Proteomics 2004, *4,* 1235–1240 DOI 10.1002/pmic.200300686

Report

The Human Proteome Organization Plasma Proteome Project pilot phase: Reference specimens, technology platform comparisons, and standardized data submissions and analyses

Gilbert S. Omenn

University of Michigan, Ann Arbor, MI, USA

A comprehensive, systematic characterization of cirolating proteins in health and disease will greatly facilitate development of biomarkers for prevention, diagnosis, and therapy of cancers and other diseases. The Human Proteome Organization Plasma Proteome Project pilot phase aims to (1) compare the advantages and limitations of many technology platforms; (2) contrast reference specimens of human plasma (ethylenediaminetetra acetic acid, heparin, citrate-anticoagulated) and serum, in terms of numbers of proteins identified and any interferences with various technology platforms; and (3) create a global knowledge base/data repository.

Received	9/10/03
Revised	31/10/03
Accepted	5/11/03

Keywords: Human Proteome Organization / Plasma proteome project / Serum plasma biomarkers

1 Introduction

The Human Proteome Organization (HUPO), formed in 2001, has launched several major initiatives. These are focused on the plasma proteome, the liver proteome, the brain proteome, protein standards/bioinformatics, and certain technologies, including large-scale antibody production. Overall, HUPO aims to accelerate the development of the field of proteomics and to stimulate and organize international collaborations in research and education [1].

The HUPO Plasma Proteome Project (PPP) has three major long-term scientific goals: (1) a comprehensive analysis of the protein constituents of human plasma and serum; (2) the identification of biological sources of variation within individuals over time, leading to validated biomarkers. These sources are physiological (age, sex, menstrual cycle, exercise), pathological (various diseases, special cohorts) and pharmacological (common medications); and (3) the determination of the extent of variation across populations and across individuals within populations (genetic, nutritional and other factors).

Blood samples have the virtues of being highly accessible sources of human specimens. There are extensive specimen archives (banks), some with extensive clinical annotation. The circulating proteins are a dynamic reflection of organ functions in health and disease.

2 Planning for the Plasma Proteome Project

In the initial planning meetings in Bethesda, Maryland, 29 April 2002, and in Ann Arbor, Michigan, 5–6 September 2002, highly interdisciplinary groups of experts from academe, government and industry proposed that there be a pilot phase to address the following ten scientific issues: (1) sensitivity of various techniques to deal with the huge dynamic range of concentrations of proteins and peptides in the circulation; (2) many technical aspects of specimen collection, handling, storage, and thawing, aiming for standardization; (3) methods of depleting or prefractionating the most abundant several proteins; (4) comparisons of the advantages and limitations or interferences associated with serum *vs* plasma, and alternative anticoagulation methods for plasma (EDTA, heparin, citrate); (5) enumeration and categorization of proteins visualized and identi-

Correspondence: Gilbert S. Omenn, M.D., Ph.D., Professor of Internal Medicine, Human Genetics, and Public Health, University of Michigan, A510 MSRB-1, Ann Arbor, MI 48109-0656, USA
E-mail: gomenn@umich.edu
Fax: +1-734-647-8148

Abbreviations: CAMS, Chinese Academy of Medical Sciences; **NIBSC**, National Institute for Biological Standards and Control; **PPP**, plasma proteome project

fied, with special attention to post-translational modifications and tissue of origin; (6) separation of intact proteins *vs* separation of peptides from digested proteins; (7) use of gel-based *vs* liquid phase multi-dimensional separation methods; (8) parameters for high-throughput links with mass spectrometry; (9) comparisons of MALDI *vs* direct MS (SELDI) methods; and (10) assessment and advancement of specific labeling chemistries.

The aims for the pilot phase are presented in Table 1 and Figure 1.

Table 1. Aims for pilot phase of PPP

1. Compare a broad range of technology platforms for the characterization of proteins in human plasma and serum. Assess resolution, sensitivity, time, cost, volumes of sample required and practicality with reference specimens.
2. Clarify influence of various technical variables in specimen collection, handling, and storage, especially anticoagulation and plasma *vs.* serum.
3. Determine whether the most abundant plasma proteins should be depleted, and whether antiprotease cocktails are necessary or desirable.
4. Develop a database structure and repository for HUPO PPP results.
5. Lay groundwork, through evaluation of technology platforms and specimen handling and through established international collaborations, for studies of plasma or serum biomarkers in health and disease across major ethnic groups.

One of the intriguing questions is the number of distinct proteins expected to be identifiable in plasma or serum. Based on the literature, progressively more proteins can be visualized and identified with advancing techniques or combinations of techniques. Ueno *et al.* [2] resolved some 350 spots of silver stained proteins on 2-D gels and identified 73 proteins with a combination of *N*-terminal sequencing and immunostaining after electroblotting onto nitrocellulose membranes. Anderson and Anderson [3] presented a thorough review which yielded 289 proteins from 2-D gels, of which 117 are used as clinical analytes. Building on the classic work of Putnam [4], they classified plasma proteins into proteins secreted by solid tissues (primarily liver and intestines) that act in plasma, immunoglobulins, long-distance receptor ligands (protein and peptide hormones), local receptor ligands (cytokines), temporary passengers (lysosomal proteins), tissue leakage proteins (diagnostic biomarkers), aberrant secretions (possibly cancer biomarkers) and foreign proteins (from infectious agents). Their 1991 paper [5] had resolved 727 spots and identified 376 spots as 49 different proteins. Combining prefractionation techniques reveals many more proteins. Adkins *et al.* [6] at Pacific Northwest National Laboratory used microcapillary liquid-phase chromatography, digestion to peptides, and ion-trap mass spectrometry on samples in which immunoglobulins were depleted using binding proteins (proteins A/G) and reported 490 distinct proteins. Then Pieper *et al.* [7] fractionated serum proteins by immunoaffinity chromatography to remove eight highly abundant proteins (albumin, haptoglobin, transferrin, transthyretin, alpha-1 antitrypsin, alpha-1 acid glycoprotein, hemo-

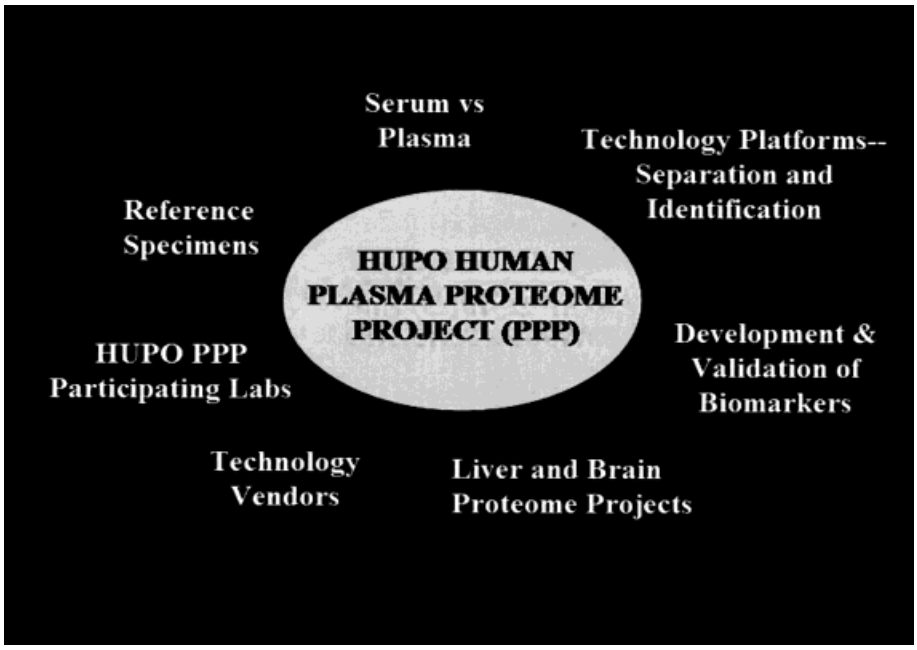

Figure 1. Scheme showing aims and linkages of the HUPO Plasma Proteome Project.

pexin, alpha-2 macroglobulin), followed by sequential anion-exchange and size exclusion chromatography, before 2-DE. They resolved about 3700 protein spots and identified 1800 by MS, which were recognized as 325 distinct proteins after sequence homology and similarity searches to eliminate redundancies [7]. The potential importance of inhibition of protease activity was demonstrated by Marshall *et al.* [8].

At the First HUPO World Congress on Proteomics in Versailles in November 2002, we presented plans for the PPP, elicited interest from many laboratories interested in participating, and established an organizational structure with active Technical Committees (Table 2).

Table 2. PPP Technical Committee structure

Reference Specimens and Specimen Handling Issues Chair Daniel Chan, Johns Hopkins
Technology Platforms and Protocols Chair Richard Simpson, Ludwig Institute, Melbourne
Database Development and Links with EBI (HUPO/PSI) Chair Henning Hermjakob, European Bioinformatics Institute, UK; Co-chair David States, University of Michigan
Population Cohorts/Specimen Banks Chair Gerard Siest, Nancy, France
Education and Training Committee Chair Peipei Ping, UCLA, Los Angeles
Executive Committee (including Partnerships) Chair Gilbert Omenn, University of Michigan

3 PPP pilot phase activities and milestones

3.1 Development of reference specimens

In order to compare the attributes of various technology platforms, it is essential to have reference specimens available for use with each platform. Extensive discussions, beginning with the September 2002 workshop, reviewed potential sources of specimens, methods of collection and division into alternate protocols, and tradeoffs in having many or few donors. The range of the latter options extended from a potential single individual to the vast American Red Cross donor pool. A special meeting on needs for standardization in proteomics convened by the US Food and Drug Administration Division of Biologics 10 January 2003, revealed another available source at the UK National Institute for Biological Standards and Control (NIBSC). Meanwhile, BD Diagnostics (Franklin Lakes, NJ, USA) stepped forward to offer extensive assistance in the preparation of reference specimens custom tailored to the needs of the HUPO PPP. The Chinese Academy of Medical Sciences (CAMS) made a similar offer. Given the lack of a standardized protease inhibitor cocktail and anecdotes about interferences arising from such proteins and chemicals, we decided to omit use of protease inhibitors in the preparation of these specimens. Instead, we proposed that a subset of laboratories, led by Daniel Chan and Alex Rai of Johns Hopkins, specifically investigate this parameter and potentially optimize the approach.

Our Specimens Committee and Executive Committee chose the following specimens:

(1) UK NIBSC lyophilized citrated plasma, previously prepared as a reference specimen for hemostasis and thrombosis studies for the International Society for Thrombosis and Hemostasis/Standards Committee. This pool was prepared from 25 donors as whole blood anticoagulated with citrate-phosphate-dextrose; double-spun, HEPES added to 0.05 M; aliquot tested for human immunodeficiency virus (HIV), hepatitis B virus (HBV), hepatitis C virus (HCV); 1 mL aliquots in 5000 ampoules frozen at −70°C. Thawed at room temperature to avoid cryoprecipitation, then freeze-dried at −35°C for 4 d and desiccated over P_2O_5 for 6 d, filled with dry N_2. Tested for longterm stability. We supply 1 mL ampoules.

(2) BD Diagnostics sets of four reference specimens for each of three ethnic groups: Caucasian-American, African-American and Asian-American. Each pool consisted of one unit of blood each from one male and one postmenopausal female healthy, fasting donor, collected in a standard donor set-up after informed consent, and immediately pooled, then divided into four equal volumes in bags with appropriate concentrations of K-EDTA, lithium heparin, or sodium citrate for plasma and without clot activator for serum. This procedure required 2 h at room temperature. Each pool was then aliquoted into numerous 250 μL portions in vials which were then frozen and stored at −70°C. Aliquots were tested for HIV, HBV and HCV. We supply sets of 4×250 μL aliquots for each of the four plasma/serum specimens. These vials plus the NIBSC ampoules were shipped frozen on dry ice *via* courier in early May 2003.

(3) The Chinese Academy of Medical Sciences sets of four reference specimens. This pool was prepared, after review by the CAMS Ethics Committee and informed consent by donors, according to the BD protocol in #2, including tests for viral infections. We supply sets of 4×250 μL aliquots for each of the four plasma/serum specimens. Following various approvals and time for controlling the severe acute respiratory syndrome (SARS) outbreak, these specimens were shipped in September 2003.

3.2 Commitment of participating laboratories

A standard HUPO PPP questionnaire was sent to all established proteomics laboratories whose investigators had expressed interest in participating, either at the workshops and HUPO World Congress or after learning about the Plasma Proteome Project through colleagues, the HUPO website (www.hupo.org), or press coverage. At the present time, 47 laboratories in 14 countries have committed to participate in the PPP. Of these laboratories, 28 are in the United States (including 17 academic, 6 US federal and 5 US corporate labs) and 19 are in other countries, including 7 in Europe, 1 in Israel, 9 in Asia and 2 in Australia. Of these labs, 41 requested the UK NIBSC specimens, 43 the BD Caucasian-American specimens, 18 the BD African-American and Asian-American specimens, and 18 the CAMS specimens. In addition, 27 volunteered for the proposed multiparameter specimen handling protocol, using reference specimens and/or their own specimens, and 30 for testing depletion or prefractionation protocols. With regard to different kinds of technology platforms, 31 indicated that they would run 2-D gels, 29 liquid chromatography separations, 25 protein digestion first, 30 MALDI or LC-MS, MS/MS, 15 direct MS, 18 protein labeling and 9 other platforms. Combinations of technologies are certain to be required in order to move down the dynamic range of concentrations, spanning about nine orders of magnitude from albumin (40 mg/mL) to PSA or cytokines (pg/mL).

3.3 Guidance from Technical Committees and the Bethesda Workshop

There has been extensive communication among the chairs and some members of the various committees (see Table 2). In May 2003, in preparation for the July 16–17 Bethesda Workshop, all participating laboratories were sent draft formats and guidance for submission of information about experimental protocols and results from the various studies of specimens. These formats were prepared by a PPP working group led by Richard Simpson of Ludwig Institute, Melbourne. Testing the limits of the methods, we proposed that results be expressed in accordance with the output of each type of technology platform, with emphasis on confidence of identification of proteins. Further, we proposed that investigators estimate the relative abundance of the proteins from spot density or absorbance peak or other methods (see Section 3.5).

The formats were reviewed in detail by investigators from 35 of the laboratories and members of the technical committees and sponsors at the Bethesda Workshop. Investigators were given the option of submitting data in Excel to the Bioinformatics Database Core, where the results would be entered; of inviting technical assistance from the Core, in order to deploy XML for submission directly into the data repository; or to use XML themselves, with which some laboratories were already familiar. The Protein Experimental Data Reporting tool (PEDRo) is being prepared for linking with the XML submission. Labs were requested to identify the protein database and search engine they utilized; following discussions in the Bioinformatics Committee and the full meeting, investigators were strongly encouraged to use the International Protein Index (IPI) search engine.

Participating laboratories and technology vendors presented updates on several platforms: Agilent – immunoaffinity column to remove the six most abundant proteins: albumin, immunoglobulin (Ig) G, IgA, haptoglobin, transferrin, and alpha-1 anti-trypsin; GenWay – chicken IgY antihuman antibodies for depletion or detection of specific proteins; BD Diagnostics – development of new collection tubes with protease inhibitors; Gradipore – fractionation cartridges; Invitrogen – zoom electrophoresis/microscale solution IEF/MALDI; and Tecan – free flow electrophoresis. Early results with HUPO reference specimens were presented by Nakamura (Yamaguchi University, Japan) with 2-D gels and by Petricoin (FDA, USA) with SELDI direct mass spectrometry.

3.4 Database development

The HUPO PPP has based its primary data repository at the European Bioinformatics Institute (EBI), which has the lead for Protein Standards (PSI) and Bioinformatics for all HUPO initiatives under the direction of Henning Hermjakob. The aim is to have a consolidated inventory of plasma/serum proteins, linked with organ-derived proteomes of other HUPO initiatives and with SWISS-PROT and other databases maintained at EBI. The specimen tracking, descriptions of experimental protocols, and related study-specific database are at the University of Michigan under the direction of David States (see Section 3.3).

Following the Bethesda Workshop, guidance was sent to all laboratories reflecting the group decisions. Most laboratories have registered at the intra-PPP websites, and many have submitted protocol information. Several laboratories submitted initial or extensive data on the reference specimens in preparation for a HUPO PPP Working Meeting immediately following the HUPO World Congress in Montreal in mid-October 2003. The Bioinformatics Core at the University of Michigan is responsible for cross-checking the protein identifications and interacting with each laboratory to refine the submissions.

An EBI-led HUPO group has already published a proposed standard for protein-protein interactions (see http://psidev.sourceforge.net/meetings/2003-01/2003-01-report.html).

3.5 Calibration of reference specimens

In order to permit assessment of the limits of detection by various methods and combinations of methods, our Specimen Committee developed a plan for quantitative immunoassays of selected proteins. These assays of aliquots of all the reference specimens are being performed under the direction of Frank Vitzthum and Harald Ackermann at Dade Behring (Marburg, Germany), Stanley Hefta at Bristol Myers Squibb (Princeton, NJ, USA), and Dan Chan at Johns Hopkins (Baltimore, MD, USA). The Dade Behring results, for example, cover 37 proteins over a range of at least seven orders of magnitude (Fig. 2). The identity of the proteins assayed will be retained centrally until all three testing labs have completed their work and their cross-comparisons and until the participating laboratories have submitted their data, protein IDs, and abundance estimates. These assays also permit direct comparisons across the sets of specimens (EDTA, heparin, citrated plasma and serum) and donor pools.

It was agreed that enumeration of proteins should be tied to the gene accession number, hence a minimal estimate of the number of biologically important proteins. A larger estimate of proteins visualized and identified would include specific modified forms of each protein. Of course, care must be taken to avoid as much as possible multiple modifications *ex vivo.* More work will be done collaboratively with EBI/HUPO-PSI to categorize protein isoforms systematically, relying less on narratives. One example is shown in Table 3.

Table 3. An illustrative classification of protein isoforms

1.0 Primary amino acid backbone sequence of gene-coded protein (including splice variants and SNPs)
1.1 Modification of backbone: cleavage (activation, inactivation)
1.2 Modification of side-chains
1.2.1 Phosphorylation (sites) and dephosphorylation
1.2.2 Glycosylation (sites)
1.2.3 Other classes of post-translation modifications
1.3 Aggregation (dimers, *etc.*)
1.4 Interactions: protein-protein, -nucleic acid, -others

3.6 Administration and dissemination of findings

The HUPO PPP has its administrative hub at the University of Michigan. Funding has been developed both from corporate sponsors and from a trans-National Institutes of Health (NIH) consortium of several NIH institutes. These funds will support extensive bioinformatics services and the database, distribution of reference specimens, workshops and associated travel, administrative services, and some support for collaborative experimental and data mining protocols. Liaison with the other HUPO initiatives is assured. Intellectual property plans have been clarified: the PPP is committed to public

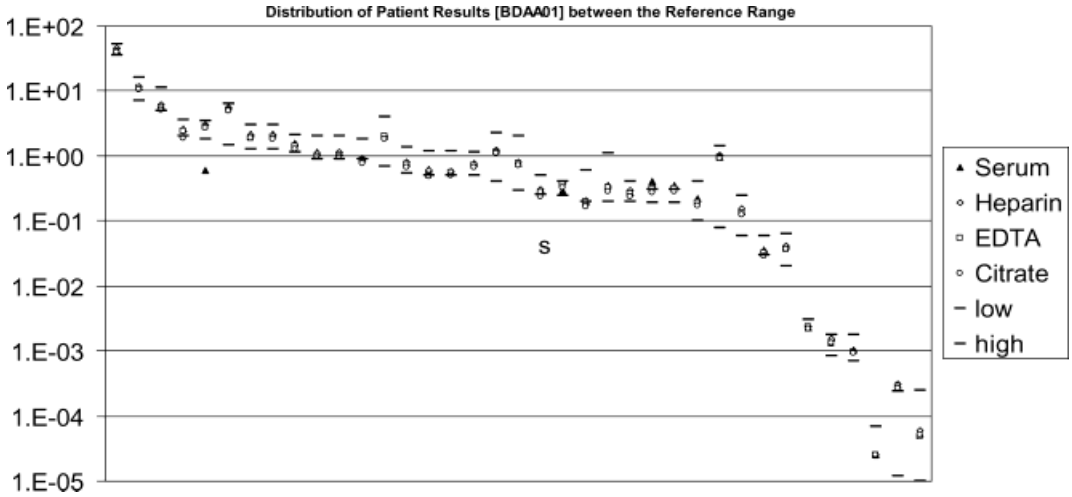

Figure 2. Quantitative determination of 37 coded proteins by immunoassay from set of four specimens from same donor pool (EDTA-, heparin-, and citrate-anticoagulated plasma plus serum), with reference ranges

domain release of its results, with ample opportunities for academic, government, or corporate laboratories to undertake proprietary development of biomarkers and targets/agents with the methodologies evaluated in the pilot phase of the PPP. We are planning an intensive workshop, perhaps modeled on the Drosophila Genome Jamboree, for cross-technology platform, cross-reference specimen, cross-laboratory analyses of the extensive findings of this project.

During the completion of the pilot phase during 2003–2004, initial planning will be undertaken for large-scale population cohort studies in collaboration with investigators responsible for many such existing and planned disease-oriented cohorts. A special issue of Proteomics dedicated to findings from the HUPO PPP is planned for late 2004, and other publications and electronic media have invited papers from this project.

4 References

[1] Hanash, S., Celis, J. E., *Mol. Cell Proteomics* 2002, *1*, 413–414.

[2] Ueno, I., Sakai, T., Yamaoka, M., Yoshida, R., Tsugita, A., *Electrophoresis* 2000, *21*, 1832–1845.

[3] Anderson, N. L., Anderson, N. G., *Mol. Cell. Proteomics* 2002, *1*, 845–867.

[4] Putnam, F. W.,*The Plasma Proteins: Structure, Function, and Genetic Control*, Academic Press New York, 1975–1987.

[5] Anderson, L., Anderson, N. G., *Electrophoresis* 1991, *12*, 883–906.

[6] Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 947–955.

[7] Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S. *et al.*, *Proteomics* 2003, *3*, 1345–1364.

[8] Marshall, J., Kupchak, P., Zhu, W., Yantha, J. *et al.*, *J. Proteome Res.* 2003, *2*, 361–372.

