

# The need for national centers for proteomics

Ruedi Aebersold and Julian D. Watts

Last year, the biological sciences witnessed a historical landmark—the announcement of a completed first draft of the human genome<sup>1,2</sup>. This information promises to transform human biology in the same way that the yeast genomic sequence has transformed yeast biology. With the raw sequence in hand, most of the work still lies ahead—is biological interpretation of sequence information. We believe this will rely heavily on the fast-emerging field of proteomics. Before proteomics can fulfill this potential, however, access to the technology will have to be dramatically improved.

Two recent papers<sup>3,4</sup> published in *Nature* describe the application of large-scale proteomics technology to protein interactions in yeast. Both reports come from the private sector and represent significant investments in manpower and technology. Both present catalogs (databases) of proteome-related data and therefore represent a direct projection of the “genome-sequencing” way of thinking onto proteomics. The generation of such databases is useful as a source of reference information, provided that they are publicly accessible and contain high-quality information. However, they do not explain biology and thus represent only a (small) part of what proteomics promises.

Although such large-scale, high-throughput approaches may seem to be akin to genome sequencing, this is a misconception. Proteomics is not an exercise in protein cataloging. Instead, it is increasingly used as a biological assay<sup>5</sup> in which specific properties of a biological system, and how they might change, are investigated on a global (proteome-wide) scale. An exciting part of proteomics is that proteins isolated from any given model system can yield much information not currently interpretable from genomic sequence data. This includes data on protein abundance, activity, complex composition, and structure, parameters that in principle can be determined and interpreted either individually or in combination. The power of proteomics over genomics is that changes in these properties can also be determined with respect to time, or some other introduced perturbation.

Ruedi Aebersold is cofounder and James D. Watts is a senior research scientist, Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103  
(raebersold@systemsbiology.org).

If we accept the view that proteomics is a central means of accessing the biological significance of genomic data, several things follow. First, as large numbers of proteins are generally measured in each single experiment, high-sample-throughput facilities for data collection and processing are required. Second, there is no end point because the assay can be applied to an essentially infinite number of questions. And third, given this, we need to allow the scientists working on the most interesting model biological systems—that is, academic researchers—easy access to the technology. At present, this last requirement is problematic.

Proteomics facilities are technically and operationally elaborate entities. By nature, they require a large monetary investment and substantial lead time for establishment. This makes them inaccessible to most academic scientists. Minimally, such facilities require the expertise to process complex protein samples, to collect data on all the proteins in the samples, and to computationally process and interpret the data. It is therefore not practical, or cost effective, to establish such technology within individual academic investigator laboratories, or in the form of core facilities, at most research institutions. Although academics may gain access to proteomics technology through contracts with industrial partners, this may also prove prohibitively expensive, and additionally require signing away some or all of any intellectual property rights for data produced. We believe a solution is to establish publicly supported facilities that would make proteomics accessible to (typically) publicly funded academic researchers.

To this end, there first needs to be a recognition at the various levels of government, and within the various public funding agencies, that setting up such programs would benefit many of the research programs they are currently funding. Given this, how might such facilities look? Where would they be located? How would they be organized? And what programs would they house? Drawing on the examples of existing US national laboratories and technology centers, one could envision the establishment of “national centers for proteomics,” either as primarily stand-alone entities or as new programs within existing multidisciplinary national laboratories.

This is nothing new. Existing programs that make costly and sophisticated technology available to publicly funded academic researchers include the various high-energy

particle physics (cyclotron) laboratories and astronomical observatories (telescopes) dotted around the United States and the world. Continued support for such centers represents a recognition that their existence is in the interest of the general scientific community. Systems in place for funding, operating, and providing access to these centers might serve as a starting point for creating new centers for proteomics. One could thus imagine proteomics centers being added as new divisions of existing institutions, potentially saving some time and money because of the administrative infrastructures already in place. Alternatively, and in our opinion preferably, free-standing proteomics facilities could be established from scratch or under the umbrella of academic institutions willing to undertake the responsibility for their oversight.

Finally, it will be important to discuss what such proteomics facilities might do and what services they should make available. Hardware, software, and methodologies for proteomics are evolving at a great rate, and there is no reason to think this will change any time soon. There are also multiple platforms for proteomics, with no single platform being superior. Rather, the biological question being asked determines the best platform to apply for maximal return. As one can imagine, this requires a judgment call, and analytical platforms are constantly being optimized and redesigned to improve their performance and allow them to tackle new problems. Thus, it is essential that any “national centers for proteomics” not be just service providers, but include significant research and development components, ideally tied to some in-house biological research programs. This would ensure that the hardware, software, and analytical approaches that make proteomics the powerful biological approach it is, remain both available and state-of-the-art.

We believe that in a few years access to such centers would transform many of the established publicly funded biological and medical research programs currently underway at universities throughout the United States and the world. With many genome sequences already completed and many more nearing completion, the time to start working on this is now.

1. Lander, E.S. *et al.* *Nature* **409**, 860–921 (2001).
2. Venter, J.C. *et al.* *Science* **291**, 1304–1351 (2001).
3. Gavin, A.C. *et al.* *Nature* **415**, 141–147 (2002)
4. Ho, Y. *et al.* *Nature* **415**, 180–183 (2002).
5. Haynes, P.A., Gygi, S.P., Figeys, D. & Aebersold, R. *Electrophoresis* **19**, 1862–1871 (1998).