albumin depletion, elution buffer: 1 M KCl in PBS

Albumin and immunoglobulin depletion, regeneration buffer: 8 M guanidine hydrochloride

Immunoglobulin depletion, binding buffer: 0.5 M potassium sulphate

Immunoglobulin depletion, elution buffer: 50 mM sodium phosphate, pH 8.0

Ammonium sulphate: 35%, 45%, 65%, and 75% saturation

Cation exchange chromatography, sample buffer: 50 mM phosphate buffer, pH 5.5

Cation exchange chromatography, elution buffer: 0-0.5 M KCl in 50 mM phosphate buffer, pH 5.5

Anion exchange chromatography, sample buffer: 20 mM TRIS-HCl, pH 8.5

Anion exchange chromatography, elution buffer: 0-0.5 M KCl in 20 mM TRIS-HCl, pH 8.5

Hydrophobic interaction chromatography, sample and elution buffer: 1 and 2 M ammonium sulphate in 20 mM TRIS-HCl, pH 7.1

Dot blot, Western blot, blocking buffer: 0.5% polyvinyl pirrolide and 0.05% Tween 20 in PBS, pH 7.4

Dot blot, Western blot, washing buffer: 0.05% Tween 20 in PBS, pH 7.4

Western blot, blotting buffer: 25 mM TRIS, 192 mM glycine, 20% methanol

Amido Black staining solution: 0.05% Amido Black, 10% acetic acid

4. RESULTS

4.1. Analyte Library generation

The Analyte Library generation was a multi-step plasma fractionation process applying specific depletion, protein precipitation and chromatographic methods.

As the first step toward the generation of the comprehensive Analyte Library, the *large human serum albumin content* of 500 mL human plasma was *depleted* by Blue Sepharose chromatography. This affinity chromatography step resulted in (1) the albumin-depleted flowhtrough along with the (2) eluted albumin and (3) regeneration fractions. According to the SDS-PAGE analysis of the obtained fractions, the HSA content in the flowthrough significantly decreased. The elution step resulted in a fraction containing the majority of the depleted albumin and proteins showing affinity to Blue Sepharose. The regeneration eluate still contained some remaining HSA and other proteins not eluted by the 1 M KCl.

After the albumin removal, *immunoglobulin depletion* was conducted by thiophilic interaction affinity chromatography. After this step, a large volume (~3.6 L) of albumin- and immunoglobulin depleted flowthrough was obtained for further separation. The bound immunoglobulins were eluted from the adsorbent beads and were collected for storage. The amount of immunoglobulins significantly decreased in the flowthrough and the proteins were detected in the eluate.

Centrifugal concentration and dialysis of the albumin- and immunoglobulin depleted plasma solution were performed to remove K_2SO_4 and as a result of these operations, its total volume decreased to 335 mL.

The albumin- and immunoglobulin depleted flowthrough was *pre-fractionated* by *salting out with ammonium sulphate* in four consecutive concentrations of 35, 45, 65, and 75%. At lower ammonium sulphate concentrations (35, 45%), mostly the hydrophobic/less water-soluble proteins were salted out, while at higher ammonium sulphate concentrations (65, 75%), the less hydrophobic/more water-soluble proteins were precipitated. The supernatant of the 75% precipitation step did not contain detectable amount of proteins, thus this fraction was not used for further processing.

The next step of the fractionation process was the *size exclusion chromatography (gel filtration)* of the four ammonium sulphate precipitates. The fractions were analyzed by SDS-PAGE and the fractions with similar profile were pooled, i.e. similar number of bands of the same molecular weight range and their approximate amounts. This step generated 14 interim Analyte Library fractions: 3–3 from the 35 and 45% ammonium sulphate precipitation steps, and 4–4 from the 65 and 75% ammonium sulphate precipitation steps, respectively.

Every (14) size exclusion chromatography fraction was further separated by *strong cation exchange chromatography (SCX)*. The flowthroughs containing the unbound proteins were collected and the bound ones were eluted by an ascending linear KCl gradient and 322 final Analyte Library fractions were obtained.

Each (14) individual cation exchange chromatography flowthrough was fractionated by *strong anion exchange chromatography (SAX)*. The bound proteins were eluted by an ascending linear KCl gradient and 316 final Analyte Library fractions were collected, whose SDS-PAGE analyses revealed higher complexity and less overlaps between neighbouring fractions than observed in the case of the SCX fractions. The flowthroughs did not contain any detectable amounts of proteins.

The 20 most concentrated ion exchange chromatography fractions (>8.0 mg/mL) – all anion exchange ones – were further separated by *hydrophobic interaction chromatography (HIC)* with a phenyl type stationary phase. The bound proteins were eluted by a descending ammonium sulphate gradient and 145 fractions were collected to have altogether 783 final Analyte Library fractions. The SDS-PAGE analysis of the fractions suggested that the separation of the more hydrophobic fractions (45% ammonium sulphate saturation level) resulted in better resolution than that of the rather hydrophilic ones (65 and 75% ammonium sulphate saturation). The flowthroughs did not contain any detectable amounts of proteins.

4.2. Characterization of the Analyte Library I: investigation of the separation route of vitamin D-binding protein

To assess the applicability of the Analyte Library, as a representative example, the separation route of vitamin D-binding protein was followed through the different levels of fractionation by dot blot and Western blot assays.

Altogether 46 fractions were selected from all levels of the separation process on the basis of their complexity (defined by SDS-PAGE) and protein concentrations. The selection included eight interim fractions from three early levels of the Library generation: the albumin- and immunoglobulin-depleted plasma, the four ammonium sulphate precipitates, three size exclusion chromatography fractions; and 38 from the higher fractionation levels: 14 cation exchange, 16 anion exchange and 8 hydrophobic interaction chromatography fractions.

A *dot blot assay* was conducted spotting the 46 Analyte Library fractions on the membrane (antigens), using vitamin D-binding protein (VDBP) specific monoclonal antibody for the reaction. The increasing chemiluminescent signal intensities were scored from 0.5 to 5.0.

Out of eight interim fractions seven evolved signals with intensities between 1 and 2, while out of the 38 final fractions only 19 contained the antigen – signal intensities varied between 0.5 and 1 –, mainly the ones derived from the 45 and 65% ammonium sulphate precipitates, suggesting that during the Library generation process the complexity of the fractions gradually decreased. The highest enrichment of VDBP – signal intensity reached 5.0 – was observed in the hydrophobic interaction chromatography fraction 65%-G3-A17-H10.

The identity of VDBP in the fraction 65%-G3-A17-H10 was verified by *Western blot*. In addition, its neighbouring fractions and also the ones on its separation lineage were studied. The analysis revealed the enrichment of the antigen from the most complex fraction (plasma) to the less complex (hydrophobic interaction chromatography) final Analyte Library fractions.

4.3. Characterization of the Analyte Library II: high-throughput screening of the Library

The Analyte Library was further characterized with mAb-s specific for known antigens by means of a *dot blot array system* to study and compare the separation routes of plasma proteins by screening with their corresponding monoclonal antibodies. The selected plasma proteins – fibrinogen, vitamin D-binding protein, haptoglobin, complement component C9, alpha-2-macroglobulin, hemopexin, human serum albumin – were abundant ones to obtain intensive signals in favour of following their accumulation. Their known physical, chemical and biochemical traits were also of great importance in terms of comparing the observed distribution patterns to the expected ones.

Out of the total 783 final Analyte Library fractions, 638 were used selected by concentration and complexity: 244 cation exchange, 273 anion exchange, and 121 hydrophobic interaction chromatography fractions. The signal intensities were evaluated and fractions above 3 were considered as positives.

In the case of *fibrinogen*, almost discrete accumulation of the protein was observed in cation exchange chromatography fractions of the 35% ammonium sulphate precipitation (35%-G1, the highest in molecular mass).

In the case of *vitamin D-binding protein*, in addition to the HIC fractions described earlier, some positive signals were observed in anion exchange fractions of the 45, 65 and 75% ammonium sulphate precipitates in every molecular weight range.

In the case of *haptoglobin*, *hemopexin* and *human serum albumin*, 'smear'-like distribution was experienced, the antigens performed positive signals in factions of different characteristics in terms of hydrophobicity/hydrophilicity, size and shape.

Strong positive reactions with the *alpha-2-macroglobulin* specific mAb were observed mainly in fractions of the 35 and 45% ammonium sulphate precipitates in the higher molecular weight range and less strong in 45%-G2, 45%-G3 SAX and 45%-G2 HIC ones.

No positive final fractions were detected after screening the Analyte Library for the presence of *complement component C9*. As a result of the dot blot study on the interim fractions – HSA depleted plasma, KCl eluted HSA fraction, regeneration

fraction of the Blue Sepharose column, albumin and immunoglobulin depleted plasma and the first wash fraction of immunoglobulin depletion –, strong signals were experienced in the KCl eluate (albumin and other proteins, score 5) of the albumin depletion and less positive in the regeneration fraction (score 3).

4.4. Monoclonal antibody proteomics-based antigen identification

The *monoclonal antibody proteomics-based identification* was conducted as a multi-stage process of (1) high-throughput dot blot array screening of the above mentioned 638 fractions with the mix of selected monoclonal antibodies raised against protein mixtures in which the exact antigen was not known and (2) individual antibody screening of the most pronounced positive fractions of the mass screening stage, also with dot blot. Altogether 21 mAb-s were examined divided into mixes #1, #2 and #3 of 6, 10 and 5 antibodies, respectively. In the case of mix #2, an intermediate screening step of three sub-mixes (3, 3 and 4 mAb-s) was performed prior to the individual screening step.

Positive fractions were selected after reaction with individual antibodies on the basis of their complexity and concentration and they were then prepared for antigen identification in cooperation with BioSystems International Kft. The selected fractions were analyzed with Western blot to study the molecular weight of the protein reacted with the antibody. The fractions were subjected to immunoprecipitation and the eluate was examined by SDS-PAGE. After the evaluation of the results, the immunoprecipitated proteins were subjected to either in-gel or “in-solution” trypsin digestion and mass spectrometric analysis. The studies resulted in the cognate protein identification of eight monoclonal antibodies, which represents approximately 38% of the examined mAb-s:

Bsi0051 – vitamin D-binding protein, Bsi0080 – alpha-1-antitrypsin, Bsi0099 – ceruloplasmin, Bsi0115 – alpha-2-antiplasmin, Bsi0142 – alpha-1-antichymotrypsin, Bsi0144 – alpha-1-antichymotrypsin, Bsi0177 – ceruloplasmin, Bsi1521 – transthyretin.

5. DISCUSSION

5.1. Analyte Library generation

A 783-fraction containing Analyte Library was generated a by comprehensive fractionation of the human plasma proteome. Multi-step fractionation was applied, which technique is usually used to enrich proteins of interest for downstream characterization. Different depletion – Blue Sepharose affinity chromatography and thiophilic interaction chromatography for the depletion of albumin and the immunoglobulins, respectively –, protein precipitation (salting out with ammonium sulphate) and chromatographic – size exclusion, cation exchange, anion exchange, hydrophilic interaction – techniques were used. The different separation steps were based on several physical, chemical and biochemical features (solubility, size, charge) of proteins in favour of obtaining fractions of various characteristics. All chromatography and precipitation steps were carefully designed aiming to maintain the native forms of the intact proteins throughout the fractionation process. In some instances, the results of the SDS-PAGE analyses of the fractions suggested overlaps between certain neighbouring fractions. However, the protein bands detected at apparently identical molecular masses could correspond to different proteins. Detection methods, which are more sensitive than Coomassie staining, could reveal minor proteins in even one or only a few fractions. As expected, gradual decrease in the complexities of the fractions was observed with the progression of the fractionation. Based on these results, the Analyte Library was considered as applicable for further studies.

5.2. Characterization of the Analyte Library

The characterization and the evaluation of the feasibility of the Analyte Library for mAb proteomics was carried out by two different methods. The separation route of vitamin D-binding protein (VDBP) was analyzed in 46 fractions derived from the different levels of the fractionation process by dot and Western blots using VDBP specific mAb-s. As a result of the dot blot studies, one fraction (65-G3-A17-H10) was found in which the antigen accumulated almost discretely. The Western blot

analysis of neighbouring and downstream interim fractions showed the gradual accumulation of the protein in the fractions of decreasing complexity. This confirmed the initial theory of the existence of final fractions with an almost discrete accumulation of a protein providing higher specificity for the protein identification process than that of more complex samples, such as whole plasma.

The second step of the characterization was a high-throughput dot blot screening of the Library with mAb-s of known plasma components. The plasma proteins were abundant ones with known and specific physical, chemical and biochemical traits (hydrophobicity/hydrophilicity, size, shape, associative character) and were also potential biomarkers.

The distribution of *fibrinogen* matched the expectations. Fibrinogen can be salted out with ammonium sulphate of lower concentration (22.5%), its molecular weight is 340 kDa thus, it accumulated in fractions deriving from the 35% ammonium sulphate precipitate and of the highest molecular weight (G1). It was detected only in the 35%-G1 cation exchange fractions, therefore its accumulation was considered almost unique.

Vitamin D-binding protein (VDBP) performed almost distinct accumulation in fractions matching its molecular weight of 52-59 kDa (G3, G4 fractions). According to its known acting binding character, it could be observed in the medium (G2) size range, while due to its binding with very-low density lipoproteins (VLDL), it could be found in the highest molecular weight range (G1). VDBP has three polymorphic variants in humans. Differences in the structures of isoforms involved biochemical consequences and changes in the hydrophobic-hydrophilic character, solubility, thus they could accumulate in fractions of different physical, chemical and biochemical characteristics (fractions deriving from the 45, 65 and 75% ammonium sulphate precipitate).

The ‘smear’-like distribution in the case of *haptoglobin* can be a result of its genetic polymorphism. In humans, three phenotypes of this protein exists and their molecular masses vary between 86 and 900 kDa. The differences in the structures of the three isoforms could explain its presence in all four ammonium sulphate precipitation levels, while the size differences influenced the size exclusion

chromatographic profiles. In addition, haptoglobin binds haemoglobin with great affinity. Molecular complexes could perform different characteristics than individual proteins, therefore they could accumulate in Library fractions of different traits.

The study of *complement component C9* revealed the possibility of binding to the albumin or immunoglobulin depleting media and/or albumin and consequently the lack of detectability of an antigen in the final Analyte Library fractions. Co-depletion of non-albumin type proteins by the Blue Sepharose is a known fact discussed in several studies. The investigation of the initial fractions obtained during the depletion steps could make the identification of potential biomarkers that cannot be detected in the final fractions possible.

As a result of the comprehensive screening with *alpha-2-macroglobulin* specific mAb, distribution of the antigen matched the expectations predicted based on its characteristics. As a globular protein, alpha-2-macroglobulin was assumably precipitated at lower ammonium sulfate concentrations, and as expected, it was represented mainly in the 35 or 45% ammonium sulfate precipitation fractions. As one of the largest plasma glycoproteins (molecular weight: 720 kDa) it mainly accumulated in the higher molecular weights size exclusion chromatography fractions (G1) as expected. Its presence in the lower molecular weight range could be the consequence of e.g. proteolysis or some other unknown phenomenon.

Hemopexin is a 57 kDa glycoprotein which binds haem with the greatest affinity (dissociation constant $<10^{-9}$ M), with equimolar stoichiometry. The size of the complex would not change to a great extent by the binding of one haem molecule, thus corresponding signals were observed in the 45%-G3 (lowest molecular weight) fraction suggesting the presence of individual hemopexin molecules or hemopexin/haem complexes. In certain proteins, e.g. matrix metalloproteinases (MMPs) there is a sequence called hemopexin domain, which can be homologous to hemopexin to various extents. It could be possible that the hemopexin specific monoclonal antibody recognized the homologous domains of MMPs and therefore reacted with the metalloproteinases of various sizes (e.g. positive signals in fractions 45%-G1 and 45%-G2).

Human serum albumin is known to be associated with several proteins and also binds fatty acids, steroid hormones, vitamins, amino acids and metabolites. The phenomena that albumin existed in the Library bound to these molecules and caused the 'smear'-like distribution of HSA – 28% of all fractions were considered as positive – should be taken into account. It is considered that some complexes formed by albumin were not depleted by Blue Sepharose chromatography.

Different outcomes were experienced in terms of the distribution of the proteins, but the majority of the results was in accordance with the expected distributions and/or was explainable by data in the literature. Therefore, the Analyte Library was considered to be appropriate for monoclonal antibody proteomics-based antigen identification.

5.3. Monoclonal antibody proteomics-based antigen identification

The selection and preparation of the antigens for mAb proteomics-based antigen identification was conducted as a multi-step experiment. The Library was screened with mAb-s of unknown proteins by means of a dot blot array system. The fractions considered as positives were subjected to the workflow of Western blot, immunoprecipitation, SDS-PAGE and mass spectrometric analyses. Approximately half (8) of the antigens corresponding to the studied mAb-s were successfully determined; this can be considered as a very good ratio in terms of the utilization of the workflow in the future as a possible tool of biomarker research.

5.4. Concluding remarks

The results suggested that the method developed during my PhD work was applicable for monoclonal antibody proteomics-based antigen identification. By the multi-step fractionation of the human plasma (Analyte Library generation), a sufficient amount of fractions was obtained of much less complexity than plasma for the subsequent studies. The dot blot screening of the Library was fast, simple and high-throughput. The plasma proteins distributed in the Analyte Library fractions in accordance with their physical, chemical and biochemical features, which also confirmed the successfulness of the separation and the applicability of the Library

for antigen identification. The monoclonal antibody proteomics-based antigen identification was successful since the antigens corresponding to eight antibodies could be identified. These findings are encouraging with respect to the application of the workflow in biomarker or other protein identification studies.

6. SUMMARY

Human plasma contains a wide concentration range of proteins (12 orders of magnitude) with diverse biological functions. Due to its high complexity, the characterization and identification of medium or low abundant proteins often represent a difficult and time-consuming task and require a large amount of sample. Monoclonal antibody proteomics uses a large number of monoclonal antibodies of unknown antigen specificity raised against complex biological samples. The antigen identification process includes the reaction of biological samples with antibodies, immunoprecipitation, SDS-PAGE analysis, Western blot validation and mass spectrometric identification.

The aim of this work was to generate and characterize an Analyte Library that contained hundreds of fractions of native plasma proteins, and to use the Library for antigen identification in conjunction with mAb proteomics. First the albumin and immunoglobulins were depleted from 500 mL normal human plasma followed by pre-fractionation by ammonium sulphate precipitation steps. Each precipitate was then further fractionated by size exclusion chromatography, followed by cation and anion exchange chromatography and hydrophobic interaction chromatography. To demonstrate the characterization process of the Analyte Library, the separation route of vitamin D-binding protein was followed in all major fractionation levels by dot blot and Western blot. Comprehensive screening of the Library with seven mAbs of known abundant antigens of biomarker potential was also part of the study to reveal the distribution of these proteins. The identification of antigens corresponding to the mAb-s was conducted by the workflow of monoclonal antibody proteomics.

A 783-fraction containing Analyte Library comprising of native protein fractions was generated. The individual fractions contained some 10 or 100 proteins. As a result of the characterization, the Library was considered as applicable for monoclonal antibody proteomics-based antigen identification; and the antigens corresponding to eight mAbs have been successfully identified.

My studies demonstrated that the Analyte Library and the developed method was applicable for monoclonal antibody proteomics-based antigen identification, and

suggests the potential of the developed workflow in biomarker and protein identification research.

7. NEW RESULTS, FINDINGS AND THEIR SIGNIFICANCE

1. Starting from 500 mL human plasma, a 783-fraction containing Analyte Library was generated by the multi-step fractionation of the human plasma proteome. Besides the overlaps between the closely related ones, the compositions of the Library fractions differed from each other in accordance with the applied specific depletion methods and the precipitation and chromatographic techniques based on the various physical, chemical and biochemical features of the proteins. All chromatography and precipitation steps were carefully designed aiming to maintain the native forms of the intact proteins throughout the fractionation process.
2. By the characterization of the Analyte Library, it was considered as applicable for monoclonal antibody proteomics-based antigen identification. The accumulation route of vitamin D-binding protein was traced in the different levels of the fractionation process. Gradual decrease in the complexities of the fractions in parallel with the progression of the fractionation was experienced and final fractions were found in which the protein accumulated in almost exclusively. After the high-throughput screening of the Library with monoclonal antibodies specific for known plasma proteins, the distributions of the antigens matched the expectations based on physical, chemical, biochemical traits and literature data.
3. The Analyte Library fractions were successfully utilized for monoclonal antibody proteomics-based antigen identification. By applying the workflow of dot blot, Western blot, immunoprecipitation and mass spectrometric analysis, the cognate antigens corresponding to eight monoclonal antibodies could be identified.

8. PUBLICATIONS

8.1. Publication list

UNIVERSITY AND NATIONAL LIBRARY UNIVERSITY OF DEBRECEN
KENÉZY LIFE SCIENCES LIBRARY

Register Number: DEENKÉTK/312/2013.
Item Number:
Subject: Ph.D. List of Publications

Candidate: András László Kovács
Neptun ID: XQQ9C4
Doctoral School: Doctoral School of Molecular Medicine

List of publications related to the dissertation

1. **Kovács, A.**, Patai, Z., Guttman, A., Kádas, J., Takács, L., Kurucz, I.: Fractionation of the human plasma proteome for monoclonal antibody proteomics-based biomarker discovery 2: Antigen identification by dot blot array screening.
Electrophoresis. "accepted by publisher", 2013.
DOI: http://dx.doi.org/10.1002/elps.201200677
IF:3.261 (2012)

2. **Kovács, A.**, Guttman, A.: Medicinal chemistry meets proteomics: Fractionation of the human plasma proteome.
Curr. Med. Chem. 20 (4), 483-490, 2013.
DOI: http://dx.doi.org/10.2174/0929867311320040001
IF:4.07 (2012)

3. **Kovács, A.**, Sperling, E., Lázár, J., Balogh, A., Kádas, J., Szekrényes, Á., Takács, L., Kurucz, I., Guttman, A.: Fractionation of the human plasma proteome for monoclonal antibody proteomics-based biomarker discovery.
Electrophoresis. 32 (15), 1916-1925, 2011.
DOI: http://dx.doi.org/10.1002/elps.201100018
IF:3.303

H-4032 Debrecen, Egyetem tér 1. e-mail: publikaciok@lib.unideb.hu

List of other publications

4. Bódi, Z., Pepó, P., **Kovács, A.**: Morphology of tassel components and their relationship to some quantitative features in maize.
Cereal Res. Commun. 36 (2), 353-360, 2008.
DOI: http://dx.doi.org/10.1556/CRC.36.2008.2.15

5. Bódi, Z., Pepó, P., **Kovács, A.**: Investigation on direct and reciprocal crosses in maize (Zea mays L.).
Acta Agron. Hung. 56 (2), 179-185, 2008.
DOI: http://dx.doi.org/10.1556/AAgr.56.2008.2.7

6. Bódi, Z., Pepó, P., **Kovács, A.**, Széles, É., Györi, Z.: Macro- and microelement contents of blue and red kernel corns.
Cereal Res. Commun. 36 (1), 147-155, 2008.
DOI: http://dx.doi.org/10.1556/CRC.36.2008.1.15

7. Pepó, P., **Kovács, A.**: Histological aspects of the cryopreservation of potato.
Acta Agron. Hung. 56 (3), 341-348, 2008.
DOI: http://dx.doi.org/10.1556/AAgr.56.2008.3.10

8. Pepó, P., **Kovács, A.**, Tóth, S., Bódi, Z.: Investigation of maize hybrids based on direct and reciprocal crosses.
Cereal Res. Commun. 35 (2/2), 913-916, 2007.
DOI: http://dx.doi.org/10.1556/CRC.35.2007.2.187
IF:1.19

9. Mádi, A., Kárpáti, L., **Kovács, A.**, Muszbek, L., Fésüs, L.: High-throughput scintillation proximity assay for transglutaminase activity measurement.
Anal. Biochem. 343 (2), 256-262, 2005.
DOI: http://dx.doi.org/10.1016/j.ab.2005.05.034
IF:2.67

10. Page, A.R., **Kovács, A.**, Deák, P., Török, T., Kiss, I., Dario, P., Bastos, C., Batista, P., Gomes, R., Ohkura, H., Russell, S., Glover, D.M.: Spotted-dick, a zinc-finger protein of Drosophila required for expression of Orc4 and S phase.
EMBO J. 24 (24), 4304-4315, 2005.
DOI: http://dx.doi.org/10.1038/sj.emboj.7600890
IF:10.053

Total IF: 24.547

Total IF (publications related to the dissertation): 10.634

The Candidate's publication data submitted to the Publication Database of the University of Debrecen have been validated by Kenezy Life Sciences Library on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

26 September, 2013

8.2. Other publications

8.2.1. Articles in Hungarian

Bódi Z, Pepó P, **Kovács A** (2007) A hektolitertömeg értékének változása eltérő genotípusok esetén kukoricánál (Zea mays L.) [Change of test weight in different maize (Zea mays L.) genotypes]. *Acta Agronomica Óváriensis* 49, 569–573.

Bódi Z, Pepó P, **Kovács A** (2007) Kék és vörös szemszínű kukoricák nemesítési értékelése [Evaluation of blue and red corn landraces]. *Agrártudományi Közlemények/Acta Agraria Debreceniensis* 26, 239–243.

Bódi Z, Pepó P, **Kovács A** (2007) Keresztezési irányok vizsgálata kukoricánál [Investigation of crossing directions in maize (Zea mays L.)]. *Agrártudományi Közlemények/Acta Agraria Debreceniensis* 27, 43–48.

8.2.2. Lectures related to the subject of the dissertation

Kovács A: Generation and characterization of a comprehensive analyte library representing the human plasma proteome. *PhD Symposium of the Doctoral School of Molecular Medicine, University of Debrecen*, Debrecen, 1 June 2010.

Kovács A: Fractionation of the human plasma proteome for monoclonal antibody proteomics based biomarker discovery. *Annual Symposium of the Doctoral School of Molecular Medicine, University of Debrecen*, Debrecen, 6 June 2011.

8.2.3. Other lectures

Kovács A: Ismert transzglutaminázokkal homológ fehérjék vizsgálata különböző rovarfajokban [Study of proteins of different insect species homologous to known transglutaminases]. *Scientific Student Meeting of the Faculty of Medicine at the Medical and Health Science Center of the University of Debrecen*, Debrecen, Hungary, 4 April 2000. In Hungarian.

Kovács A: Ismert transzglutaminázokkal homológ cDNS-szakasz vizsgálata ecetmuslicában [Study of a fruit-fly cDNA strand homologous to known transglutaminases]. Scientific Student Meeting of the Faculty of Medicine at the Medical and Health Science Center of the University of Debrecen, Debrecen, Hungary, 19 February 2001. In Hungarian.

Mádi A, Kárpáti L, **Kovács A**, Muszbek L, Fésüs L: High-throughput scintillation proximity assay for transglutaminase activity measurement. *8th International Conference on Protein Crosslinking and Transglutaminases (PCL8)*, Lübeck, Germany, 2 September 2005.

Pepó P, Bódi Z, **Kovács A**: A ciklotron alkalmazásai a növénygenetikában és nemesítésben [Application of the cyclotron in plant genetics and breeding]. *Scientific meeting on the application of research results in physics*, Debrecen, Hungary, 15 November 2006. In Hungarian.

8.2.4. Posters related to the subject of the dissertation

Kovács A, Sperling E, Takács L, Kurucz I, Kádas J, Guttman A: Generation and characterization of a comprehensive analyte library representing the human plasma proteome. *25th International Symposium on Microscale Bioseparations (MSB 2010)*, Prague, Czech Republic 21–25 March 2010.

Kovács A, Sperling E, Lázár J, Balogh A, Kádas J, Szekrényes Á, Takács L, Kurucz I, Guttman A: Humán plazma frakcionálása monoklonális ellenanyag-proteomikai alapú biomarker-felfedezéshez [Fractionation of the human plasma proteome for monoclonal antibody proteomics-based biomarker discovery]. *1st National Conference of the Hungarian Chemical Society*, Sopron, Hungary, 22–25 May 2011. In Hungarian.

Kovács A, Sperling E, Mittermayr S, Lázár J, Balogh A, Kádas J, Szekrényes Á, Takács L, Kurucz I: Fractionation of the human plasma proteome for monoclonal antibody proteomics based biomarker discovery. *36th International Symposium on High-Performance Liquid Phase Separations and Related Techniques (HPLC 2011)*, Budapest, Hungary, 19–23 June 2011.

Kovács A, Patai Z, Sperling E, Lázár J, Balogh A, Kádas J, Szekrényes Á, Takács L, Kurucz I, Guttman A: Biomarker screening of the human plasma proteome with disease specific monoclonal antibodies. *8th International Interdisciplinary Meeting on Bioanalysis (CECE 2011)*, Brno, Czech Republic, 3–4 November 2011.

Kovács A, Patai Z, Kádas J, Takács L, Kurucz I, Guttman A: Monoclonal antibody proteomics: Antigen identification by mini dot-blot array screening. *27th International Symposium on MicroScale Bioseparations and Analyses (MSB2012)*, Geneva, Switzerland, 12–15 February 2012.

Patai Z, **Kovács A**, Kurucz I, Takács L, Guttman A: Monoclonal antibody proteomics based characterization of a human plasma proteome library. *38th International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC 2012)*, Anaheim, CA, USA, 16–21 June 2012.

8.2.5. Other posters

Kovács A, Blaskó B, Mádi A, Fésüs L: Ismert transzglutaminázok génjeivel homológ cDNS-klón vizsgálata ecetmuslicában [Investigation of a fruit-fly cDNA clone homologous to known transglutaminase genes]. *6th Workshop of the Molecular Biology Section of the Hungarian Biochemical Society*, Sárospatak, Hungary, 14–17 May 2001. In Hungarian.

Blaskó B, Mádi A, **Kovács A**, Fésüs L: Protein diszulfid izomeráz jellemzése Caenorhabditis elegansban [Characterization of the protein disulfide isomerase from Caenorhabditis elegans]. *6th Workshop of the Molecular Biology Section of the Hungarian Biochemical Society*, Sárospatak, Hungary, 14–17 May 2001. In Hungarian.

Blaskó B, **Kovács A**, Mádi A, Fésüs L: Protein diszulfid izomeráz jellemzése Caenorhabditis elegans-ban [Characterization of the protein disulfide isomerase from Caenorhabditis elegans]. *10th Days on Cell and Developmental Biology*, Siófok, Hungary, 27–29 March 2002. In Hungarian.

Kovács A, Blaskó B, Mádi A, Fésüs L: Ecetmuslica-transzglutamináz klónozása és vizsgálata [Cloning and investigation of the fruit-fly transglutaminase]. *7th Workshop of the Molecular Biology Section of the Hungarian Biochemical Society*, Keszthely, Hungary, 14–17 May 2002. In Hungarian.

Blaskó B, **Kovács A**, Mádi A, Fésüs L: Funkcionális transzglutamináz-e a C. elegans protein diszulfid izomeráza? [Is the protein disulfide isomerase of C. elegans a functional transglutaminase?] *7th Workshop of the Molecular Biology Section of the Hungarian Biochemical Society*, Keszthely, Hungary, 14–17 May 2002. In Hungarian.

Kovács A, Blaskó B, Mádi A, Fésüs L: Cloning and characterization of the fruit-fly transglutaminase. *7th International Conference on Transglutaminases and Protein Crosslinking Reactions*, Ferrara, Italy, 14–17 September 2002.

Kovács A, Blaskó B, Mádi A, Deák P, Fésüs L: Transzglutamináz jellemzése ecetmuslicában [Characterization of the fruit-fly transglutaminase]. *8th Workshop of the Molecular Biology Section of the Hungarian Biochemical Society*, Tihany, Hungary, 12–15 May 2003. In Hungarian.

Kovács A, Pál M, Blaskó B, Mádi A, Deák P, Fésüs L: Az ecetmuslica-transzglutamináz jellemzése [Characterization of the fruit-fly transglutaminase]. *9th Workshop of the Molecular Biology Section of the Hungarian Biochemical Society*, Sopron, Hungary, 10–13 May 2004. In Hungarian.

Erdei É, Pepó P, Tóth Sz, **Kovács A**, Szabó B: A silócirok (Sorghum dochna L.) beltartalmi értékeinek vizsgálata alternatív energia előállítás céljaira [Investigation on the nutritional values of the sorghum (Sorghum dochna L.) with the purpose of alternative energy production]. *14th Scientific Day on Plant Breeding*, Budapest, Hungary, 12 March 2008. In Hungarian.

8.2.6. Educational, popular

Kovács A, Fésüs L (2000) Cell death in plants compared to animals. *Proceedings of EMBO Lecture Course Molecular and Cellular Biology: From Plant to Human*, Debrecen, Hungary (Editors: Krasznai Z, Mátyus L, van Duijn B): 8–22.

Pepó P, Tóth Sz, Bódi Z, Kovácsné Oskolás H, **Kovács A**, Erdei É, Szabó E (2007) Szántóföldi növények genetikája, nemesítése és biotechnológiája [Genetics, breeding and biotechnology of field plants] (Editor: Pepó P). University of Debrecen, Debrecen, Hungary. ISBN 978-963-9732-18-6. In Hungarian.

Pepó P, Bódi Z, **Kovács A** (2007) Növénybiotechnológiai praktikum [Practical guide to plant biotechnology] (Editors: Pepó P, Bódi Z). University of Debrecen, Debrecen, Hungary. ISBN 978-963-9732-19-3. In Hungarian.

Pepó P, **Kovács A** (2009) Environmental danger of genetic engineering. *Pollution and Water Resources Columbia University Seminar Proceedings* (Editor: Halasi-Kun GJ) XXXVIII-XXXIX (2008-2009 Scientific and Social-Institutional Aspects of Central Europe and USA): 286–296.

Kovács A (2009) Szőlőmag olaj – hidegen sajtolt szőlőmag olaj [Grape seed oil – cold pressed grape seed oil]. *Naturland Egészségmagazin* [*Naturland Health Magazine*] IV/4, 25. In Hungarian.

The doctoral training program was supported by the TÁMOP-4.2.2/B-10/1-2010-0024 project. The project is co-financed by the European Union and the European Social Fund.