

HUPO Liver Proteome Project

Workshop

*Marriott Hotel Pooks Hill
Bethesda, Maryland
July 17-18, 2003*

Recommendations for action plan

This document summarizes the recommendations that resulted from the workshop.



The strongest and most compelling argument for a proteome initiative that focuses on the liver is the essential and multi-function role of the liver in human health and disease. Liver disease afflicts more than 10% of the world population. Liver pathogenesis remains however largely undefined and most liver diseases remain poorly diagnosed, staged and treated. In addition, many proteins in plasma are synthesized in the liver and change with disease in the liver, as well as systemic responses to inflammation and disease processes in other organs. Finally, protein signatures in preclinical study models and in human liver samples could be used as an effective way to understand pharmacology and toxicology. Understanding the liver proteome will accelerate the development of diagnostics and therapeutics towards the diseases of this organ and also facilitate drug discovery. A proposal for an action plan for the initial phase of an international liver proteomics initiative has been sketched out and funding to that effect has been committed by several governments and agencies (\$10M from Chinese Government, \$2.5M from Genome Canada, \$2M from INSERM and ARC in France). Other participating countries include Japan, Korea, Germany, Australia and the United-Kingdom. The importance of liver proteomics is therefore reflected by the commitment of countries worldwide to support, in a major way, such an initiative.

It was clear from the NIH-HUPO workshop that several NIH institutes have substantial interest in a liver proteome initiative. In addition, some US investigators have ongoing liver proteome programs and are committed to actively participate in the HLPP project. It is therefore timely to develop a plan for US participation in this international effort and for the US to join our Chinese, Canadian and French colleagues as a major participant and in leading the effort to finalize the action plan.

The number of proteins in an organ like the liver is very large and their characterization is complicated due to post-translational modifications and splicing variants of the gene products. Localization of proteins in the cell and protein-protein interactions are also of critical importance. Mobilizing the resources available in different countries will facilitate the task of executing a major liver proteome project. As highlighted during the workshop, there are in the US many outstanding groups with expertise in liver science and a unique liver tissue and cell distribution system, with availability of controls and standards. The HLPP would benefit from this expertise and synergy via the involvement of the NIH would help ensure the success of the project.

A critical success factor for this initiative will be the ability to assure that protocols, data capture and analysis are standardized. An immediate role of the US would be to participate in the establishment of protocol standardization, assuring that rigorous and standardized protocols are created and followed. Selection of standards and of a site for database management would assure that the tremendous amount of data and resources generated could be put into use by the research community for a multitude of liver-related projects.

The immediate goal for HLPP is to demonstrate what can be done in a Pilot Project. The key to the Pilot phase is to lay a foundation for future studies. The proteome will be studied for the next 10-20 years or more. The hope is to stimulate that research now.

Scope of U.S. participation to HLPP

The major recommendation is to focus on the biology. The field of proteomics is developing rapidly and therefore techniques are at the present time not the most important aspect. The technology is a tool to get us where we want to go as biologists. Defining the biology and the questions researchers want to ask are most important. Therefore the scientific objectives need to be defined first.

The scientific objectives proposed by the co-directors for the pilot phase are encompassed in three sub-projects that focus on:

- Normal Human Liver
- Hepatocellular Carcinoma (HCC)
- Normal Rat Liver

Common activities between these sub-projects include:

- Collection and banking of human liver tissue specimen
- Proteome expression profile of human liver
- Elucidation of protein modification profiles of liver
- Protein interaction maps of human liver
- Intracellular localization of human liver proteins
- Antibody bank for human liver proteins
- ORF bank of human liver proteins

Normal Human Liver

- China will contribute in a major way in defining the proteome of human normal liver, with 19 participating laboratories. It may be difficult to convince leaders at NIH to support a similar effort in the US if it can't be linked to clinical issues. However, it is important to consider that once the scientific basis of the liver proteome is known then that knowledge can be expanded to study liver diseases. Mapping and identifying the proteins in the human liver may be the key to many diseases. One approach would be to define the range of the proteome in the normal liver and do a subtractive approach between normal and diseased liver, identifying added or subtracted or altered protein species. In addition, most of the programs on genomics and proteomics of liver diseases or on hepatotoxicity utilize normal liver tissue or isolated liver cells as controls. Therefore, the study of normal liver proteome is an obligatory step to build basis for current and future research.

- Recommendations are to analyze both liver tissues and isolated cells. For cellular protein profiles, two approaches were proposed: 1/ Identify proteins in isolated cells. Need to develop tissue culture models that minimize *in vitro* related changes and 2/ Identify proteins in whole liver tissue followed by cellular localization studies using histological or other techniques. This may be preferable to identifying proteins in isolated hepatic cells as the isolation technology induces changes in protein expression.

Diseased Liver

- HCV-related liver disease including HCC is an important public health issue and a priority in many countries including the US. The study of HCV-related HCC should be extended to all stages of liver disease progression following from chronic hepatitis to liver fibrosis to cirrhosis to HCC. France and Japan in particular had indicated their priorities in including this project in HLPP. This project will clearly be also a priority and part of the contribution of US investigators to HLPP. Because of the diverse geographical distribution of HCV genotypes and co-factors such as race and alcohol consumption, it would be of great interest to compare the data generated by different countries. In addition, through HLPP, large cohort studies will be identified in different countries, in which some promising proteomic biomarkers for liver cancer early detection could be tested.
- The study of liver diseases should also include HCC, liver fibrosis and cirrhosis associated with other etiologies (other viral hepatitis, alcohol, environmental agents, etc..)
- Because of the US leading role in HUPO and the commitment of NIH to the plasma proteome project, strength of the US participation to HLPP will be the identification and subsequent validation of protein biomarkers for early diagnosis of risk, onset, and progression of many important liver diseases.

Animal Models

It is necessary to undertake the study of both animal and human simultaneously and to retain and capture information in a way to serve future generations. The choice of the rat versus the mouse is debatable as the most appropriate animal model. Rat models have been used for years without translation to the human. However, the rat proteome model could help develop technology and establish a basis from which future research can build. For any methods development, using rat liver as a surrogate for the pilot phase would indeed be a great choice. In addition, the pharmaceutical industry relies heavily on the rat. Good mouse models already exist. Therefore, there is probably a need to profile the liver proteome in both the rat and mouse. Because, there are currently many animal models, there is a need to collaborate in animal models.

Others

The initiative should not exclude projects focusing on individual or groups of proteins. In certain disease states, one group of proteins may be important. The challenge is to pursue them on a case-by-case basis.

NIH support of investigator research is hypothesis driven and linked to clinical issues. In this context, consensus scientific objectives may not be necessary between the different countries as well as between US investigators. There is a need to divide and conquer the task. The projects listed above can all go in parallel. Investigators could develop tools and models and propose technologies to apply to a liver problem during the next 3 years. The real linchpin is to coordinate the studies and make them accessible in a database. If that data is shared in 3 years then HLPP will have 1/ the possibility of identifying data of immediate biological interest and 2/ useful information about what works in studying the liver. Projects in the US would largely benefit from the efforts in other countries and vice versa.

Platform diversity and protein standards in bridging the proteomic technologies in HLPP

(separation, identification, characterization)

The diverse nature of proteins leads to a requirement for platform diversity and for a full array of existing and emerging technologies included in the project. The choice of platform depends upon the project and the level of proteins of interest and effort to one prescribed technology or method of research should not be reduced. The major proteomic technology platforms that are currently used are 1) 2D PAGE-mass spectrometry (MS) with protein identification by peptide fingerprinting with MALDI-ToF MS or sequencing tagging and post-translational modification determination by LC-MS/MS; 2) multi-dimensional protein identification technology LC-MS/MS or 'MuDPIT', 3) stable isotope labeling of proteins such as isotope affinity tag technology or 'ICAT'; and 4) protein retentate-MS such as surface-enhanced laser desorption ionization or 'SELDI'. These are the primary technologies that would be used in HLPP.

The many attributes of proteins have spawned many technologies to measure these attributes in a comparable manner. Proteomic platforms in the HLPP would be expected to be 1) global, 2) discovery-enabled, 3) reproducible and 4) high throughput. Once data is collected, it is difficult to compare across platforms and experiments without the proper physical and data standards. The formulation of a set of Proteomics Protein Reference Standards, a set of about ten proteins, that could a) be representative of a desirable cross-section of protein properties of human liver proteins or proteins in general, b) be used for instrument calibration and c) normalization of data across proteomic technology platforms, is proposed. This set of proteins would be used as a proof-of-principle to determine whether the limitations of the various platforms and methodologies could be observed. This reference material can then be enhanced over time by substitution or addition of other proteins and may eventually develop into a Standard Reference Material. Workshop attendees could be queried for specific proteins or properties considered important for such a group of protein standards and then a means could be found to create such proteins by recombinant expression from *E. coli* for purification, characterization, aliquoting and eventual distribution to participating laboratories. NIST can make standards or work with the proteomics community on organizing standards. The objective nature and long history of analytical standards production of the NIST agency is viewed as a promising source for creating and storing proteomics protein standards.

Cell and Tissue Resources

(well characterized samples)

Major focus should be on the biological tissues researchers are going to study because that's where most of the variation will be. It is to stress that the pathology and the quality control of the liver tissues are of fundamental importance. There is fantastic material available to study liver proteomics. Whether the existing material can cover the scope of this HLPP project is however not yet known.

In the US, there is already, in place, a mechanism to procure liver tissue from normal livers as well as diseased cases. The Liver Tissue Procurement and Distribution System. (LTPADS) is extremely effective in this area and can furnish the needed tissue for proteome analysis. The LTPADS program is an NIH funded tissue procurement and banking system for liver tissue. Investigators at 3 medical centers procure tissue from normal and diseased liver tissue. Diseased tissue is usually procured from the explanted tissue at the time of whole organ transplantation. Diseased tissue is snap frozen in liquid nitrogen in a manner which retains RNA and DNA integrity. With the diseased tissue, the recipient is provided with a summary of the age, sex and ethnicity of the donor and the type of liver disease which caused the pathology. A short pathology report is also available when needed. Snap-frozen normal liver tissue and isolated human hepatocytes are also provided. Normal liver is defined as an organ procured with the intent for transplant from an organ donor, but which is not used for the transplant. The most common reason for not transplanting the organ is steatosis (fat) in excess of 30%. There is a complete donor history, with clinical chemistries and serological testing for HBV, HCV, and HIV. In some cases, hepatocytes are isolated from the normal livers and the cells are available for research purposes in suspension, on wet ice, or plated on flasks or 6,12, or 24 well plates. LTPADS has an excellent infrastructure and experience in addition to a good track record. Serum/plasma is not available but there is current discussion to store serum/plasma as well.

In addition to LTPADS, nationally funded cell and tissue repository will also be available in France and China. France has recently established a multi-center collection of liver tissues and proposed protocols for standardization of approaches. China is initiating the development of a Chinese Liver Tissue Bank, involving major liver transplantation centers in China. To this regard, they wish to follow protocols suggested by experimented "tissue bankers". Therefore, there is a need to agree upon routes of tissue collection, stage of disease, definition of the stage of disease to study. A committee to determine the standardized protocols to be used for HLPP across the participating centers will include:

- From US: Stephen Strom (coordinator)
- From France: Chantal Housset (coordinator)
- From China: Xiaohang Zhao (coordinator)

Specific questions will include:

* blood contamination

How should researchers factor into proteomic profiling of biopsies the amount of modifications in liver proteins related to contamination in the biopsy?

- * accompanying serum/plasma samples
- * are there sufficient samples (samples from large number of patients representing spectrum of demographics and stage of liver disease) on all the diseases of interest?
- * Definition of normal
- * Criteria for diagnosing and staging liver disease
- * There are a large number of protein changes in response to any specific signals and therefore there is a need to control for factors that may affect expression of proteins in normal liver
 - Cell type – purity of cell population
 - Cell isolation technique – may affect cell cycle, physiologic state
 - Proliferative status of liver
 - Location of cell type in hepatic acinus
- * Collection, handling and storage
 - Time elapsed from tissue removal to processing and storage
 - Degree of blood and bile contamination
 - Chemicals used in cell isolation
 - reservation solution
 - Storage condition
- * Verification of cell purity and viability
 - Minimal standard required for different types of studies
 - Methods used to verify cell purity and viability
- * Standard format for accompanying patient's record and clinical information
 - Demographics
 - Physiologic state
 - Diagnosis and staging of liver disease
 - Comorbid illness
 - Medications
- * Ethical standards for consent, documentation and oversight
 - Consent
 - Source of materials
 - Collection, processing and storage
 - Verification of cell purity and viability
- * Standard screens need to be applied at the site of the tissue distribution since the liver is often contaminated with hepatitis viruses, etc. Agreements limiting liability also need to be established.

Data Capture, Analysis and Sharing

The power of this initiative will stem from the integrative analysis of experimental data from a large number of laboratories. Participating organizations will apply their expertise and key technologies to common sample sets, and will contribute their results to a shared, central comparative analysis effort.

To enable the comparative analysis of experimental results from a broad variety of participating organizations, across different and fast developing experimental approaches, procedures and infrastructure for sharing and analyzing data need to be a key aspect of the overall project.

A central results database needs to be developed to provide easy and consistent access to the results provided by participating organizations, and to allow quick linking to existing resources. However, not all details can and should be stored centrally, LIMS-specific and raw data should only be stored locally, but should be linked from the results database. For the data communication between project partners, the central results database, and the scientific community at large, data representation standards need to be defined. The HUPO Proteomics Standards Initiative (PSI) is currently developing such standards in a collaborative effort with partners from both the public and private sector, to ensure consistency of data management approaches across all HUPO initiatives (<http://psidev.sf.net>). Formats have already been developed for the representation of protein interaction and mass spectrometry data, a general framework for the representation of proteomics data is under development in collaboration with the PEDRO project (***). PSI will also collaborate with the proteomics group of the National Institute for Standards and Technologies (NIST), in particular in the domain of electrophoresis data (<http://bioinfo.nist.gov:8080/examples/servlets/index.html>).

In a large scale project like HLPP it is essential to standardize not only data formats, but also contents. This comprises the use of shared data resources, for example the Gene Ontology (<http://www.geneontology.org>) for annotation, the use of a common protein sequence database for protein identifications, e.g. the International Protein Index (<http://www.ebi.ac.uk/IPI>), as well as the development and use of shared controlled vocabularies for the classification of experimental technologies and results. Without the use of common, standardized resources, the comparative analysis of results, e.g. protein identifications, becomes virtually impossible.

Proteomics projects currently often lack a rigorous statistical validation of results, and statisticians often only become involved in a very late phase of the project, usually once the raw results have been generated. To ensure statistically viable and scientifically relevant results, statisticians need to be involved already in the early, planning stages of HLPP, and need to be consulted throughout the project.

To allow comparative analysis of results, all project partners need to have early access to the results of other consortium partners, To allow scientific scrutiny of HLPP approaches and results, the scientific community needs to be able to access and verify intermediate results of the project as quickly as possible. To avoid project delays due to data access discussions, clear data release policies need to be defined and to be reconciled with intellectual property issues prior to the start of the project. The Bermuda principles used in the Human Genome Project might serve as an example for the successful definition of data release policies.

Summary of recommendations

- Integrate results into a centralized database, with links back to local LIMS systems for access to very detailed and raw data
- Use standardized data formats and contribute to their development
- Use standardized data resources as a basis for e.g. protein identifications
- Ensure not only format, but also contents standardization through the use of controlled vocabularies and ontologies
- Assign sufficient resources for the complex task of central data analysis
- Involve data analysis teams and in particular statisticians from the beginning of the project, not only in the interpretation of results
- Develop a well-defined data release policy pertaining both to project partners and the scientific community

Building bridges with the HUPO Plasma Proteome Project (PPP)

The scientific objectives of this effort to build bridges between the plasma and the liver proteome projects are:

- To determine which proteins in plasma are synthesized in the liver
- To determine the dynamic of the changes affecting these proteins in diseases of the liver
- To determine the dynamic of the changes affecting these proteins in diseases in other organs
- To determine the dynamic of the changes affecting these proteins as indicators of hepatotoxicity

Because of the US leading role and the commitment of NIH to the plasma proteome project, strength of the US participation to HLPP will be the identification and subsequent validation of protein biomarkers for early diagnosis of risk, onset, and progression of many important liver diseases.

CONNECTIONS OF PPP and HLPP

- Collection of serum/plasma and liver specimens from same patient or experimental animal. Plasma or serum samples should be collected in parallel to liver tissues, in order to screen for potential circulating early onset biomarkers of liver diseases. One of the immediate objectives of the plasma proteome project is to clarify the influence of various technical variables in specimen collection, handling, and storage; especially anti-coagulation and plasma versus serum. Preliminary data will be presented at the HUPO congress in October and therefore it is anticipated that recommendations for collection of serum/plasma will be available to the centers involved in liver tissue collection for HLPP.
- Evaluation of technology platforms, with several labs participating in both projects
- Joint planning and analyses of liver-disease-related population cohorts to develop biomarkers for liver diseases in serum/plasma
- Production of high specificity, high affinity probes such as antibodies to detect and quantitate these potential markers.
- Common standard tools for data capture and analysis

NIH-funding support for HLPP

Liver disease is an important cause of morbidity and mortality in the United States, affecting persons of all ages, but most frequently individuals in the productive years of life, between the ages of 40 and 60 years. A human proteome initiative would allow the rapid translation of findings from basic research to practical means of prevention, control and cure of liver diseases. To date approximately 15 M USD has been secured for the initial phase of HLPP (\$10M from Chinese Government, \$2.5M from Genome Canada, \$2M from INSERM and ARC). It is therefore timely to consider a US participation in this effort and to provide input in the development of an action plan for the HLPP. To assure the success of the project, NIH should contribute to this effort, using various mechanisms and sources of support for research and for enhancing cooperation and coordination. HLPP is a large-scale biology project that is suited for a trans-NIH initiative. NIDDK is developing an *Action Plan for Research in Liver Disease* to address the burden of liver diseases in the United States. The Action Plan will address the broad range of liver disease research, and will help guide NIH initiatives in liver disease research. The HLPP initiative goes beyond a single Institute's interest and is also part of the priorities for: NCI (increasing incidence of HCC); NIAID (HCV is the most common chronic blood-borne infection in the US); NIAAA (alcohol-related liver disease and interactions between alcohol and HCV in cirrhosis and HCC); NIDA (hepatitis C, interactions between HIV and HCV, HIV and HBV); NIEHS (hepatotoxicity); as well as NIST (data interoperability in large scale projects) and the FDA (drug toxicity). Therefore, a recommendation for providing funding to accomplish the objectives of the HLPP would be through a trans-Institute RFA. A trans-Institute commitment would make available sufficient funds for the RFA, avoid duplication of efforts, permit cross-fertilization, assemble relevant expertise. This mechanism will enable researchers working in liver proteomics to submit at the same time for review by an assembled panel with expertise in proteomics, proposals to accomplish objectives as defined in the RFA. NIH should also ensure that the resulting research effort is a united one and not a loose collection of individual R01 grants without integration. A cooperative agreement mechanism with oversight by NIH and regular meetings to foster exchange is necessary. Funding for a centralized resource center (tissue collection, proteomic platforms, data analysis) may assure that the start of the pilot phase is not delayed.

The NIDDK should bring this initiative to the attention of relevant trans-NIH working group for coordination with Dr. Zerhouni's roadmap. The RFA concept should be presented to NIDDK's Council as a high priority initiative for 2004 funding. The workshop participants recommended the formulation of a US HLPP working group to facilitate communication among investigators in the US as well as with international colleagues. The propose composition of this group consists of:

- *Coordination with HLPP*: Laura Beretta (coordinator), Jake Liang, George Michalopoulos, Allan Wolkoff
- *Cell and tissue resources*: Stephen Strom (coordinator), Harvey Sharp, Anna Lok
- *Proteomic platforms*: Alex Merrick (coordinator), Frank Witzmann, Greg Vasquez
- *Data analysis and sharing*: Ziding Feng, Veerasamy Ravichandran

- *Coordination with PPP:* Gil Omenn (coordinator), Sam Hanash, Sudhir Srivastava
- *Consultants:* Jose Serrano, Francis Chisari