© 1998 Nature Publishing Group http://www.nature.com/naturebiotechnology

from the gene into which the DNA was inserted. The Lexicon researchers refer to these sequence tagged sites as Omnibank sequence tags (OST), which are maintained in a proprietary database. The authors have shown that although there is a slight bias for integration of retroviral vectors carrying the selectable gene at the 5′-end of genes, integration into all parts of genes occurs. Forty percent of the OSTs correspond to genes and ESTs in databases. The remaining 60% also represent genes, as determined by expression analysis, and reveal the current bias of the ESTs, most of which are derived from 3′-end sequences.

At time of publication, the authors had identified 2000 OSTs and they claim to be generating additional OSTs at a rate of 500/week. At this rate, one genome equivalent of OSTs will require a mere 200 weeks, and additional automation might reduce this time further. Lexicon generously offered ES cell clones corresponding to the 2000 OSTs to investigators for noncommercial research. This generosity would presumably expand to include additional ES cell clones as they are generated.

Does this technology solve all of the needs of geneticists anxious to study gene mutations? To have every gene represented in the Omnibank, it is necessary to have 300,000 or more gene insertion events. This may be beyond the capacity of Lexicon. Some skeptics might argue that their approach will yield only null mutants and therefore has limited utility. Although there is some merit in this argument, it has to be recognized that the approach of generating null mutants by gene targeting in mouse ES cells has served, and continues to serve, the community well. In many laboratories, the time required to generate ES cells containing any one desired mutation may take 6–12 months of effort. The Omnibank resource would undoubtedly save those laboratories a lot of effort.

Omnibank provides a collection of mutations with no information about their phenotype. Efforts to develop a large collection of phenotypically characterized mouse mutants obtained after chemical mutagenesis are underway in Europe, and the United States is contemplating similar efforts. Genes involved in the individual phenotypes have to be identified by positional cloning methods. It would be valuable to have methods that can efficiently introduce point mutations into specific locations of genes. The use of chimeric DNA–RNA oligonucleotides[4] may prove to be a valuable tool to accomplish this goal.

High-throughput functional genomics is just getting underway. It is clear that the daunting task of deciphering gene function on a genomic scale will require the adoption of new and audacious approaches. In this respect, Omnibank leads the way and shows how to think big in genome science and biology.

1. Adams, M.D. et al. 1991. *Science* **252:**1651–1656.
2. (http://www.ncbi.nlm.nih.gov/)
3. Zambrowicz, B.P. et al. 1998. *Nature* **392:**608–611.
4. Cole-Strauss, A. 1996. *Nat. Med.* **4:**285–290.

Postgenomic protein analysis: The next bend in the river

Andrew R. Mendelsohn and Roger Brent

It seems that every few months brings another postgenomic conference, and with it a view of the next bend in the river rushing toward the future of genomics research. A recent meeting in San Francisco* may have given us a glimpse of one such bend. If that view holds true, the next few years will likely see physical techniques like mass spectrometry begin to make a significant contribution to gene functional analysis.

Given that functional genomics—the large-scale, systematic production of information about gene function—is barely more than three years old (the first "After the Genome" conference occurred only in October 1995), it is not surprising that new types of genomic information are appearing rapidly. The first such information type was DNA sequence data, of which their is now enough that subscribers to commercial databases can get at least some protein sequence for most human genes. The second information type to mature will likely be gene expression monitoring data; that is, data about the presence and abundance of different mRNA species in different cell types, tissues, and disease states.

Andrew R. Mendelsohn and Roger Brent are at The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, California 94704 (brent@sequence.molsci.org).

A third information type—inventories of the interactions among proteins and other molecules in humans and model organisms—will become available within the next decade. It is likely that this will first emerge from interaction-mating yeast two-hybrid experiments. At the core, these experiments rely on robust biological technologies that, although sophisticated in terms of the cellular engineering involved, can practically be performed in small laboratories. The value of this information is likely to be augmented by other biotechnologies that facilitate the dissection of protein networks by disrupting or enhancing particular intermolecular interactions. But the meeting clearly indicated that mass spectrometry techniques will at least complement these yeast-based interaction methods and may well supersede them.

Mass spectrometry allows precise measurement of the mass-to-charge ratio of ions, so that, if the ionization state of an analyte is known, one can directly calculate its mass. If the analyte is a protein or peptide, and its size matches that predicted from DNA sequence data, one has identified that peptide as surely as if one has sequenced its amino acids. If two proteins interact, continue to interact through some purification steps, and can be identified afterwards, one has identified that interaction. For mass spectrometry to help identify interactions, then, the first enabling condition is complete knowledge of the sequence of the proteins encoded by a genome—information already available for *Escherichia coli* and yeast. The second enabling condition is the development methods for achieving relatively controlled ionization of proteins.

Two such gentle methods, which come down to us from the 1980s, are electrospray ionization mass spectrometry (ESI-MS) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). These techniques are applicable to the study of interactions between biological molecules such as proteins because they tend to preserve the structure of an analyte during ionization.

In ESI-MS, small drops of liquid containing the analyte are sprayed through a charged needle into a vacuum, where the solvent evaporates, leaving the molecular ions to be analyzed. Ions of a desired mass-to-charge ratio are focused by a quadrupole filter into a detector. The filters are usually arranged in tandem, so that, for example, proteins of a

© 1998 Nature Publishing Group http://www.nature.com/naturebiotechnology

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Condition n

Figure 1. A SELDI chip. The ablative surface is derivatized at each spot on the chip with different functional groups commonly used in protein purification. The chemical identity of some of these groups is shown in the cones. In a typical fractionation, a protein mixture that sticks to hydrophobic groups (top left), can be further fractionated by varying the concentration of hydrophobic solvents in conditions 1, 2, 3, and 4 (first column), prior to laser desorption/ionization.

given mass can be isolated, fractionated by passage through argon, and then peptide fragments of given masses isolated. Sensitivity to a given ion borders on the attomolar, although to achieve that sensitivity one must know the mass of the ion one is trying to detect[1,2].

In MALDI-TOF-MS, a protective matrix compound is coprecipitated with the analyte. A laser vaporizes the matrix, which absorbs most of the beam energy (thus protecting the protein). The proteins are simultaneously blasted out of the matrix and singly protonated. Here there is no tuning, and the detector simply registers all the ions as they come in. The higher the mass, the longer the ions take to reach the detector. Femtomolar sensitivities are possible[3,4].

An interesting new technology reported at the conference, arguably an outgrowth of MALDI-TOF, allows copurification with identification. William Hutchens of Ciphergen (San Diego, CA) described a technology termed SELDI, for surface-enhanced laser desorption/ionization.[5] Unlike MALDI, SELDI interrogates sample on an ablative surface rather than in a matrix. This surface contains discrete spots derivatized by the same functional groups that are used chromatographically to separate proteins, and which can vary in their concentration of these functional groups. Rapid liquid handling allows separation of protein mixtures along different chromatographic dimensions.

If two proteins interact, colocalize on the chip, and can be identified, the interaction is identified. Alternatively, in principle, use of discrete spots derivatized with specific proteins could allow identification of interacting partners from complex mixtures[5]. Whether ESI, MALDI-TOF, SELDI, or other acronyms to be developed emerge as the mass spectrometry method of choice, it appears near certain that approaches that combine copurification and mass spectrometry will be used to generate interaction data.

The next bend in the river may show us still more types of functional data, and, perhaps its conjunction with computational approaches to aid their conversion to biological knowledge. The accumulation of these new types of functional information into databases large enough to be useful is going to be costly, and the more exotic the generating technology, the greater the expense is likely to be. On the one hand, this will allow the private sector to fuel the creation of these new data types, but on the other, there may be some unsettling repercussions for the wider research community. Biologists whose employers can pay for them to access data generated in private facilities will be at a distinct advantage over biologists who's employers cannot. Not only will they be able to make more functional inferences about a protein, but also they will be able to make them quicker. These trends, like the river itself, seem to be speeding up.

**2nd Annual Gene Functional Analysis Conference, February 9–11, Fairmont Hotel, San Francisco.*

1. Yamashita, M. and Fenn, J.B. 1984. *J. Phys. Chem.* **88**:4451–4459.
2. Schwartz, J.C. and Jardine, I. 1996. *Methods Enzymol.* **270**:552–586.
3. Karas, M. and Hillenkamp, F. 1988. *Anal. Chem.* **60**:2299–2301.
4. Capriole. R.M., Farmer, T.B., and Gile, J. 1997. *Anal. Chem.* **69**:4751–4760.
5. Hutchins, T.W. and Yip, T.T. 1993. *Rapid Commun. Mass Spectr.* **7**:576–580.

What are you looking for?

New business opportunities?
Fast growing markets?
Strategic R+D partners?
Well-trained people?
Unique public support?

We got it all in a great landscape.

Biotechnologies Mecklenburg-Vorpommern Germany

Come and meet us at:
Bio '98,
New York City, Hilton Hotel,
exhibition stand 208–217,
Rhinelander Gallery

Bioregio Greifswald Rostock
hotline: ++49 (0) 38 34-51 51 08
Fax: ++49 (0) 38 34-51 51 02
http://www.uni-greifswald.de/~bioregio/

READER INQUIRY NO. 1076