DRUG DISCOVERY TODAY BIOSILICO

Data management solutions for protein therapeutic research and development

Jost Vielmetter, Jeff Tishler, Marie L. Ary, Peter Cheung and Richard Bishop

Protein therapeutics, including monoclonal antibodies, are a growing focus of drug discovery research organizations. High-throughput screening of large libraries of protein variants is therefore becoming increasingly important in R&D. As a result, there is a need to link large numbers of variant protein sequences with chemical and biological assay data. This integration will allow more efficient data mining and facilitate decision-making regarding hit identification, lead optimization and drug development. In this paper, we present an implementation in which a widely used small-molecule high-throughput screening data management system has been adapted to meet the unique needs of protein drug discovery and development.

With the rapid advance of high-throughput screening (HTS) technology and the creation of large and diverse chemical libraries, the number and size of datasets created from routine screening campaigns has grown considerably in biopharmaceutical research and development. Managing and extracting valuable information from such datasets has therefore become a high priority for many pharmaceutical companies. Several data management platforms are now commercially available to enable small-molecule drug discovery by facilitating the organization, analysis and subsequent querying of high-volume HTS data. This technology allows efficient data mining and reporting and is widely used in the pharmaceutical industry to assist decision-making regarding hit identification, lead optimization and drug development.

Therapeutic proteins, including monoclonal antibodies, are gaining increased attention in drug discovery research organizations. Global sales in 2003 exceeded US$30 billion compared with US$12 billion in 2000. In this three year period, 30 new protein drugs were approved by regulatory agencies in the US and Europe, accounting for more than a quarter of all new drug approvals [1]. With hundreds more proteins in clinical trials, these numbers will continue to grow, especially since biopharmaceuticals tend to move through clinical development more quickly than small-molecule drugs [2,3]. The expanding interest in protein therapeutics stems in part from their superb affinity and specificity for their clinical targets. In many cases, natural proteins serve as excellent 'lead' compounds. However, evolution has not shaped natural proteins to function as medicines; consequently, they often lack many of the characteristics desired in a therapeutic. Optimization through protein engineering can therefore be quite beneficial, and is now frequently employed in lead development and in the creation of second-generation products [4–7]. Engineering can be used to alter multiple protein properties including efficacy, specificity, solubility, stability, pharmacokinetics and immunogenicity. These can impact drug safety, potency, dosing frequency and route of administration; other considerations that can be affected include manufacturing cost and intellectual property.

Jost Vielmetter*
Marie L. Ary
Peter Cheung
Richard Bishop
Xencor,
111 W. Lemon Avenue,
Monrovia,
CA 91016,
USA
*e-mail: jvielmet@xencor.com
Jeff Tishler
ID Business Solutions Inc.,
1900 Powell Street,
Suite 1070,
Emeryville,
CA 94608,
USA

1359-6446/05/$ – see front matter PII: S1359-6446(05)03495-1

TABLE 1

Examples of companies generating protein libraries

Company	Focus	Technology/method of generating protein libraries
Abgenix	Therapeutic antibodies	High-throughput screening of human antibodies developed in transgenic mice, epitope mapping (XenoMouse™, XenoMax™)
Amgen	Therapeutic proteins and antibodies	Protein engineering, glycosylation engineering, antibody display
Applied Molecular Evolution/Eli Lilly	Therapeutic proteins and antibodies	Directed evolution (AMEsystem™ technology)
Biovation	Therapeutic proteins and antibodies	T cell epitope identification and removal (site-directed mutagenesis) (DeImmunisation™ technology)
Cambridge Antibody Technology	Therapeutic antibodies	Antibody libraries, phage display, ribosome display
Centocor/Johnson and Johnson	Therapeutic proteins and antibodies	Monoclonal antibody technology, structure-based design
Diversa	Therapeutic proteins and antibodies, industrial enzymes	Directed evolution (Gene Site Saturation Mutagenesis™, Tunable Gene Reassembly™)
DNA2.0	Custom gene synthesis, expression optimization, protein engineering	Bioinformatics-based protein design (DeNovo Genes™ technology, custom gene synthesis)
Egea Biosciences/ Johnson and Johnson	Therapeutic proteins and antibodies	High-throughput gene synthesis, rational protein design
FivePrime	Proteomics, target discovery	High-throughput cloning and expression of human cDNAs (ProScreen technology)
Genentech	Therapeutic proteins and antibodies	Bioinformatics/genomics (Secreted Protein Discovery Initiative, SPDI), phage display, structure-based design
Maxygen	Therapeutic proteins, vaccines	Directed evolution, PEGylation, glycosylation (Molecular Breeding™, Family Shuffling™)
Novozymes	Industrial enzymes	Directed evolution, combinatorial protein engineering/phage display, site-directed mutagenesis
Protein Design Laboratories	Humanized therapeutic antibodies	Rational structure-based design, antibody humanization
Roche	Therapeutic proteins and antibodies	High-throughput cloning and expression, site-directed mutagenesis
Xencor	Therapeutic proteins and antibodies	Rational structure-based design, rational PEGylation (Protein Design Automation®, ImmunoPDA™, XmAb™, and Rational PEGylation™ technologies)

Improved properties can be achieved by modifying the protein's primary structure through sequence changes, or by incorporating chemical or post-translational modifications such as PEGylation (the attachment of polyethylene glycol) or glycosylation. The oligomerization state of the protein can also be altered, or the protein can be fused to other entities such as albumin or the Fc region of antibodies [4,5,7]. Strategies for implementing these changes include site-directed mutagenesis, random mutagenesis, recombination and other directed evolution methods, as well as rational protein design and structure-based computational approaches [5,8–10].

Many pharmaceutical and biotechnology companies are now using these and other strategies to generate large protein libraries and are taking advantage of HTS technology to rapidly assay these libraries for desired properties (Table 1). This creates the need to link large numbers of variant protein sequences with assay data. Ideally, one must analyze the results efficiently and display them so that patterns, particularly sequence–activity relationships, can be identified. Although database systems with HTS data management and structure–activity relationship (SAR) analytical capabilities are commercially available, they are specifically designed to handle small-molecule chemical and biological data; they do not support the basic bioinformatics functions that proteins require, such as sequence registration and comparison. There is no integrated and efficient data management system available that provides all the capabilities required for protein sequence data.

One solution is to adapt an existing data management system to handle high-volume protein sequence data. In this article, we describe an implementation in which a widely used small-molecule screening, data management, and SAR analysis system has been adapted to serve the unique needs of protein drug discovery and development.

Data management requirements for protein drug discovery

In addition to linking protein sequences with HTS data and providing powerful analysis tools, a useful data management system for proteins should support graphics as well as numeric and string-based data types, enable connections between various sources of data, and allow efficient querying and reporting. To be effective, the system must not only include the required infrastructure and