

# REPORT

## The HUPO Plasma Proteome Project: A report from the Munich congress

Gilbert S. Omenn<sup>1</sup>, Young-Ki Paik<sup>2</sup> and David Speicher<sup>3</sup>

<sup>1</sup> University of Michigan, Ann Arbor, Michigan, USA

<sup>2</sup> Yonsei Proteome Research Center, Seoul, Korea

<sup>3</sup> The Wistar Institute, Philadelphia, Pennsylvania, USA

The Human Proteome Organization has several major collaborative research initiatives, including the Plasma Proteome Project. A major feature of the HUPO World Congress in Munich in August 2005 was the release of the special issue of PROTEOMICS with 28 articles from the pilot phase of the Plasma Proteome Project. An open Workshop and a presentation in the closing plenary session of the congress focused on next phases for the Plasma Proteome Project.

Received: November 9, 2005

Accepted: November 11, 2005

### Keywords:

Plasma proteome / HUPO / Proteotypic peptides / False-positive rates / Protein databases

## 1 Introduction

The 4<sup>th</sup> HUPO World Congress on Proteomics in Munich August 28 until September 1, 2005 featured progress on the several collaborative research initiatives organized under the aegis of HUPO during the past three years. Major progress was reported from the Plasma Proteome Project (PPP), supplemented by the release of the August 2005 special issue of PROTEOMICS with 28 papers on many aspects of the human plasma proteome, about half collaborative analyses and annotations and half laboratory-specific studies. Altogether 35 laboratories in 13 countries provided datasets for these analyses. Public PPP datasets can be accessed at <http://www.bioinformatics.med.umich.edu/hupo/ppp>, [www.ebi.ac.uk/pride](http://www.ebi.ac.uk/pride), and [www.peptideatlas.org/repository](http://www.peptideatlas.org/repository). Additional publications are in process. The PPP investigators and HUPO leaders are keen to have others utilize these data for additional analyses and insights.

**Correspondence:** Dr. Gilbert S. Omenn, Internal Medicine, University of Michigan, MSRB 1, 1150 W. Medical Center Dr., Ann Arbor, Michigan, 48109-0656, USA

**E-mail:** gomenn@umich.edu

**Fax:** +1-734-647-8148

**Abbreviations:** PPP, Plasma Proteome Project; SOPs, standardized operating procedures

## 2 Goals

The long-term scientific goals of the HUPO PPP are (i) a comprehensive analysis of plasma and serum proteins in people, linked with comparative analyses of other species, (ii) identification of physiological, pathological, and pharmacological sources of variation within individuals over time, leading to development and validation of biomarkers for diagnosis of diseases, and (iii) determination of the extent of variation across individuals and populations. Within HUPO, there is an agreed-upon priority for collection of plasma/serum specimens and analysis of those specimens alongside organ-specific proteomic analyses, to build a pathway to useful biomarkers for diseases of those organs.

## 3 The Munich PPP workshop

After a welcome from HUPO Initiatives Committee Chair Sam Hanash and a summary of highlights from the PPP Pilot Phase from Gil Omenn [1], there were multiple presentations and active discussions on strategies for the next phase of the PPP. Young-Ki Paik presented extensive protocols for depletion, fractionation, and mass spectrometry analyses of plasma and other fluids, including use of chromatofocussing with the PF2D instrument to capture basic

and membrane-embedded protein fractions. David Speicher evaluated the performance of MS/MS approaches in light of the well-recognized challenges of such biological specimens as plasma and serum: enormous complexity (numbers) of proteins and their many isoforms, extreme dynamic range of concentrations, and inherently incomplete sampling of peptides (and proteins) by MS/MS. He presented data from depletion of 6 or 20 most abundant proteins and pixelation of multi-dimensional protein fractions. Eugene Kapp presented an ongoing analysis of the comparative performance of several different search engines, raising many questions about the search engines themselves and the rapidly evolving protein and gene databases, including the appropriate methods of estimating and reducing false-positive rates, while not losing too many true-positive protein identifications.

Ruedi Aebersold introduced an emerging strategy to build a glycosite database of the proteins and peptides containing N- and O-glycosyl side chains. This approach is part of a broader strategy to identify unique peptide sequences for as many proteins as possible, then synthesize these peptides with heavy isotope labeling, followed by spiking of complex specimens. The result would be mass pairs for the peptides (heavy and light isotopes) that would permit specific, sensitive identification and reliable relative quantitation. Arie Admon and Joel Pounds presented a potential high-throughput approach based on accurate mass tags and highly reproducible retention times in LC-MS after extensive LC-MS/MS characterization of these variables. Brian Haab evaluated several antigen-capture methods with antibody microarray and Sam Hanash advocated multi-dimensional fractionation of intact proteins.

Richard Simpson emphasized the importance of standardized operating procedures for specimen collection, handling, storage, and thawing, and for each of the depletion, fractionation, analysis, and peptide and protein identification steps of these complex methods. We have ample evidence from specific inquiries that many investigators around the world are keen to have consensus protocols from HUPO that they can apply to their studies without becoming involved in methods development. Michael Mueller and Henning Hermjakob described the plans at the European Bioinformatics Institute to facilitate linkage of data submissions from specific organ proteomes and the plasma proteome.

A very positive general discussion with many participants was led by Matthias Mann and Ruedi Aebersold.

## 4 Next phase of human PPP

The participants considered (i) standardized operating procedures (SOPs) and issues of replicability of results, (ii) production and availability of antibodies for microarray studies and for validation studies, (iii) quantitation

of proteins identified with new methods and advanced instruments, (iv) databases and bioinformatics, and (v) clinical chemistry requirements and disease-related studies.

### 4.1 SOPs

A consensus process, enhanced by critical experiments, would support development of standardized procedures for specimens, fractionation, and analyses. Experimental demonstration of intra-specimen coefficients of variation in relation to sensitivity of detection of proteins of various concentrations is a necessity. Such information is required for clinical studies and testing. It was suggested that HUPO collaborate with clinical chemistry organizations (American Association for Clinical Chemistry, AACC; International Federation for Clinical Chemistry, IFCC). (Dan Chan organized HUPO PPP participation in the AACC meeting in Washington DC, October 24–25, 2005.)

### 4.2 Antibodies

We have an opportunity and responsibility to submit lists of PPP proteins for priority consideration in the HUPO Antibody Production Initiative led by Matthias Uhlen in Stockholm. The value of the Antibody Initiative was highlighted by the release at the Munich Congress August 29 of the Protein Atlas by Uhlen and colleagues. Immunohistochemistry patterns for some 660 proteins have been made available in the web ([www.proteinatlas.org](http://www.proteinatlas.org)) for 48 human tissues from three individuals and for 20 tumor types from (usually) 12 patients each. The prioritization needs to be linked with requests from the tissue proteome projects and emerging biofluids projects. The PPP seems to be particularly well advanced. Then disease-related plasma or serum studies can be conducted with antibody arrays.

### 4.3 Deeper mining of the plasma proteome

Several methods are attractive for this wide-ranging effort. First, the complexity of the plasma can be reduced by initial fractionation into subproteomes, such as glycoproteins (which eliminates albumin and other proteins lacking sugar groups), using either hydrazide chemistry or lectin chromatography, or both. Phosphoproteins are a subproteome of particular interest due to the importance of phosphorylation and de-phosphorylation in signaling pathways, though the reliability of measures in the circulation, let alone in stored samples from the circulation, will need to be established. The Aebersold approach of defining and preparing proteotypic peptides seems particularly promising, as do attention to low molecular weight proteins and peptides, more powerful MS instruments (such as multiple mass spectrometry, MS<sup>3</sup>), protein labeling methods (including iTRAQ), and intact protein fractionation (IPAS) methods.

#### 4.4 Informatics and databases

Much work is needed to clarify the additions, subtractions, nomenclature, and errors in existing, evolving gene and protein databases. Error estimation requires further development and comparison of search engine probability outputs with statistical methods like PeptideProphet/ProteinProphet and the PPP algorithms [2]. A further PPP refinement is introducing more stringent criteria for inclusion of protein identifications based on adjustments for protein length and for multiple comparisons testing (like the Bonferroni adjustment, with 43 700 entries in the International Protein Index at EBI). Speicher showed that “one-hit wonders” can become multi-peptide hits with a linear ion-trap LTQ, rather than the LCQ Deca XP+, mass spectrometer [3]. Mann has shown that highly precise mass determinations with MS-FTICR-MS can enhance confidence. These informatics and database questions are well suited to cross-initiative and cross-technology collaborations within HUPO.

#### 4.5 Disease-related studies

Hanash introduced the potential for an over-arching Disease Biomarker Initiative that would embrace many organ-directed proteomics studies, disease-specific biomarker development, and, of course, plasma proteome findings as the final common pathway for accessible specimens for screening and monitoring tests. The PPP could be an integrating unit within the Disease Biomarker Initiative, could continue as a stand-alone initiative, or, in fact, could function in both modes for several of the components discussed above. The

HUPO Initiatives Committee and Executive Committee are expected to make decisions about the structure and funding of such initiatives in the near future.

### 5 Concluding remarks

The “Holy Grail” for proteomics is high-throughput with high resolution and high sensitivity to permit large-scale clinical and epidemiological studies. Combinations of technologies and advances in databases and informatics will be critical to achieving the considerable promise of proteomics in medicine and public health.

*We warmly acknowledge the productive collaboration of the very many PPP laboratory investigators and analysts (see authors of the 28 PROTEOMICS papers of issue 13), our technical committees, and our sponsors and vendors, as well as the HUPO leadership.*

### 6 References

- [1] Omenn, G. S., States, D. J., Adamski, M. R., Blackwell, T. W. *et al.*, *Proteomics* 2005, 5, 3226–3245.
- [2] Adamski, M., Blackwell, T., Menon, R., Martens, L. *et al.*, *Proteomics* 2005, 5, 3246–3261.
- [3] Tang, H.-Y., Ali-Khan, N., Echan, L. A., Levenkova, N. *et al.*, *Proteomics* 2005, 5, 3329–3342.