Bioinformatics: a holistic approach to drug discovery

'The vision of the near future is of bioinformatics bridging the gap between traditionally diverse disciplines.'

Scott Kahn, General Manager and Senior Vice President – Life Sciences, Accelrys

One of the most difficult challenges for the pharmaceutical industry is to transform the mountains of complex biological data into workable knowledge. Pharmaceutical and biotechnology companies believe that somewhere in this biological information lie the answers to questions that could prevent and cure disease. It is a matter of piecing together the clues in the hopes of using the deluge of this information to develop new, more effective drugs directed against as yet undiscovered functional targets. Arguably, this integration across the formal disciplines of biology and chemistry – involving the access and understanding of the raw data, the resultant information, and the derived knowledge – is the aspect that is both the most challenging and the most ripe for breakthrough advances.

A major challenge has resulted from the trend for research teams to develop proprietary bioinformatics solutions to suit their particular research endeavors, and this has resulted in a scattering of niche computational solutions with few current standards in the field and little or no compatibility [1–3]. Piecing together these diverse technologies into a streamlined and fully integrated drug discovery process is a formidable task, requiring a unification of what traditionally have been separate R&D efforts. The ability to establish and dynamically maintain the many relationships between the disparate content sources in an automated fashion will be key; human curation is stretched beyond its capacity even when only existing content is taken into consideration. Furthermore, one must not overlook the broader integration needed between biology and chemistry.

Overall, the key challenges facing bioinformatics, and indeed modern biology, are to harness the mass of accumulating data and use it to create new understandings and new knowledge. The long-term value of bioinformatics lies not in the tools developed and used, but in the conversion of data into the practical knowledge required for the delivery of better therapeutics [4]. Success has profound implications for drug discovery by improving overall productivity at various points throughout the drug discovery process and its constituent sciences. The vision of the near future is of bioinformatics bridging the gap between traditionally diverse disciplines. With this in mind, pharmaceutical companies are taking great strides to make bioinformatics as standardized and easy to use as possible, foreseeing a day when a bioinformatics platform will be as commonplace as computer operating systems.

Benefits of transforming data into knowledge

Bioinformatics methods have already enhanced the discovery process in significant ways. The Human Genome Project and dozens of other genome sequencing projects have spurred target discovery and identification. Whole genome sequence comparison, functional annotation, and advances in automated methods in three-dimensional (3D) structure determination are uncovering usable targets for drug discovery faster and more reliably than ever.

Related to bioinformatics, and faced with the similar data integration issues, is lead identification and optimization (LIDO). LIDO has been greatly enhanced by advances in virtual screening and cheminformatics, which increase its efficiency by focusing in on compounds for synthesis, reducing the amount of assays, increasing the parallelization of screening steps, and optimizing the screening process altogether. ADME/Tox prediction models have further helped by enabling researchers to quickly identify compounds destined to fail in clinical or preclinical trials. By estimating the toxicological properties, ADME/Tox prediction significantly reduces drug discovery costs without increasing development and evaluation time for successful candidates. In essence, it enables researchers to 'fail faster' so that they can focus on those compounds that ultimately pass regulatory approval.

Viewed individually, each area has its own set of tools, databases and standards. These bio-niches can no longer be viewed in isolation as separate entities – efficiency demands an integrated drug discovery solution and a more global view of the data, the information, and the processes by which drug discovery and optimization take place.

1359-6446/02/$ – see front matter PII: S1359-6446(02)02295-X

Ideally, integrated drug discovery solutions will combine leading-edge technologies such as genomics, HTS, structure-based drug design and cheminformatics into a robust information management system. This holistic approach to drug discovery understands the flow of data, information and knowledge between the technologies of the different disciplines they serve, realizes their interdependence, and incorporates them to guide the decision-making process.

Bioinformatics incorporated in such a way would enable researchers to link, for example, chemical information with target active sites and substrate specificity. Chemical information is of little use if chemists doing drug design cannot access information on targets and the biological and structural assumptions associated with binding and activity. By integrating bioinformatics with biological screening and cheminformatics, scientists will be provided with the technologies to allow them to streamline and maximize the speed and efficiency of the design, discovery and optimization of lead compounds [5].

Meeting IT requirements in a bioinformatics solution

The key to an effective enterprise-wide bioinformatics solution is developing the necessary infrastructure and mining tools for biological data, while supporting the needs of life scientists performing discovery research. For the data to be useful, researchers must know where information resides and how it can be accessed and used to enhance productivity and add value, not only to current practices but to meet the evolving needs of life science research.

Relational database technology has gained widespread use because of its flexibility, security and expandability. It provides a unified infrastructure with which to store and access data from standard development tools such as Microsoft® VisualTools, Open Database Connectivity (ODBC), Java, Java Database Connectivity (JDBC) and Oracle Objects for OLE [6].

To be of value to the discovery scientist, once a database schema, or map of how and where data are stored in a database, is established, in-house data can easily be incorporated into the database and can also form connections between the various different sources of data. In the case of DNA or protein sequences, standard relational database tools centralize and maintain sequence data, along with supporting data such as structural and functional annotation, clusters, alignments and reference information. Acting as a centralized repository of information, integrated data management links compound, textual, proteomic, genomic, clinical, toxicology, gene expression and other data sources. Such a scenario envisages bioinformatics expanding to the desktop, where everyone in a company can access data and everyone will have access to scientific content to enable management, analysis, and communication of all information.

Time and cost savings

Although faced with the daunting task of keeping pace with the proliferation of data and integrating with in-house biological and chemical data, pharmaceutical companies need to realize the long-term implications that bioinformatics will have on drug discovery. Analysts predict that bioinformatics could help cut the cost of developing a drug by 50% and prune as many as three years off its time to development [7]. The economic opportunity created by reducing drug discovery costs and time to market has implications that will revolutionize the drug discovery industry as we know it.

Bioinformatics offers the pharmaceutical industry a huge saving in the cost of drug discovery. With the cost of drug development tipping the scales at US$880 million per drug, and taking as long as 15 years to take a drug to market [1,7], drug companies are eager to streamline their discovery and development pipelines. With an estimated 75% of the cost resulting from failures along the way, bioinformatics aims to reduce this failure rate.

Individually, the different areas of drug discovery have benefitted from bioinformatics approaches. Data integration of the different research disciplines involved in drug discovery will allow scientists to begin to see a more streamlined and efficient discovery process. The overall effect of these will be to speed up R&D through better, more globally aware decision-making. Bioinformatics can facilitate the incorporation of larger amounts of better quality information early on, enabling them to make better decisions on targets and lead compounds. Ultimately, this means higher success rates and faster drugs to market.

References

1 Overby, S. (2001) Drug companies on speed. *CIO Magazine* 15 October 2002 issue (available online at: http://www.cio.com/archive/101501/speed.html)

2 Saviotti, P.P. *et al.* (2000) The changing marketplace of bioinformatics. *Nat. Biotechnol.* 18, 1247–1249

3 Silico Research (2001) Data integration in biopharmaceutical research and development processes. *Silico Research Data Integration Insight Paper* July 2001 (available online at: http://www-3.ibm.com/solutions/lifesciences/pdf/Silico.PDF)

4 Persidis, A. (1999) Bioinformatics. *Nat. Biotechnol.* 17, 828–830

5 Larvol, B.L. and Wilkerson, L.J. (1998) *In silico* drug discovery: Tools for bridging the NCE gap. *Nat. Biotechnol.* 16 (Suppl.), 33–34

6 Hudson, P. (2001) External solutions help organisations meet bioinformatics needs. *Eur. Biopharmaceut. Rev.* 96–98

7 Flanagan, A. *et al.* (2001) A Revolution in R&D: How genomics and genetics are transforming the bipharmaceutical industry. *Boston Consulting Group Report* 26 November, 1–64 (available online at: http://www.bcg.com/publications/files/eng_genomicsgenetics_rep_11_01.pdf)

Scott Kahn

General Manager and Senior Vice President, Life Sciences

Accelrys, San Diego, CA 92121, USA

tel: +1 619 546 5310

fax: +1 619 458 0136

e-mail: scott@accelrys.com