

DNA stored at room temperature...

HUPO Plans Ambitious 10-Year, \$1B Project to Map Entire Human Proteome

May 1, 2008
By Tony Fong

The Human Proteome Organization plans to launch an ambitious initiative to map the entire human proteome, a project that is anticipated to last up to a decade and cost \$1 billion.

In an interview with *ProteoMonitor* last week, John Bergeron, immediate past president of HUPO and a professor of anatomy and cell biology at McGill University, broadly outlined the goals of the initiative, called the Human Proteome Project, which HUPO began thinking about in 2006.

HUPO is preparing a paper for publication detailing its plans for the initiative, and Bergeron said it will focus on three technologies: those that are based mass spectrometry, antibodies, and protein interactions.

Bergeron said he hopes to launch the project in August during HUPO's annual conference in Amsterdam, and to begin doing the actual work next year. In total, he said he believes the project would last a "maximum of 10 years," but with the rapid development of new technology it could even be finished in five years, he added.

"The big picture for us is that it will fill the void between genotype and phenotype," Bergeron said. With the human genome mapped, "all large-scale projects right now are ... focused on trying to indicate what genes or what proteins might be linked to some sort of a phenotype — baldness, stroke, cancer, you have it," he said. "The objective here is to say what proteins are actually present in the protein-coding genes, where these proteins are expressed, what they are doing, and what their partners are."

The first phase of the project would collect all the "high-quality" proteomics-related mass-spectrometry data in order to evaluate "where we are in terms of the global community with matching high quality tandem mass spectra to the protein-coding genes of the human genome," Bergeron said. The amount of such spectra would number in the millions, he said.

"From [those] tissues, organs, organelles, and body fluids [where] we see the highest quantity of tandem mass spectra of high quality assigned unambiguously to peptides and their cognate unique proteins, each representative of a single protein-coding gene would be delivered," Bergeron said.

With a team of five full-time personnel, that task would take about six months, he said. Afterward, a pilot project would map out the protein-coding genes of human chromosome 21, which at about 47 million nucleotides is the smallest in the genome.

Because chromosome 21 contains only between 195 and 299 protein-coding genes, Bergeron said that researchers would be able to "easily" make proteotypic peptides for it, characterize the proteins, raise the antibodies in the chromosome, and evaluate the effort it took to map the chromosome correctly by using both antibody-based and mass spec-based technologies.

"From doing chromosome 21 as a pilot phase, we'll know exactly what the two sets of communities can do," Bergeron said, adding that the effort could be completed within 18 months.

He said that in order to keep the project manageable, HUPO would not set out to map every single protein with all its post-translational modifications and isoforms, a task that many think may be unachievable given the dynamic nature of the proteome. Rather, the goal would be to determine one “representative” protein for each protein-coding gene — the major protein expressed.

The thought is that while both antibody and mass-spec analysis pick up splice variants in proteins, the number of such variants is relatively low and don’t significantly affect the complexity of the genome, Bergeron said.

Two forms of post-translation modifications would be looked at as part of the project, however — glycosylation and phosphorylation — as HUPO is already conducting research into those areas, he said.

To ensure the results are valid, all data generated from the Human Proteome Project and made public will have to be cross-validated by an independent lab.

The \$1 billion price tag was estimated based in part on the cost of the Human Protein Atlas and HUPO’s Human Antibody Initiative to date, as well as personnel that will be needed for performing bioinformatics, computational biology, and data analysis tasks.

According to Tommy Nilsson, co-chair of HUPO’s Mouse Models Initiative, the cost is “peanuts.”

Compared to most academic endeavors, the level of funding needed for the project may appear outlandish, said Nilsson, who is also scientific manager at the Center for Cellular Imaging and Center for Proteomics at the Sahlgrenska Academy at the University of Gothenburg. But he said that “rather than comparing it to academic funding, I think one should put it in the category of human achievement — and this will be a human achievement.”

Beyond the Genome

The idea for the Human Proteome Project was first proposed in 2006 while HUPO was readying to launch its protein-standard mixture initiative when officials received a call from members of the European Commission inquiring about large-scale projects HUPO might be undertaking, Bergeron said.

Over the ensuing months, HUPO officials explored the idea of a proteomics equivalent of the Human Genome Project. The following year researchers further debated and fleshed out the project.

Along the way, two other HUPO projects became crucial to pushing the Human Proteome Project ahead: the protein-standard mixture initiative [See [PM 07/20/06](#), [04/19/07](#), and [09/06/07](#)] and a study to test the reproducibility of 2D gel technology [See [PM 11/15/07](#)].

“You have to accept the complexity [of the human proteome] and just deal with it. That’s the only way I think you’re really going to gain knowledge.”

The protein-standard mixture initiative proved the functionality and accuracy of mass spectrometry. That ongoing work indicates “at least on the mass spec side, what the limitations are and what bottlenecks there could be in doing a large-scale project,” Bergeron said. “We really do think that we have tested many of the most obvious bottlenecks ... and have resolved them.”

Meanwhile, the 2D gel study, carried out by HUPO’s Industry Advisory Board, addressed another major bottleneck in proteomics: reproducibility. In the study, participants showed that, contrary to the prevailing opinion, proteomics can be done in a reproducible manner across different laboratories as long as proper protocols were followed.

“If the will is there, if people set their mind to it, it can be done,” said Will Dracup, CEO of Nonlinear Dynamics, who was involved in the study and was at a HUPO meeting in January in Barbados, which was largely dedicated to planning out the Human Proteome Project.

While the initial study looked at 2D gel, the fact that it could be done with that technology indicates it should be doable with any proteomic technology, Dracup said, adding the IAB hopes to do a similar study with mass specs.

In further developing its proposal for the project, HUPO was careful to consider input from the proteomics-wide community, Nilsson said.

“HUPO has been asked to come up with a proposal and it has responded and done so in the best possible way, taking a very inclusive approach where everyone got together in Barbados to put together this draft proposal on the table,” he said. “It has a very big, solid support in the community.”

Indeed, the work of the Human Proteome Project will need to be carried by proteomics researchers and laboratories throughout the world.

It took 13 years to complete the Human Genome Project, and mapping the human proteome, by comparison, is significantly more difficult, even under HUPO's plan to identify only the major protein for each protein-coding gene. Taking on the additional responsibility of identifying post-translational modifications, isoforms, and splice variants would increase the complexity of the task by even more orders of magnitude, several researchers said.

Yet some believe that for the project to be truly beneficial, that's the path HUPO should take and within the organization, there is ongoing debate on the issue. Albert Heck of the Netherlands Proteomics Center and chair of HUPO's Proteome Biology of Stem Cells project said in an e-mail to *ProteoMonitor* that a mass spec-based approach to identify every protein in the human proteome is still "far from being able to unravel the human proteome in its full glory." Knowledge about modifications and protein processes are "essential to understand the human proteome at the functional level."

Michael Washburn, director of proteomics at the Stowers Institute for Medical Research, said that the "representative protein" approach that HUPO is proposing can generate a huge database of proteins, but "the utility of all of that in the end is hard to say.

"The point of it all is to basically provide scientific knowledge and discovery that facilitates research," he said. "Trying to reduce something to its simplest form ignores its complexity of what it is to be a human proteome ... You have to accept the complexity and just deal with it. That's the only way I think you're really going to gain knowledge."

Funding Headwinds

At this point, it is unclear how the project will be funded as no funding agency has yet committed any money to it. In the US, given the current funding environment, it will not be easy to get federal financial support, said Washburn.

"The main thing is you have to prove what this will enable," he said. HUPO will have to clarify what "this type of project [is] going to provide in the short term and in the long term that wouldn't happen otherwise: How is it going to enable disease research, just basic human biology, and biochemistry, and cell biology? What are the outcomes going to be and how are they really going to change the way people are able to pursue biological research? I think that's pretty much the bottom line."

Sudhir Srivastava, chief of the cancer biomarkers research group, division of cancer prevention at the National Cancer Institute, said that the NCI is not yet convinced about the need for the Human Proteome Project. Colleagues at the institute have said that the project is too broad and some of the work that HUPO wants done as part of the initiative is already being carried out by other researchers, he said.

His advice to HUPO has been to make the project disease-specific so that funding agencies can digest what aims are being set forth and can see the potential benefit from such a project.

The European Commission is in listening mode and measuring what needs could arise among European researchers if the project gets off the ground, said Tomasz Dylag, a scientific officer at the EC who is in charge of proteomics at the commission.

"Certainly this Human Proteome Project is something that goes far beyond the possibilities of funding of any single funding agency," he said. "Also, certainly this \$1 billion [price tag] is something that is beyond the scope of the European Commission. Let that be clear."

Aside from the funding, obstacles exist in terms of the quality of the work that has been done as well as methodological bottlenecks. After promising a decade ago that proteomics would lead to new diagnostics and therapies, the sense now is that proteomics has oversold itself and failed to live up to expectations, an assessment that could hurt HUPO in its efforts to round up financing.

"That works against [the Human Proteome Project] because [proteomics hasn't] shown anything," Srivastava said.

The field also still lacks clear standards in many aspects. Though HUPO's Proteomics Standards Initiative has been working to rectify this, it remains an issue that has thwarted proteomics results.

Rolf Apweiler, president of HUPO, said even if not every obstacle has been overcome yet, there is "some light at the end of the tunnel." It should also not be a problem to work on the Human Proteome Project in parallel with smaller quality- and workflow-related projects, he said.

"You prepare for a bigger project and at the other end you try to sort out what you still see as important work which needs to go on,"

Apweiler said.

Regardless of any potential headwinds the project may face, however, Bergeron was steadfast in his conviction that it will proceed.

“You know how much money was spent on the sequencing of the human genome, and there clearly is a need to know what the protein-coding genes are ... if we want to make headway in terms of mapping these genes to human disease,” he said.

In addition to filling the gap between the genotype and phenotype, the Human Proteome Project will provide “what I think proteomics is meant to do, and that is to go after all that part of the genome that is poorly characterized in terms of what it’s doing — the protein-coding genome,” Bergeron said. “There is a significant portion of the protein-coding genome whose predicted proteins are just unknown, so this project will just automatically characterize them.”