Journal of Proteomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Proteomic analysis of first trimester maternal serum to identify candidate biomarkers potentially predictive of spontaneous preterm birth

Arlene M. D'Silva[a], Jon A. Hyett[b], Jens R. Coorssen[c,d,*]

[a] Department of Molecular Physiology, The Molecular Medicine Research Group, School of Medicine, Western Sydney University, Campbelltown, NSW 2150, Australia
[b] Department of High Risk Obstetrics, RPA Women and Babies, Royal Prince Alfred Hospital, University of Sydney, Sydney, SW 2050, Australia
[c] Department of Health Sciences, Faculty of Applied Health Sciences, Brock University, St. Catharines, ON L2S 3A1, Canada
[d] Faculty of Mathematics and Science Brock University, St. Catharines, ON L2S 3A1, Canada

ARTICLE INFO

Keywords:
Premature labour
Two-dimensional gel electrophoresis
Post-translational modifications
LCMS/MS
Protein species
Proteoforms

ABSTRACT

Spontaneous preterm birth (sPTB) remains a major clinical dilemma; current diagnostics and interventions have not reduced the rate of this serious healthcare burden. This study characterizes differential protein profiles and post-translational modifications (PTMs) in first trimester maternal serum using a refined top-down approach coupling two-dimensional gel electrophoresis (2DE) and mass spectrometry (MS) to directly compare subsequent term and preterm labour events and identify marked protein differences. 30 proteoforms were found to be significantly increased or decreased in the sPTB group including 9 phosphoproteins and 11 glycoproteins. Changes occurred in proteins associated with immune and defence responses. We identified protein species that are associated with several clinically relevant biological processes, including interrelated biological networks linked to regulation of the complement cascade and coagulation pathways, immune modulation, metabolic processes and cell signalling. The finding of altered proteoforms in maternal serum from pregnancies that delivered preterm suggests these as potential early biomarkers of sPTB and also possible mediators of the disorder. *Biological significance:* Identifying changes in protein profiles is critical in the study of cell biology, and disease treatment and prevention. Identifying consistent changes in the maternal serum proteome during early pregnancy, including specific protein PTMs (e.g. phosphorylation, glycosylation), is likely to provide better opportunities for prediction, intervention and prevention of preterm birth. This is the first study to examine first trimester maternal serum using a highly refined top-down proteomic analytical approach based on high resolution 2DE coupled with mass spectrometry to directly compare preterm (< 37 weeks) and preterm (≥ 37 weeks) events and identify select protein differences between these conditions. As such, the data present a promising avenue for translation of biomarker discovery to a clinical setting as well as for future investigation of underlying aetiological processes.

1. Introduction

Preterm birth (PTB), defined as delivery < 37 weeks' gestation, remains the leading cause of perinatal mortality and morbidity worldwide [1]. An estimated 15 million babies are born preterm each year and 28% of deaths in infancy are related to prematurity [1,2]. In addition to the risk of immediate neonatal complications, preterm infants have increased risks of neurodevelopment complications (such as cerebral palsy and autism) and of cognitive, cardiovascular and metabolic disorders [3–5]. Understanding and predicting PTB has been a challenge as the risk factors and pathophysiological pathways associated with PTB appear to be complex and poorly understood [6,7]. Consequently, there has been little success in predicting or preventing PTB.

Proteoforms are of fundamental importance to all biological processes [8,9]. Exploring and defining the serum proteome in pregnancies that are later affected by PTB should improve our insight into the evolution and progression of this disease. Quantification of disease-associated protein alterations can be achieved by using two-dimensional gel electrophoresis (2DE) coupled with mass spectrometry (MS) [10–12]. 2DE was specifically chosen because it is the only available proteomic technique that can simultaneously resolve hundreds-to-thousands of intact protein species in a single analytical run, while also enabling multiple parallel analyses. As the only such routine top down analytical protocol, it is thus the only available approach that enables quantitative profiling of large sets of complex mixtures of protein species; that is, as part of the routine analytical protocol, this approach

* Corresponding author at: Department of Health Sciences, Faculty of Applied Health Sciences, Brock University, St. Catharines, ON L2S 3A1, Canada.
E-mail address: jcoorssen@brocku.ca (J.R. Coorssen).

https://doi.org/10.1016/j.jprot.2018.02.002
Received 15 August 2017; Received in revised form 29 January 2018; Accepted 2 February 2018

resolves protein isoforms, splice variants, and the vast range of post-translationally modified protein species that define biological functionality. The ability to resolve protein species, including those involving specific post-translational modifications (PTMs) is likely to prove quite important; for example, changes in glycosylation have been reported in trophoblast tissue in pregnancies affected by both pre-eclampsia and PTB [13]. Identification of consistent alterations in the abundance of particular protein species may thus provide new opportunities for prediction of PTB.

A number of aetiological pathways have been implicated in PTB including infection (chorioamnionitis), uteroplacental ischaemia, haemorrhage and mechanical over-distension of the uterus [14]. Risk factors for PTB have been identified but screening on the basis of maternal history fails to predict the majority of women who present in spontaneous preterm labour or with preterm prelabour rupture of membranes [15,16]. Current investigational tools, such as ultrasound assessment of cervical length and measurement of vaginal fetal fibronectin appear to focus on a final common pathway toward PTB [17–19]. In contrast, obstetricians screen for a variety of pregnancy complications at $11–13^{+6}$ weeks' gestation and have demonstrated that early prediction enables effective therapeutic intervention and disease prevention [20,21]. In this study, we identify changes in the proteome of first trimester maternal serum collected from pregnancies that subsequently delivered preterm.

2. Materials and methods

2.1. Study design

Proteomic analyses were carried out on a cohort of maternal serum samples prospectively collected and stored during first trimester screening at $11–13^{+6}$ weeks' gestation. The samples were collected between 2011 and 2014 and serum was separated within 4 h of collection. Aliquots of residual serum, available after measurement of free βhCG and PaPP-A, were stored immediately at −80 °C. Details of pregnancy outcome were collated at the end of pregnancy allowing identification of a cohort of women who laboured spontaneously before 37 weeks' gestation (n = 10). These were matched (one to one) to a cohort of women who laboured spontaneously and delivered at term (≥37 weeks' gestation) (n = 10). Control samples were matched by maternal age, gestational age, BMI, parity, smoking status, sex of the fetus and sample storage time. Samples were excluded if preterm delivery was elective, had occurred after spontaneous rupture of membranes or involved a multiple pregnancy or a pregnancy affected by chromosomal or structural abnormality. The study was approved by the local hospital Human Ethics Committee (Protocol No: X11-0305).

2.2. 2-Dimensional gel electrophoresis

A 'top down' proteomic approach was used to resolve and identify maternal serum proteins and PTMs via 2DE and liquid chromatography mass spectrometry/mass spectrometry (LC/MS/MS). As described previously, 500 µg total serum protein was resolved on a 17 cm 3–10 non-linear immobilised pH gradient strip in the first dimension followed by second dimension resolution on 7–20% gradient gels with a combination of lithium dodecyl sulfate (LDS) and sodium dodecyl sulfate (SDS) detergent [22]. All analyses were carried out using three technical replicates to ensure reproducibility (Fig. 1).

Resolved proteins were subsequently detected in-gel using the current gold standard protocol with Colloidal Coomassie Brilliant Blue (cCCB) as a near-IR dye [23]. To enable resolution of co-migrating proteins that appear as hyper-abundant spots following the second dimension of resolution, and to facilitate detection of low abundance proteins, a post-fractionation approach was used, combining 2DE with third dimension electrophoresis (3DE) and a well-established 'deep imaging' protocol, as previously described [22,24–27]. 3DE is used to further resolve co-migrating proteins that appear as hyper-abundant spots after initial resolution by 2DE, using a gradient gel customized to provide optimal resolution within the target molecular weight range [24,25]. Deep imaging involves excising saturating spots/regions prior to imaging the gel again, thus enabling detection of lower abundance protein species [25,26]. The specific sub-proteomes associated with protein phosphorylation and glycosylation were identified using the fluorescent stains Pro-Q Diamond and Pro-Q Emerald, respectively, according to the manufacturer's protocols (Life technologies, Carlsbad, CA). After detection of phosphoproteins and glycoproteins, gels were finally stained with cCCB for total protein as previously described [23,28,29]. Stained gels were imaged using the FLA-9000 (FujiFilm/GE Health Sciences) for phosphoprotein (555/580 nm excitation/emission (ex/em)) and glycoprotein detection (510/520 nm ex/em; 500 V PMT and 100 µm pixel size). Optimal near-IR imaging of cCBB-stained gels using the FLA-9000 was carried out at 685/ > 750 ex/em with a PMT setting of 600 V and pixel resolution set to 100 µm [23,26].

2.3. Image analysis and mass spectrometry

Image analysis was carried out using the software Delta2D V4.0 (DECODON Gmbh, Greifswald, Germany) in accordance with the manufacturer's specifications. Protein patterns revealed by each stain were used to create three unique maps for each patient sample. The 'phosphorylation' and 'glycosylation' images were directly superimposed over the total-protein stain image to enable unambiguous matching of protein spots. This enabled quantitation of changes in both the PTM and amount of each resolved protein species. Images were divided into sPTB and control groups. For the purpose of this analysis, we used union setting by aligning each gel image to form a consensus pattern. This pattern was then transferred to all raw images to enable spot comparisons between the sPTB and control groups. The union setting enabled selection of only those spots that were 100% reproducible within all three replicate gels. The total number of resolved protein species for each proteome (i.e. total proteome, phosphoproteome and glycoproteome) were calculated using the three individual raw gel images for every patient (i.e. n = 3 per patient) and are reported as mean ± SEM (Table 2). To enable calculation of the observed molecular weight (MW) and pI of a given protein spot, a set of 2DE standards (Bio-Rad) was routinely resolved over the course of the study. A standard curve was created based on the average migration of the protein standards.

For inclusion in the analysis, changes in spot volume (i.e. the abundance of a resolved protein species) had to (i) differ significantly between samples from sPTB and matched control patients (*t*-test; $p < 0.05$, n = 3) and (ii) have a fold change of > 1.5 [30]. To account for patient-to-patient variability and ensure that only consistently changing protein species would be included for identification by MS, greys values for individual spot volumes were calculated across the sPTB and control sample set for all technical replicates. This was further confirmed by visual inspection of both the gels and imaging data. Protein species meeting the criteria were manually excised from the gels and identified using LC/MS/MS; peptides were isolated for LC/MS/MS analysis and data analysed as described previously [23,26]. The peptide sequences from the MS/MS spectra were identified by correlation with the peptide sequences of proteins in the SwissProt database (version 2011_06) using the Mascot Daemon search algorithm (V2.2.2) (Matrix Science, Boston, MA, USA). The PANTHER (Protein Analysis Through Evolutionary Relationships) classification system and UniProt database were employed to assess the identified proteins for biological context, involvement in various physiological pathways and association with disease pathophysiology.

3. Results

The demographic and clinical characteristics of the sample set are