

**HUPO Workshop**

**Defining the proteome agenda**

**Oct 7, Leesburg, Virginia**

**Meeting Report (First published in the Dec 2001 issue  
of Proteomics)**

On Oct 7, 2001, a planning meeting was convened by the newly founded Human Proteome

Organization (HUPO). The meeting was sponsored by the National Cancer Institute

and the Food and Drug Administration, and took place in Leesburg, Virginia.

The purpose was to review the state of the art in proteomics and to consider

various opportunities and initiatives to further our knowledge of the human

proteome. The participants strived to develop a vision, and to consider the

challenges that lie ahead. A summary of the discussions and recommendations

resulting from the meeting are presented.

It was acknowledged that proteomics is in an exponential growth phase. This

is due to a great extent to the fact that the major undertaking of sequencing

the human and other important genomes has largely been accomplished, which has

opened the door for proteomics by providing a sequence-based framework for mining

the human proteome. It was evident that proteomics has attracted a substantial

following, with an influx of new academic, pharmaceutical and biotechnology

investigators.

### **Human proteome project**

The goal of proteomics was ambitiously defined as the identification of all

the proteins encoded in the human genome followed by the determination of a)

their range of expression across the different cell types that constitute the

various human tissues, b) their sub-cellular localization, c) their post-translational

modifications, d) their interactions with other proteins, and e) their structure-function

relationships. Additionally, there is a need to develop an understanding of

global and temporal protein expression patterns in different developmental,

physiologic and pathologic states. The implications of such an undertaking for

biomedicine are tantalizing, from accelerating drug discovery, as pointed out

at the meeting by the Food and Drug Administration commissioner Bernard Schwetz,

to the early diagnosis and therapy of cancer and other diseases as pointed out

by Carl Barrett, scientific director, National Cancer Institute.

It is obvious that whereas the genome sequence is fairly uni-dimensional and

finite, the proteome is multi-dimensional, with quasi infinite dimensions stemming

from the large number of cell lineages and subtypes and additionally, from the

various body fluids, each with its own proteome. Moreover the proteome is dynamic

and constantly changing in response to various environmental factors and other

signals, thus giving rise to near-infinite dimensions of states. It was pointed

out that the absence of a simple focal goal analogous to the sequencing of the

human genome, and the complexity and diversity of both the technologies and

the resultant proteome data, make the initiation of a single large international

human proteome project much more difficult compared to the human genome project.

Additionally, the need for sophisticated tools for proteome data integration

is much greater than for genomic data. Consequently many participants felt that

it was unrealistic to initiate a human proteome project that would exhaustively

tackle all aspects of the human proteome. Instead, it would be more appropriate

to begin a planning and discussion phase to identify appropriate short-term

milestones and measures of success, as we prioritize the stages of the unfolding

human proteome project. It was noted that the decision to launch the human genome

project was made after no less than two years of discussion, which was followed

by three years of pilot effort.

Francis Collins emphasized that technology should be in place before scale

up occurs. He emphasized that technology development and production lines should

be separate, with production lines relying on facilities already in place. Several

major pharmaceutical companies were represented at the meeting and it was felt

that private-public partnerships would be highly desirable in specific areas

where compelling scientific opportunities exist at an affordable cost, and for

specific projects that have a finite duration. Lee Hood remarked that an ambitious

goal should be set for a proteome project with an emphasis on integration and

systems biology.

It was clear that proteomics employs a mix of high tech approaches plus old-fashioned

protein chemistry and biochemistry. As a result, there is a need and an opportunity

to engage scientists in disciplines and in countries that have not been heavy

participants in genome initiatives. It was noted that several countries have

begun to develop proteomics agendas. For example, a center for proteome research

has been set up in Moscow, under the direction of professor Alexander Archakov,

that takes advantage of expertise already available. The center is developing

programs in research and education. A French delegation to the meeting led by

Jean Rossier, conveyed the enthusiasm of INSERM and its director, Christian

Brechot, for a substantial French participation in various components of a proteome

project, perhaps with a focus on proteomics at the single cell level.  
Ian Humphery-Smith

reviewed other proteomics initiatives in Europe. Young-Ki Paik, head  
of the

Korean Proteome Organization which includes 700 members, pointed out  
a substantial

interest in his country in contributing to a proteomics effort and to  
related

technological innovations.

Ed Liu, director of the Genomics Institute in Singapore pointed out  
that in

addition to efforts in individual countries, proteome initiatives  
with a multi-national

backing from governments, industry and philanthropic organizations  
could contribute

substantially to rallying and assembling the international community  
into an

effective and powerful alliance to tackle the human proteome. The  
success of

international proteomic efforts would be enhanced by nucleating  
working "hubs"

within the larger proteomics community. These points of nucleation  
can be based

on disciplines (structural biology or protein interactions),  
technology platforms

(databases, bioinformatics, or bioanalytical chemistry), or  
geographical regions

(Asia, Europe, Americas).

Three breakout sessions focused on opportunities in expression proteomics,

functional proteomics and the prospects of developing a proteome knowledge base.

### **Expression proteomics**

Expression proteomics involves the identification and quantitative analysis

of all proteins encoded in the human genome and assessment of their cellular

localization and their post-translational modifications. The sensitivity and

throughput of current technologies for expression proteomics were important

discussion items. Jean Rossier noted that advances in mass spectrometry have

resulted in substantial sensitivity, such that a detailed protein profile could

be obtained starting with no more than a thousand cells that can be isolated

from a tissue using laser capture microdissection, as pioneered by Lance Liotta's

group at the National Cancer Institute. However current technologies fall short

of the goal of providing a complete proteome expression profile at the level

of a few hundred cells or less. Lance Liotta and Emanuel Petricoin outlined

and described the FDA-NCI Clinical Proteomics Program and stressed the critical

need to bring proteomic technologies to the bedside and use proteomics for direct

patient benefit as quickly as possible. Tim Reilly, president, Micromass, noted

that further improvements in mass spectrometry throughput will likely emerge.

However additional technologies are needed at the front end of mass spectrometry

to reduce sample complexity, prior to mass spectrometric analysis. Several participants

pointed out the potential of antibody-based microarray technology for large

scale, high throughput, high sensitivity, protein expression profiling.

### **Functional proteomics**

Discussions dealing with functional proteomics focused on current strategies

to define protein-interaction, to determine the role of individual proteins

in specific pathways and cellular structures, to elucidate their structure and

determine their function. It was noted that there are currently funded initiatives

aimed at elucidating protein structure and there are private efforts at determining

protein-protein interaction on a large scale. However the current technologies

for protein-protein interaction analysis suffer from a lack of standardization

and a limited ability to elucidate interactions in a physiologic context on

the necessary scale. Additionally, there is interest on the part of several

NIH institutes and other government agencies to fund programs in proteomics.

Joshua LaBaer reviewed effort at creating a comprehensive cDNA repository that

enables protein expression in different experimental formats. There was much

interest in functional proteomic approaches that are pathway driven and that

focus on a particular set of proteins. Ron Taussig described the effort of the

Alliance for Cell Signaling in this regard, with a focus on proteins involved

in cell signaling and which could serve as a model for approaches that target

specific protein sets. The challenge of elucidating protein function and the

contribution of various protein post-translational modifications was considered

most daunting.

### **Knowledge base**

A critical issue pertaining to proteome mining efforts has to do with organizing

proteome related data into a knowledge base. Fred Cohen discussed the feasibility

of such an undertaking. It was felt that a project aimed at developing infrastructure

resources and databases that regroup currently scattered proteome related data

and that would result in an annotated human proteome has substantial merits.

Many participants felt that such an effort should engage, in addition to informaticians,

a community of investigators that is as diverse as is needed to tackle the ensemble

of human proteins. The constructed protein encyclopedia would be constantly

updated with additional knowledge. Several participants felt that a protein

encyclopedia project may well be our best shot at a unified human proteome project.

It is a project that would be viewed enthusiastically by both the public and

the private sectors as it would transform scattered information that is generated

with public funding or that exists in the public domain, into an organized resource

to everyone's benefit.

Ed Liu pointed out that a decentralized approach to a proteome knowledge base

effort is feasible with current technologies and has the added benefit of extending

buy-in from government agencies in different countries that might fund platforms

established in home institutions. Moreover, this decentralized approach will

help those nations that are less developed in proteomics to learn through direct

participation.

## **Recommendations**

There was consensus that a proteome project should be developed that combines

the elements of expression proteomics, functional proteomics and a proteome

knowledge base. Technology development should be an integral component of the

project. There should be open access to the data resulting from the project

at a pre-competitive level with data accessible to all users. Strong consideration

should be given to integrating proteome knowledge with other hierarchical data

at the genome level. The project should have definable milestones and deliverables.

A planning and piloting phase should precede the production phase of the project.

During the planning phase, emphasis should be placed on model systems such as

lymphoid cells, body fluids, and/or other well defined cell populations to prototype

multiple approaches to proteome analysis. Additionally, in the planning phase,

a focus could be placed on defined subsets of proteins that may represent a

sub-cellular compartment or a specific pathway, for their comprehensive characterization

across different cells and tissues. Most importantly, there need to be substantial

support for technology development to achieve the goals of the proteome project.

During the planning phase, funding will be needed to support the creation of

proteome centers of excellence that have both a technology development and a

biological application focus. Additionally, support for alliances of investigators

following the model of the Alliance for Cell Signaling is highly desirable.

The lack of a sufficient pool of researchers trained in proteomics was also

noted, with the recommendation to support a variety of training activities.

Protein reference standards, either as cocktails of recombinant proteins, cell

lysates, or complex protein mixtures such as human plasma or sera, will need

to be identified so that experimentalists from a variety of institutes can compare

technologies and results. It was pointed out that quality control and quality

assurance procedures should be developed and implemented early on in the proteome

project. There were also some specific recommendations for the role of HUPO.

In the early phases of consensus building, HUPO should encourage and support

nucleating groups and hubs to form, support common formats to facilitate large

scale scientific interactions between hubs and encourage scholarly proteome

related activities through scientific meetings, workshops and other educational

activities. HUPPO may want to raise funds or to provide the name recognition

to help local communities to start the nucleation process and support individual

investigators in different countries to work together in a focused and synergized

manner. It is expected that some of these efforts will be supported by both

governments and corporations.

Finally it was recommended that working groups be formed around expression

and functional proteomics and around a proteome knowledge base to begin to tackle

specific components of the proteome agenda. A follow up meeting with expanded

participation is planned for spring 2002 in Bethesda, sponsored by the National

Institutes of Health, to continue the process of planning and consensus building.

Sam Hanash

University of Michigan School of Medicine

Ann Arbor MI 48109 (USA)

E-mail: [shanash@umich.edu](mailto:shanash@umich.edu),

Emmanuel Petricoin III

Food and Drug Administration

FDA-NCI Clinical Proteomics Program

Bethesda MD 20892 (USA)

E mail: [petricoin@cber.fda.gov](mailto:petricoin@cber.fda.gov)

Lance Liotta

Laboratory of Pathology

National Cancer Institute, NIH

Bethesda MD 20892 (USA)

E mail: [lance@helix.nih.gov](mailto:lance@helix.nih.gov)