Maximizing the value of bioinformatics in drug discovery

Bioinformatics, who needs it? 'Everyone in the drug discovery business' was the consensus at a recent conference held in London to discuss ways to maximize the value of bioinformatics in drug discovery. The amount of biomedical information available to scientists is increasing exponentially; a new DNA sequence is submitted to the European Molecular Biology Laboratory (EMBL) web server every minute, a new chemical structure is published every minute and a new patent is filed every two minutes. Computing technology is rising to meet the challenge; for example, using algorithmic software from the EMBL (Heidelberg, Germany), a Silicon Graphics (Mt View, CA, USA) supercomputer array has just completed an analysis of the entire yeast genome in 72 h, pinpointing the function of over 5,000 proteins and publishing the results in real time on the World Wide Web (WWW).

Pharmaceutical companies, particularly those with large discovery programmes, simply cannot afford to ignore the impact of this explosion of technology and information. Tough decisions must be made about how to react to it in a rational way. This meeting, the second in a series on bioinformatics, focused upon technological aspects of data storage, access and analysis as well as the logistics and strategies companies might adopt in integrating bioinformatics into their drug discovery programmes.

Dollar value of bioinformatics – benchmarking

In a climate of rising R&D costs and commercial and regulatory pressure, how does a company assess whether or not its investment in bioinformatics is appropriate? Dr William Bains (PA Consulting Group, Melbourn, UK) proposed benchmarking as a tool to support good decision-making, according to the individual needs of the company.

Benchmarking is the identification and quantification of best practice, through comparison of in-house processes with those of other organizations. Drug discovery is a technology in constant flux, and competitive advantage relies upon how the various components of the technology are combined; benchmarking these components in a forward-looking manner can ensure that the company concerned becomes a 'leader' and not a 'follower'.

Process-modelling strategies to reduce and predict the costs of target discovery were described. For instance, the possible routes in the discovery of genetic markers – 'data mining' of existing databases, genetics and high-throughput sequencing – can be quantitatively compared to decide which is the most appropriate. Bioinformatics as a 'bolt-on extra' is wasting valuable data and missing opportunities for discovery.

Presenting bioinformatics to its best advantage

Dr Steve Gardner (Oxford Molecular Group, UK) addressed the issue of the interface between bioinformatics and potential users. Making the resources available to bench scientists in the most productive way involves bringing together methods, algorithms and databases in an interoperable environment – a user friendly package, which everyone can use on a variety of platforms. There are currently several cross-platform development tools that enable this kind of integration, Java and JavaScript (Sun Microsystems Laboratories, Mt View, CA, USA) being the most talked about, offering unprecedented interactivity in its scripting facilities.

'Comms Manager', due to be launched as a world-wide development standard in October by Oxford Molecular (Oxford, UK), was developed with this kind of WWW technology very much in mind. Featuring cross-platform, network-aware CAMD architecture, it is compatible with Java, HTML and CORBA, and benefits from a standards editorial board from major development partners. Designed to be used in private networks (intranets), databases like Iditis can be quizzed effectively by anyone from the most inexperienced to the expert, and the results can be routed into other applications for manipulation with ease. Personal OMIGA, Oxford Molecular's personal bioinformatics workbench, provides a similar interface, bringing together many of the tools required for sequence searching, alignment, and editing in a single package.

Internet as an external resource

As well as acting as a laboratory for the development of technological resources and effective interfaces with them, the Internet (especially the WWW) is an increasingly valuable source of biomedical information, as demonstrated by Dr Tony Parsons (Pfizer, Sandwich, UK). In his discussion of search strategies for information retrieval, attendees were provided with a fascinating list of URLs, newsgroup listservers and ftp sites to illustrate the utility of these external resources in drug/target discovery (see Box 1).

Examples included were the electronic journal *Emerging Infectious Diseases*, which can be accessed via the WWW, ftp and by e-mail through a listserver, numerous sites on Internet research, and Entrez, the unique document retrieval system of the National Center for Biotechnology Information (NCBI). From this WWW site, it is possible to conduct integrated searches of MEDLINE and all of the publicly available protein and nucleotide databases. In addition, a useful facility provided by Entrez makes it easy to examine 'neighbouring' sequence data, together with related literature references. From its inception in 1992, the bimonthly releases of Entrez on CD ROM have grown from one to five CDs, making it less popular than the Net and Web versions, which are updated daily, contain new databases and provide more full-text articles from on-line journals. However,

 1359-6446/96/$15.00

Box 1. Some URLs relating to drug/target discovery

- *The Virtual School of Molecular Sciences* – a 'virtual course' in structure based drug design: http://www.venus.co.uk/vsms
- *SWISS-PROT* – an indispensable tool for molecular biologists: http://expasy.hcuge.ch
- *Emerging Infectious Diseases* – the on-line journal: http://www.cdc.gov/ncidod/EID/eid.html
- Silicon Graphics/EMBL yeast genome analysis: http://genecrunch.sgi.com/
- NCBI's Entrez: http://www3.ncbi.nlm.nih.gov/Entrez/index.html
- EMBL: http://www.embl-heidelberg.de/
  For information about 'Syncron' tools for maintaining synchronized copies of the EMBL databases, e-mail: ditommaso@ebi.ac.uk
- For information about intranets:
  Forrester Research: http://www.forrester.com/
  Intranet Journal: http://www.brill.com/intranet/
  The Intranet exchange: http://www.brill.com/intranet/ijx/index.html

without an in-house database a pharmaceutical company would find itself sending out proprietary sequence data onto the Internet – commercially sensitive information, to say the least. Maintaining in-house copies of the public domain databases is expensive, even with automated cumulative updates, and the possibility of preserving confidentiality by submitting queries in encrypted form is under discussion among corporations and database administrations, but there are problems: the US government treats encryption tools as weapons, and applies the same legislative restrictions on their export.

Intranets

Intranets use Internet technologies within a company, deployed on an internal network of nodes. This allows the centralization of information, which can be organized and shared to save costs and time. It is relatively inexpensive; Eli Lilly is in the process of linking 16,000 of its 26,000 employees on an intranet, and the 3,000 nodes installed to date have cost just $80,000. Nearly one in four companies already uses an internal web-server, and the business will 'hit $1 billion by the year 2000' (Forrester Research; see Box 1 for URL). Protecting these intranets from uninvited guests is the subject of much discussion.

Managing the data from combichem and HTS

The goals of combinatorial chemistry and high-throughput screening (HTS) are mutually to accelerate lead compound generation and optimization. That both of these are practical technologies is the result of progress in robotics and miniaturization, resulting in high material throughput, which in turn leads to greater demands on systems of data storage and management. Eight robots processing 10,000 samples per day, five days per week for a year will generate about 20 million data points.

Dr Stephen Welford (MDL Information Systems Inc, San Leandro, USA) elaborated on the bioinformatics capabilities required in their planning, automation, screening and interpretation phases. Combinatorial peptide libraries can contain mixtures of 10^3–10^6 compounds. Problems with the identification and resynthesis of compounds of interest (and, at least in the pharmaceutical industry, with finding nonpeptide alternatives with equivalent activity) have led to the increasing use of smaller, discrete compound libraries, especially in lead optimization. Planning libraries might involve the generation of a so-called 'virtual library' on the basis of the selected strategy and building blocks, enabling the calculation of metrics such as target population diversity and physicochemical properties. An emerging research area is the development of 'virtual assay' tools, which lead to the prediction of probable hits on the basis of such a virtual library.

The successful management of library structural data requires a general database model and associated software tools to support the different library types and synthesis strategies. Easy-to-use desktop software is appropriate for managing project-level data, whereas client-server applications are needed for group and corporate projects, where data from several different centralized sources is required. MDL's combinatorial chemistry synthesis (CCS) software includes Project Library, a desktop application which meets the requirements of the former, and they are launching Central Library, a client-server package providing centralized data management support, in the autumn. MDL SCREEN supports HTS data management, and they are aiming to integrate this software with Central Library using their ISIS technology.

Genecrunch

A powerful example of what progress in computing and bioinformatics technology can achieve was demonstrated by Dr Juli Nash (Silicon Graphics Computer Systems Inc., Mt View, USA). GeneQuiz is an application developed by collaborative research teams at the EMBL institute (Heidelberg, Germany), its European Bioinformatics Institute outstation in Cambridge, UK, and former EMBL scientists at research centres in Spain, Germany and the USA. It aims to predict the functions of genes and has already been used to perform the fully automated analysis of the complete genomes of *Mycoplasma genitalium*, *Haemophilus influenzae* and, recently, the yeast *Saccharomyces cerevisiae*. Running a Silicon Graphics POWER CHALLENGE array of 64 processors in parallel, GeneQuiz identified and analysed more than 6,000 protein sequences from 20 Gb of raw data (12.5 million bp) in 72 h, the results being made available on the WWW in real time. This analysis of the

most complex genome sequenced to date is probably the most comprehensive and computationally advanced yet undertaken, and will be of direct benefit to those working on the Human Genome Project. Discoveries about genes lead ultimately to new diagnostic tools, new targets and new drugs – 'in silico' genomics is of obvious importance in drug discovery.

Dr Nash also discussed the way that new web-based tools, such as Java, convenient interfaces and new visualization technologies, such as Silicon Graphics' recently released 'data mining' application MineSet, improve the communication of data throughout international networks and across disciplines.

A concluding presentation given by Prof. Rod Hubbard (University of York, UK) highlighted the challenges to bioinformatics for the future, and lively discussions ensued about learning neural nets as 'sophisticated search engines', for example, to search 3D databases, and the use of virtual reality technology in molecular modelling.

The next bioinformatics meeting in this series takes place in October; all enquiries should be addressed to: Nicola McCall, Henry Stewart Conference Studies, Russell House, 28/30 Little Russell St, London, UK WC1A 2HN, tel: + 44 171 404 3040, e-mail: 100622.3263@compuserve.com.

Matthew Thorne

Expansion at Oxford Asymmetry

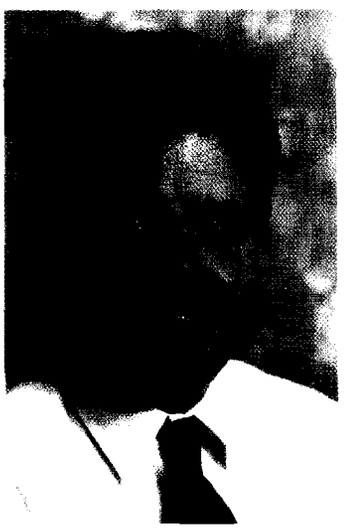
At the end of March, Oxford Asymmetry celebrated the opening of their new facilities at Milton Park in Oxfordshire by UK Deputy Prime Minister Michael Heseltine. The company has grown very rapidly and now claims to offer complete solutions to chemical needs in the pharmaceutical industry, from library generation through to commercial-scale production. Oxford Asymmetry became profitable last year, only the fourth year of operation, with a turnover of £2.8 million, and turnover is expected to double in 1996. The company sprang from the research group of Dr Stephen Davies at the Dyson Perrins Laboratories of Oxford University. Davies and a team of thirty researchers were excelling in the field of chiral chemistry.

Dr Nick Cross, Chairman of Oxford Asymmetry, believes that "the biggest single milestone in the company's development" has been the recruitment of Dr Edwin Moses, whose international credentials are already established through experience at Raggio-Italgene SpA in Italy and Enzymatix in the UK. When Moses joined as Managing Director in July 1993, turnover was only around £100,000 and the staff numbered less than ten.

Moses, in turn, points to another milestone: the deal with Pfizer in early 1995. Pfizer pumped £3.6 million into Oxford Asymmetry in a two-year deal tied in with the creation of a subsidiary called Oxford Diversity. The company is developing additional enabling technology in solid-phase synthesis and producing combinatorial compound libraries for its large partner. It will retain ownership of the enabling technology, which it will license to Pfizer, and has the right, subject to a royalty, to exploit the technology derived from the association to generate small quantities of compounds for screening and analogues for lead optimization. Oxford Diversity is headed by Dr Tony Baxter, formerly of Glaxo and Ciba-Geigy.

Moses believes that Pfizer's approach to the deal has been exemplary, offering Oxford Asymmetry scientists access to Pfizer's combinatorial experience in a very open way so that, in effect, their learning curve has been eliminated. Oxford Asymmetry not only benefits from Pfizer's technical experience, but according to Moses "a major collaboration with such a progressive company confers prestige and credibility on Oxford Asymmetry that must assist in generating new business". Pfizer are also pleased with the way the collaboration is developing. Dr Nick Terrett of Pfizer's New Leads Chemistry Group describes Oxford Diversity as "a young and innovative group, producing new solid-phase technology that is proving valuable in our drug discovery programmes".

At the other end of the scale, Oxford Asymmetry established another major deal with the US giant Cambrex at the end of 1995. This association has enabled the company to complete its range of services to the industry, from small volume library development up to multi-ton manufacture. Cambrex operate from seven centres within the USA and three in Europe. The US presence has further enabled the company to interact with new US partners (some 50% of Oxford Asymmetry business is derived from partnerships with US and Japanese companies). Cambrex will also underwrite $1 million of the company's R&D annually for the next 3 years. Cambrex, in return, has a partner that can provide flexible and responsive support for early development of its chemical technology.



Dr Edwin Moses, Managing Director of Oxford Asymmetry.