

Using Image Informatics

Focusing Proteomics

Suzanne Mattingly

Ongoing investigations into the genetic blueprints of humans and other organisms have confirmed what molecular biologists have long believed: What we call disease is nothing more than a change in the structure, location, or abundance of proteins (1). Consequently, drug research organizations have begun complementing their genomics initiatives with proteome analysis to better understand the mechanisms underlying diseases and treatments.

Proteomics currently affects nearly every stage of the pharmaceutical research and development process. Revealing protein expression in cells, tissues, and in vivo samples, two-dimensional (2D) gel electrophoresis can be used to optimize new leads identified by high-throughput screening and provide crucial insights into structure-activity relationships (SARs). Protein expression data are also valuable in preclinical and clinical testing, where analysis has identified biomarkers that can predict or elucidate in vivo outcomes (2,3). Some researchers have applied proteomics to target identification in the hope that knowing how proteins regulate cellular behavior will help focus studies toward revealing unique targets.

Since 1998, use of 2D gels and mass spectrometry in proteomics research has increased substantially (4,5). Refinements in both techniques have reduced the protein expression analysis bottleneck (6). But now, having alleviated some tedious tasks associated with processing and analyzing gel data, many organizations face the challenge of putting those data to better use. As has often been the case with emerging technologies in drug discovery, proteomics remains somewhat compartmentalized. Expression data come in one form — gel images — that complicates data management. Images are unwieldy data sources because deriving meaning from them

requires interpretation, which can be subjective and unreliable. Consequently, scientists often use expression data only passively (referencing them in reports and summaries) rather than actively to direct project team activities.

Image Informatics

Scimagix, Inc., has commercialized a set of technologies for the life science industry (called image informatics) that render image content into a form that can be stored, searched, and mined like other data. The company's Scientific Image Management System (SIMS) provides a core infrastructure technology for storing image content in Oracle databases. SIMS is enhanced and augmented by applications that incorporate algorithms for visual content searching: finding and retrieving images that are "like" each other.

ProteinMine is one visual-searching application. Developed in collaboration with Pfizer global R&D, it allows search and retrieval of protein expression patterns from databases of 2D gel images. Specifically, scientists can use a given protein expression pattern as a query, retrieving all other instances in which that pattern has been expressed. ProteinMine both improves and expands on the algorithms necessary for analysis (25 gels can be registered, segmented, quantitated, and unwarped in about 38 minutes). Such algorithms provide the tools to find expression patterns, search accurately through them, and link them with other image and nonimage data to create a comprehensive picture of drug or disease effects.

Mining Gel Images

The ability to store, retrieve, and mine 2D gel images opens up new avenues for proteomics investigation. Scientists can access data across experiments to provide insights into the mechanisms of action for

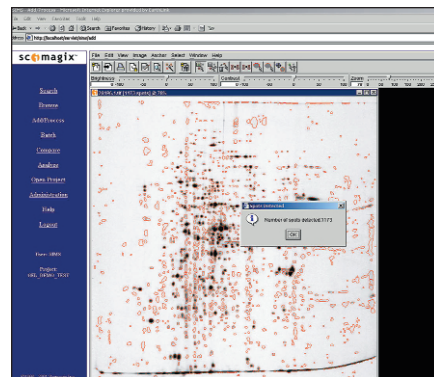


Figure 1. Gel registration. Users click on an expression pattern linked to a particular toxic effect, then use visual search parameters to define the query as shown in Figure 1b. A database search retrieves other instances of the same pattern (Figure 1c), allowing users to explore mechanisms of action and evaluate which compounds or drug-like molecules may be eliminated from or progressed to future investigations.

various compounds and drug-like molecules. When used with the SIMS open architecture, ProteinMine helps organizations associate gel and other vital research data including images generated by other research methods. Mining those data provides ways to verify or disprove hypotheses using image data and influences directly the progress or termination of development projects.

Information gleaned from a database search can help scientists correlate relationships among protein expression patterns, ultimately offering insights into the action mechanisms of drugs under investigation. At Pfizer global R&D, ProteinMine and SIMS support the identification of "proteomic signatures": empirically or heuristically derived protein expression patterns that constitute reproducible markers of specific cellular functions or responses (7).

Patterns associated with particular toxic effects, identified through rigorous experimentation, can be used to screen for compounds and molecules with similar protein expression patterns indicating that perhaps they should be eliminated from further testing. The search example depicted in Figure 1 represents Pfizer's approach in this application.

The ability to archive, reuse, and mine 2D gels helps proteomics researchers better direct R&D activities and rapidly focus investigations. Image informatics can be applied routinely to characterizing new therapeutic candidates. A variety of questions can be answered: Which molecules cause protein expression patterns that are entirely new? Would any be good candidates for follow-on drugs? Do any tend toward toxic mechanisms of action?

Real Results

At Pfizer central R&D, image informatics helped researchers determine that a compound demonstrating in vitro potential for topoisomerase inhibition actually had a different mechanism of action. In vitro data seemed to indicate that the compound acted similarly to novobiocin, a desirable result. But in vivo data clearly revealed a match between its expression pattern and that of polymyxin B, an undesirable outcome. That interpretation supported redirecting the work of six chemists at a savings of \$24,000 per week.

As the throughput of proteomics studies increases, the techniques used to make decisions about the data those studies provide must also become speedier and more accurate. Image informatics offers one mechanism for searching protein patterns and integrating gel images with other key data to rapidly provide a more complete picture of scientific outcomes.

References

- (1) N.G. Anderson et al., "Back to the Future: The Human Protein Index (HPI) and the Agenda for Post-Proteomic Biology," *Proteomics* 1, 3-12 (2001).
- (2) Lesko et al., "Optimizing the Science of Drug Development: Opportunities for Better Candidate Selection and Accelerated Evaluation in Humans," *Pharm. Res.* 17(11), 1335-1344 (2000).

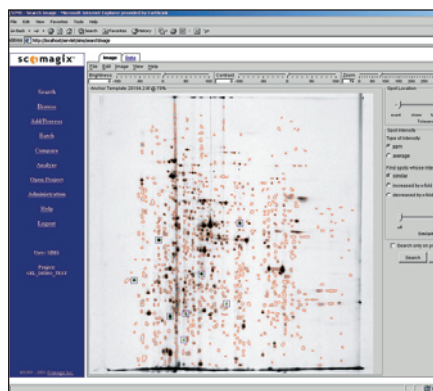


Figure 2. Searching a protein pattern

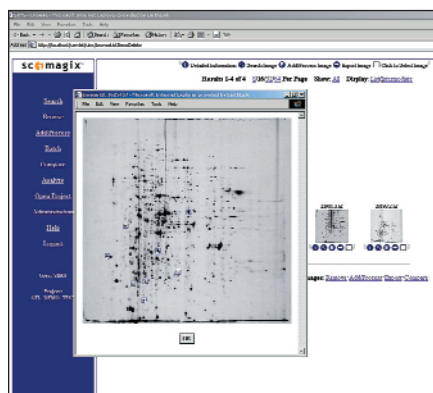


Figure 3. A database search retrieves other instances of the pattern of interest.

- (3) A.S. Vercoutter-Edouart et al., "Proteomic Analysis Reveals that 14-3-3s Is Down-Regulated in Human Breast Cancer Cell," *Cancer Res.* 61, 76-80 (2001).
- (4) S. Martin, "The PE Biosystems/Celera Proteomics Research Center Initiative: What are the Objectives?" presented at the Proteomics Conference, Boston, 22-23 February 2001.
- (5) J. Healy, Venture Capital Panel: "Identifying Promising Technologies and Companies in Proteomics" presented at the Proteomics Conference, Boston, MA, 22-23 February 2001.
- (6) K. Garber, "Proteomics Gears Up," *Signals Magazine: Online Magazine of Biotechnology Industry Analysis* (2 November 1999). Available at www.signalsmag.com.
- (7) J.D. Thomas and R.A. Van Bogelen, "Experimentalism in Proteomics," *Curr. Trends* (July 2000). **AA**

Suzanne Mattingly is vice president of marketing at Scimagix, Inc., 2855 Campus Drive, Suite 100, San Mateo, CA 94403; 650.377.2378; fax 650.377.2499; smattingly@scimagix.com.