Final publishable summary report

Grant agreement no.:	282506
Project acronym:	PROT-HISPRA
Project title:	HIgh Speed PRoteomics Analysis (Prot-HiSPRA: Solving the bottlenecks of proteomics technologies for time sensitive proteome driven medical decisions
Funding Scheme:	Collaborative Project (CP) - Small or medium-scale focused research project (STREP)

Date of latest version of Annex I against which the assessment will be made:

Periodic report:	Final publishable summary report
Period covered:	from M01 to M42 (01. October 2011 - 31. March 2015)

Project co-ordinator name:	Goran Mitulovic
Project co-ordinator organisation:	MEDIZINISCHE UNIVERSITAET WIEN

Table of Contents

Final publishable summary report....3

Executive summary....3

Summary description of project context and objectives....5

Description of the main S&T results/foregrounds....8

Potential impact, main dissemination and exploitation activities....11

4.1. Final publishable summary report

Executive summary

On October 1 2011 Prot-HiSPRA Consortium started the activities on developing hardware, software, and sample preparation methods for proteomics analysis of minute protein amounts in biological samples. The Consortium set to develop fast, reproducible, and as simple as possible on-line methods for application with biological samples relevant for clinical diagnosis. We have chosen to use spent culturing media utilized in in-vitro fertilization (IVF), analyze the secretome of fertilized oocytes, which are to be re-implanted into women's uterus during the IVF procedure, and to search for distinct proteins, which could help to better characterize positively and negatively labeled blastocysts.

The idea was based on assumption that growing embryos secrete distinct proteins during their development, and that these proteins can be used to distinguish between embryos with a higher potential for successful pregnancy upon first implantation.
Classical proteomics approach for analysis of secreted proteins involves time consuming sample preparation steps and cannot be employed for analysis of samples in situation where clinicians need rapid answers for fast proceedings. Classical proteomics analysis, usually, requires 18-24 hours to deliver the result. However, this time span is too long for the IVF procedure during which the results must be either immediately available, or at latest, within a few hours.
In order to provide the analysis of secreted proteins fast, reliable, and reproducible, new analytical approaches for sample preparation, peptide separation, and data analysis are needed. These methods should enable fast depletion of high abundant proteins form spent culturing media; in this case it was human serum albumin, fast protein digestion – minutes instead of hours – and peptide separation with highest possible separation resolution. Peptide and protein identification would be performed by means of mas spectrometry and database search.
This goal could be only reached by establishing new methods for sample preparation and developing new hardware for online depletion, online digest, and parallel chromatographic separation of multiple samples simultaneously. At the same time, software solutions for prediction of peptide retention times in HPLC and synchronization of separation and mass spectrometric identification. At the same time, software solutions for prediction of peptide retention times in HPLC and synchronization of separation and mass spectrometric identification were also develped.

The hardware needed for sample preparation (sample amounts were approximately at 40 – 60 µl) included the use of new, monolithic-based, columns bearing immobilized enzymes (trypsin and pepsin) and TiO_2 nanoparticles for sample digest, trapping of modified peptides, and nano HPLC columns with alternative stationary phases (normal phase, different HILIC phases) for separations of peptides in parallel manner and with high sample throughput, which, in this case, was less than one hour analysis time.
The technologies, which are currently available for each of these analysis steps, hav limitations or they cannot be readily automated and synchronized. In addition, new hardware is needed for operating depletion, digestion, and separation columns simultaneously and on-line. When analysis includes sample loading, column equilibration and cleaning in addition to the nano chromatographic system used, a multivalve, low void-volume switching system is needed.

Further, a model system was needed to mimic the biological system and be used for method development instead of real IVF samples. The Consortium succeeded in creating a high dynamic range model samples for mimicking the complexity of targeted biological materials and developing online digestion methods based on enzymes immobilized on monolithic support. Further, new software for retention time prediction and calibration was developed for reversed phase and HILIC separations. Although one partner (NEPAF-CELS) left the Consortium at the very beginning of the project, we were able to gain another partner (BIA Separations) and successfully integrate it into the running project.

Having developed hardware, synthesized nanoparticles from different materials, and applied the particles for copolymerization and immobilization in monolithic columns by BIA Separations, the Consortium made a significant step forward in achieving the project's goals.
The Consortium presented the results, which were achieved at different international and national scientific conferences, and also published the first findings in peer-reviewed journals. In order to make the project activities more effective, partners visited each other and cooperated in developing methods and hardware on-site at respective laboratories.

For the second reporting period, the Consortium concentrated its activities on applying developed techniques and methods for analysis of clinical samples and further optimization of online depletion and digestion methods.

BIA Separations and University of Debrecen successfully completed the synthesis and hardware development of the anti-HSA depletion column along with the monolithic column bearing immobilized nanoparticles for trapping of posttranslationaly modified peptides.

Medical University of Vienna continued optimizing the digestion strategies for clinical samples based on results obtained within the first half of the project and based on new depletion columns synthesized by BIA Separations.

During the IVF procedure, two types of IVF media are being applied. Nominally, these media should have constant composition and be reproducible from batch to batch. However, this is not the case and a range of batches of IVF media from two manufacturers were analyzed using the tools that were developed during the first stage of the project.
All experiments were performed using the hardware developed for parallel albumin depletion, online digestion, and parallel peptide separation. The hardware was fully developed and incorporated for the use with low amounts of clinically available samples.
In addition to commonly used multi-embryo approach, IVF samples derived from single-bred embryos were introduced in order to gain a better insight into the early development and secretion of possible markers.
By applying new depletion and digestion columns, it was possible to significantly improve and speed up the analysis from more than 24 hours to less then two hours total analysis time, which was on of the main goals for the clinical samples.

The analysis of identified proteins and peptides is still ongoing due to a large amount of generated data; however, some identified proteins are likely to be of great importance to developing embryos based on current results and comparison with clinical data on successful pregnancies. A further, interesting and unexpected, finding was the identification of large amount of proteins in IVF media, from two different manufacturers, which were not declared in the product sheet. It is still not clear on how much influence these proteins have on early development of the embryo since they showed variations in protein content and protein amounts present. The results are indicative because the amount and the type of proteins differ from batch to batch and there is also a difference in the success rate of the IVF procedures, which seem to be cyclical. It was not investigated on whether the type of IVF media can be connected to this information but the amount of generated data is large and this data analysis is ongoing.

Summary description of project context and objectives

The Prot-HiSPRA Consortium defined the following scientific and technological objectives:

(1) Development of the hardware needed for the high throughput on-line sample preparation, which can be applied in combination with any commercially available HPLC system.

The sample preparation is, along with the online digest, the most complex part of the project. Identification of proteins in biological samples is a complex procedure simply due to the extremely large amount of high abundant proteins that “cover and mask” the lower abundant peptides. For the biological sample selected, spent culture media for In-Vitro Fertilization Procedure (IVF), serum albumin, which makes for more than 80% of the sample (concentration is ~10 mg/ml), must be removed. The removal of such a large amount of interfering albumin can be performed either by using specific antibodies, chromatographic separation (size exclusion) or ultrafiltration using filters with different molecular weights (MW) cutoffs.

First columns with immobilized specific antibodies were purchased from Agilent, BioRad, and Millipore. Although these columns are designed to bind the albumin and (in some cases) other abundant proteins, it was not possible to achieve the goal and remove the albumin in sufficient amount due the extremely large amount of albumin (5-10mg/ml) and a second and third depletion steps were needed. Even with this method, significant amounts of albumin were detectable in processed samples resulting with strong interferences for detection of peptides from other proteins.

Ultrafiltration of IVF samples delivered mixed results. Expected removal of albumin, when using Umicon MW filters (Millipore), was not achieved with filters with pores of 50 kDa. Using filters with a MW cutoff of 30 kDa and with 10 kDa resulted with lower amount of detectable serum albumin, and it was possible to analyze the sample and detect significantly more proteins compared to results obtained with immobilized antibodies.

However, these results were also not fully satisfying and the next step in optimizing depletion procedures with antibodies was the use of monolithic columns prepared by BIA Separations (BIA) with immobilizing nanoparticles carrying antibodies developed by University of Debrecen (UDE) and directly immobilizing antibodies on the monolithic surface. Monolithic columns carrying immobilized TiO_2 nanoparticles were tested for trapping of posttranslationaly modified peptides generated in the IMER, which resulted with identifications of a number of peptides present only in positive IVF samples (positive labeled by physician prior to implantation).

The hardware (HW) developed during the project enables the use of multiple switching and selection valves combined with pumps able to deliver flows from 10 nl/min up to 300 µml/min. This combination enables fast and seamless sample loading and digest using fast flow rate of 20 – 50 µl/min switch to very low flow rates (100 nl/min or lower) for peptide digestion or trapping on monolithic columns with immobilized nano particles.

All modules of the column-switching system are completed and operational. All modules are fully controlled by the software, although currently only through Chromeleon 6.8.

The column-switching module was designed to be operated as an independent device, which can be combined with any commercially available pumping module to be applied for sample depletion and digestion. Column switching and selection module is equipped with 1/32''' valves enabling low void volume and high flexibility in terms of combinations, depending on separation problems to be approached. The pump module houses two IDEX-produced pumps for sample loading and for cleaning and equilibrating of IMER columns. These pumps can be operated at different flow rates; both are equipped with pressure sensor and the nano-flow pump with a flow rate sensor.

(2) Study on new, selective and targeted sample specific nano-particles that can be applied to sample depletion and/or enrichment methods

The final goal of this part of Prot-HiSPRA was to create nanoparticles for isolation of proteins and small molecules with molecular weights less than approximately 20 kDa. We assumed that this part of the secretome of targeted biomaterial could be rich source of putative biomarkers. Firstly, three different types of magnetic and non-magnetic ferrite based nanoparticles were synthesized at the University of Debrecen and tested at the MUW.

An alternative material, which seemed to be a better candidate for binding antibodies to its surface was TiO_2 and further studies were focused on TiO_2 sol and on binding the Human Serum Albumin. The TiO_2 sol and the solution of HSA were mixed at different ratios while pH of the solution was maintained at 5.0. Depending on the ratio of the components, three different observations were made. When the albumin concentration in the samples was low (3:1 TiO_2: HSA by weight) aggregation and sedimentation took place and no free albumin could be detected in the clear solution. The same sensibilization was observed at high HSA ratio (above 1:3). In the medium range, a stabilization effects were detected, which can be explained by steric stabilization of the adsorbed protein on the TiO_2 surface. The final dispersion was stable enough, and no aggregation or sedimentation was detected during several days. Scattering of the stable nanoparticles prevents the accurate measurement of the non-adsorbed albumin concentration.

Further, Ti/Si concentration was determined by ICP method in order to optimize the applied derivatives and conditions (media) for the highest degree of activation. The ratio determination is also important to calculate the amount of biotin, which was applied in the next reaction step. Biotinylation of the activated nanoparticles was performed in organic media using small excess of biotin derivatives in order to maximize the amount of linker on the surface. The nonbound biotins were removed by consecutive washing. The antibody conjugation was performed in aqueous media after biotinylation and the liquid phase was replaced for a buffered solution. To ensure the absence of the original organic solvents, the buffer exchange was performed in multiple steps, which were also necessary for preservation of the size distribution.

(3) Integration of high throughput (HTP) affinity enrichment methods and mass spectrometry techniques for proteins and protein complexes with an emphasis on selective protein enrichment from complex matrices with high abundance impurities.

The first aim of this part of the project was the integration of the developed analytical systems, including depletion kit for Human serum albumin (HSA) from IVF media.

The second aim was the identification of the marker candidates for further studies, testing, and validations in the post-project follow-up efforts. In the first step of meeting this objective, a quantitative analysis of protein content of the IVF media was characterized employing the procedure developed in Task 5.1. This analysis also involved determination of the minimum number of samples that were needed for validating biomarker candidates in the IVF media using developed high throughput technologies.

(4) Improvement of on-line sample treatment and minimization of manual sample preparation steps and techniques required for fast analysis of clinical samples regardless of their origin.

The online digestion system (IMER) was tested using columns prepared by BIA. Digestion system was tested in terms of digestion efficiency, digestion reproducibility, and system stability. Settings such as buffer composition, column temperature, digestion time, flow rate, pressure applied on digestion column, and the amount of proteins injected. Further, buffer concentration used for sample injection and the sample volume was assessed in terms of the most effective digestion. The carry-over of samples from one injection to the next was investigated and the method was developed for removing sample remnants from both the digestion column and the reversed phase (RP) trap and separation column. Parallel loading and separation setup was established for faster sample loading and processing.

IMERs based on monolithic columns were successfully used for online digestion of IVF media and other biological samples such as urine and processed human serum. The final method for online sample digests involved improved and optimized buffers and time settings. Due to a larger number of clinical samples, a second online system was also introduced for the digest procedure and the measurements are still in progress as of July 2015.

(5) Development of the software interfaces to combine both spatial and temporal data flow from real-time nano-HPLC-MS/MS analyses for improved efficiency and reliability in both the time scale and the quantitation.

Institute for Energy Problems of Chemical Physics (INEP) and MUW were developing model samples for simulation of the online digest of biological samples. These samples were modeled in order to mimic the biological samples in terms of dynamic range and were analyzed using gradient settings and mass spectrometric (MS) approach as used for biological samples. Further, INEP has developed the software for enabling retention time (RTP) prediction of tryptic peptides on both reversed phase (RP) and hydrophilic interaction chromatography (HILIC) columns. The software has been implemented as an open source library (libBioLCCC) and integrated within the proteomic data-mining framework called Pyteomics (http://pythonhosted.org/pyteomics/). Pyteomics is a free downloaded fully documented suite of functions in Python programming language (https://pypi.python.org/pypi/pyteomics).

Description of the main S&T results/foregrounds

The following describes the main results that have been achieved by the Prot-HiSPRA Consortium:

Sample preparation technologies and hardware development

The decision to use the spent culture media from In-Vitro Fertilization was made because it is of great medical interest, and it could shed light on the embryos' secretome. Since the spent culture media are usually discarded after the embryo has been either transferred from one type of culture media into the other or prior to the implantation or storage it is a quite cheap and reliable source of biological sample.

For successful detection and analysis of biologically important proteins in the spent culture media samples, high abundant proteins must be removed. Conventional methods are not suited for depletion of high abundant analytes from samples of very low volume typical in clinical practice. Therefore, new depletion methods tailored to analyze these samples had to be developed, optimized, and tested for optimal performance.

The model sample with high dynamic range in protein concentration was prepared in order to mimic the complexity of the biological samples intended to be analyzed. The use of nanoparticles was considered primarily for depletion of albumin because of the large active surface and because of possible immobilization in monolithic columns that can be integrated into complete analytical system.

The online digestion system (IMER) has been tested using immobilized digestion columns prepared by BIA. One of the major problems for developing online digestion is the establishing of digestion times and the optimization of the flow rate through the digestion column. Current column design, which is a subject to change, does not allow for exact time calculation and exact flow rate settings. The use of conventional digestion buffers such as ammonium bicarbonate had to be abandoned because of CO_2 development. Therefore, two additional buffers have been tested: ammonium acetate (AmAc) and triethylammonium bicarbonate (TEAB). No gas development was observed with TEAB and the pH could be set to maximize the protein digestion and the same was true for AmAc. However, the use of AmAc was not as successful as the use of TEAB. Addition of $CaCl_2$ increased the digestion efficiency but interfered ionization in the mass spectrometry and need to be further tuned.

The influence of temperature (T) on IMER, especially on digestion efficiency was investigated and, surprisingly, no large differences were observed for operations at room T and at 40°C. However, the IMER operation was set at controlled T (40°C), and the separation columns were operated at 60°C.

Albumin depletion is usually performed using large amounts of sample and applying offline digestion on large depletion columns. Small sample volumes made the use of miniaturized depletion, digestion, and separation columns a must.

High-throughput separation technologies

High throughput technologies for separation of digested proteins from biological samples were developed including peptide separations using HILIC, Normal Phase (NP), and monolithic columns combined with reversed phase HPLC. Development of HILIC and NP separation methods, as truly

orthogonal separation methods to RP, was intensively studied in order to determine the suitable separation conditions and to filter out the suitable separation phases.

The separation speed depends on columns used and the complexity of the sample. In combination with online digestion and separation on either parallel nano separation columns or monolithic separation columns, the separation times ranging from 80-180 minutes could be reduced to 60 minutes in total. Further optimization of separation conditions and operation of three separation column will decrease the separation time by further 30%. Currently, the multidimensional separation platform is operational using the hardware developed by MayLab. Retention time prediction software enabled designing gradient conditions in order to achieve the fastest possible separation of peptides of interest without information loss has been developed and tested.

The chromatographic separations system (depletion, IMER, RP or multidimensional separation) is hyphenated with the mass spectrometer for detection and analysis of separated peptides. It is very important to ensure that no salts or nonvolatile buffers from any of the separation steps are introduced into the mass spectrometer. A sample is loaded to a mass spectrometer from three separation columns and the flow from the nano separation column or from the monolithic column is directed into the MS through the nano UV cell. Prior to the nano UV cell, a flow selection valve is diverting the flow from the separation column either into the detector, or a waste.

Data mining and processing

One of the main goals was a study on the performance of different available MS/MS search engines in identification of targeted proteins. The study was started with general evaluation of the performance of three selected search engines: Mascot, X!Tandem, and OMSSA. These search engines are the most widely used in proteomics and the latter two are based on the open-source algorithms. The software tools for specific applications targeted in the project were developed using an open-source framework and the specific goal of this evaluation was an understanding the performance of the open-source search engines compared with commercial one. It was found that the simultaneous use of several search engines increases the reliability of the bottom-up analysis as one of the possible strategies, yet, for the price of loosing up to 80% of the sampled peptides. The alternative strategy employing a combination of the search results from different search engines will result in added FDR level even if the peptides passed the respective identity thresholds in each of the engines taken separately. The major finding from the studies along the Task 3.1's objectives was the realization that for the dynamic range samples (such as the samples for targeted medical problem, IVF) the majority of the true positive identifications, specifically for the lower concentration proteins are falling behind the identity threshold set by any of the tested search engines. The demonstrated losses could be up to 80% of all identified tandem mass spectrometry spectra.

A further issue was to evaluate the applicability of chromatography information such as retention times and elution orders to receiving a priori information about the sequences of targeted peptides and proteins. The specific objective of this evaluation is to probe the combination of LC and MS or MS/MS data to use two dimensional protein search space. Based on the results obtained within Task 3.1 these objectives were further extended to study on the multi-dimensional search space for the peptides that includes six peptide features (descriptors). The idea behind this extension was developing a strategy of identification assessments that achieves a higher level of sensitivity at the same FDR level. In the core of this strategy is developing filters that allow distinguishing between correct and incorrect peptide assignments that fall below the set MS/MS based confidence level. To achieve this goal a novel scoring scheme complementary to the MS/MS-based scoring

dimension has been proposed and extensively tested using several independently obtained datasets.

In addition, a main part was developing a user-friendly software that integrates the predictive chromatography for LC data dependent sample analysis and characterization with MS data mining. During the project the software package PyBioLCCC for peptide de novo retention time prediction using the chromatographic model BioLCCC has been developed and integrated into the open-source proteomics library Pyteomics. The software was made freely available at http://pythonhosted.org/pyteomics.biolccc via domain theorchromo.ru. The latter one was developed for free on-line use by the community for optimization of peptide separation protocols and de novo retention time prediction. Within the Pyteomics library the retention time prediction function are integrated with other data mining capabilities such as mass calculations, sequence parsering, etc. A full documentation for the predictive chromatography package has been completed as well. In addition to the on-line version the INEP team has developed a desktop version of BioLCCC-based retention time prediction software, TheorChromo, with graphical user interface allowing optimization of a number of chromatographic parameters such as the automatic prediction model calibration, column factor adjustment, column geometrical parameters selection, temperature, etc.

Multi-dimensional separation as one of the major project's objectives is being developed and probed by MUW, MayLab and INEP teams. With this regard the predictive chromatography approach utilized in the project needs re-calibration of the peptide retention time prediction BioLCCC model for the selected (or a combination of) separation platforms. Due to the intrinsic nature of the BioLCCC model this re-calibration requires limited set of standard peptides to re-determine all phenomenological parameters of the model. These phenomenological parameters depend on the type of selected stationary phase, mobile phase constituents, and selected ion-pairing reagents.

Nano-particles for sample treatment

The final goal of this WP was to create nanoparticles for isolation of proteins and small molecules with molecular weights less than approximately 20 kDa since this part of the proteome of targeted biomaterial is thought to be a rich source of biomarkers. Developed nanoparticles were used as stand-alone I a suspension and as immobilized and co-polyerized particles on monolithic support.

Validation of developed high throughput laboratory methodologies

The main result was the integration of developed systems (WP1 and WP2) including depletion kit for Human serum albumin (HSA) from IVF media; on-line digestion reactor cartridge for protein mixtures prior to rapid 2D separation of the resulting peptide mixture, as well as detection and quantitation of biomarker peptides utilizing information obtained using enhanced peptide identification algorithms developed within the scope of WP3.
The identification of the marker candidates for further studies, testing, and validation in the post-project follow-up efforts was done.

Monolithic column and IMER development

The development of a monolithic platform for the efficient and selective removal of highly abundant HSA protein from the IVF sample followed by a subsequent protein (trypsin and/or pepsin) digestion of the remaining proteins coupled to the MS detection.

Potential impact, main dissemination and exploitation activities (10 pages)

Prot-HiSPRA organized dissemination activities towards the scientific community and mainly industry in the relevant fields. Ten scientific publications have been published, several more are still in progress. During more than 20 oral talks at congresses, conferences and workshops more than 10.000 people have been reached and informed about Prot-HiSPRA project and results.

The Consortium, in addition, set up a project website and five platform and databases that are operated as open source for the relevant experts in bioinformatics, chemistry, and medical issues.

Exploitation and protection of results were one of the key issues within Prot-HiSPRA. Taking into account that two companies are represented within the Consortium, it was a goal to achieve project results that can be taken up by the industry and into a product development process. Several exploitable results have been defined and discussed and two inventions have been prepared for patent protection.

Prot-HiSPRA organized several workshops to inform the industry and experts and closed the project with a final Prot-HiSPRA symposium in Vienna, Austria.

Website: http//www.prot-hispra.eu
Mr. Goran Mitulovic