

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

First Human Proteome Organisation congress

ArticleInfo		
ArticleID	:	4646
ArticleDOI	:	10.1186/gb-spotlight-20021129-01
ArticleCitationID	:	gb-spotlight-20021129-01
ArticleSequenceNumber	:	312
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	5
ArticleHistory	:	RegistrationDate : 2002-11-29 OnlineDate : 2002-11-29
ArticleCopyright	:	BioMed Central Ltd2002
ArticleGrants	:	
ArticleContext	:	130593311

Laura Spinney

Email: LFSpinney@compuserve.com

PARIS - The [first annual meeting of the Human Proteome Organisation](#) in Versailles last week (November 21-24) threw up several burning issues that speakers said must be resolved if the project is to be successful.

Proteomics researchers estimate that the human body contains around one million different protein molecules. Considering there is still no agreement on the number of genes in the human genome - anywhere between 30,000 and 60,000 - it's anyone's guess as to how meaningful that estimate is.

Gene products can be altered in many different ways, for example, by alternative splicing of the gene, by post-translational modifications of the protein, not to mention tweaking by enzymes and non-enzymatic processes in different tissue types. But as estimates go it's good enough to make the point that the proteome poses a far greater challenge than the Human Genome Project (HGP) did 15 years ago.

So most researchers agree that mining the human proteome will require industrial-scale tools. The core technologies of proteomics are gel electrophoresis and liquid chromatography for separating the proteins in a given sample, followed by mass spectrometry for analysis. But although they are continually being refined, each of these technologies has its flaws: gel electrophoresis only detects the most abundant proteins, whereas chromatography and mass spectrometry are not yet ready for the study of complex tissue samples.

Speaking at the Human Proteome Organisation (HUPO) congress, Richard Simpson of the Ludwig Institute for Cancer Research and the [Walter and Eliza Hall Institute of Medical Research](#) in Melbourne, Australia summed up the feeling of many researchers when he said that expectations of proteomics went way beyond what the tools permit. "None of the technologies we have at the moment is robust enough to support a high throughput initiative," he said.

Perhaps most importantly, the lowest limits of protein detection are currently in the nanomolar (10^{-9}) or picomolar (10^{-12}) range. But as Denis Hochstrasser of the [Geneva Proteomics Centre](#) at Geneva University Hospital in Switzerland points out, that only accounts for a portion of the proteome. In blood, for instance, albumin and immunoglobulin are abundant (in the milli- or micromolar ranges), but potential disease markers such as parathyroid hormone and tumor necrosis factor are there in the low picomolar or femtomolar (10^{-15}) ranges, and hence are invisible to today's technology.

The only way to detect those tiny protein quantities in blood is to start with enormous volumes of plasma, says Hochstrasser - between five and 10 litres. That creates problems on several levels. First, as a single patient can't provide that much plasma, samples have to be pooled. That has the effect of diluting individual phenotypes and increasing the load of plentiful proteins. It means that you need efficient techniques for purifying the mixture, dividing it into workable fractions; and it means that pooled samples have to be prepared to the same standards - not always easy when they are collected in a clinical setting.

"There's a whole school of thought that says the plasma proteome is going to be the vehicle for solving a lot of clinical diagnostic problems," says Ralph Bradshaw of the [University of California](#),

Irvine, editor of the journal [Molecular and Cellular Proteomics](#). As HUPO gets ready to launch the pilot phase of its Plasma Proteome Project, researchers are increasingly concerned about the lack of standardization in sample preparation.

According to Julio Celis of the Institute of Cancer Biology and Danish Centre for Human Genome Research in Copenhagen, potential diagnostic markers that have shown promise first in cell lines and then in clinical samples now have to prove themselves reliable on a large scale. Only if a technique accurately diagnoses disease in thousands of patient samples will it have a hope of translating to the clinic. But, he warns, "To do that you have to organise the sample collection, the number of patients, the clinical information."

Standardization in bioinformatics was a major theme of the HUPO meeting. The organization launched its first major initiative, the [Proteomics Standards Initiative](#) (PSI) last April. The aim of the PSI is to define standards for representing all kinds of proteomics data so that they can be compared, exchanged and verified. To begin with, it is focusing on mass spectra and protein-protein interaction data. At the moment, new spectra tend to be compared to theoretical spectra rather than to existing experimental spectra - even though real and theoretical data are known to differ in unpredictable but reproducible ways. So the aim is to set up databases of experimental mass spectra. And although several protein-protein interaction databases already exist, their data are recorded in different formats which need to be synchronized.

Everyone agrees that rapid and easy access to all the available data will be vital. And in that spirit, the US [National Human Genome Research Institute](#) and five other National Institutes of Health bodies have awarded \$15 million to a project to combine the two largest protein sequence databases, [SWISS-PROT](#) in Geneva and [TrEMBL](#) in Cambridge, with the US-based [Protein Information Resource](#). The resulting database will be called the United Protein Database (UniProt).

SWISS-PROT and TrEMBL currently contain around 50,000 human protein sequences. By eliminating some of these due to repeated entries, and adding predictions relating to genes not yet identified, SWISS-PROT coordinator Rolf Apweiler of the [European Bioinformatics Institute](#) in Cambridge arrives at a working total of 33,000 human sequences. Of these, he says, 9000 are annotated to a high degree - that is, something is known of their function, their interactions with other proteins, diseases associated with defects in them and other details.

Meanwhile, a new European consortium launched in October, [Structural Proteomics in Europe](#) (SPINE), aims to solve the structures of 500 disease-related proteins within three years - including human proteins involved in cancer and neurodegenerative diseases - and to establish a Europe-wide network of centres working on high-throughput structure determination.

It's an important start, says Ian Humphery-Smith of the [Department of Pharmaceutical Proteomics](#) at the University of Utrecht in the Netherlands, adding that, "Long before we have a knowledge of the totality of the human proteome, research outcomes are expected in the short term to improve [diagnosis in] precocious diseases such as ovarian and prostate cancer." For that reason, and because proteomics is more dependent on industry than genomics, intellectual property has now become a burning issue.

As a founder of [GeneProt](#), one of several companies that carry out mass-spectral proteome screening on an industrial scale, Hochstrasser admits that the drive toward high-throughput analysis has already

forced him into a conflict of interest, and he believes some resolution is urgently required. Lee Hood, president of the [Institute for Systems Biology](#) in Seattle, which is developing new technology for the US National Institutes of Health's own proteomics initiative, offers a lesson from the HGP: "The key thing is: don't make the information confidential and proprietary."

In 1996 the participants of the HGP agreed the Bermuda principles, laying down the ground rules for cooperation between the public and private sectors, but no equivalent exists for the proteome. The task of defining the entire human proteome, most agree, will take decades. Yet according to Hood, the success or failure of the entire venture rests on this one issue: the resolution of intellectual property rights.

References

1. Human Proteome Organization - First World Congress, [<http://www.hupo.org/new/congress1/index.html>]
2. Walter and Eliza Hall Institute of Medical Research, [<http://www.wehi.edu.au/>]
3. Geneva Proteomics Centre, [<http://ca.expasy.org/gpc/>]
4. University of California, Irvine, [<http://www.uci.edu/>]
5. *Molecular and Cellular Proteomics*, [<http://www.mcponline.org/>]
6. Proteomics Standards Initiative, [<http://www.ebi.ac.uk/Information/meetings/psi.html>]
7. National Human Genome Research Institute, [<http://www.genome.gov/>]
8. SWISS-PROT, [<http://www.ebi.ac.uk/swissprot/>]
9. TrEMBL, [<http://www.ebi.ac.uk/trembl/>]
10. Protein Information Resource, [<http://pir.georgetown.edu/>]
11. European Bioinformatics Institute, [<http://www.ebi.ac.uk/>]
12. Structural Proteomics in Europe, [<http://www.spineurope.org/>]
13. Department of Pharmaceutical Proteomics, University of Utrecht, [<http://www.uu.nl/uupublish/homeuu/homeenglish/1757main.html>]
14. GeneProt, [<http://www.geneprot.com/scripts/index.asp>]
15. Institute for Systems Biology, [<http://www.systemsbiology.org/>]

