logistics, and robotics/automation as the key issues from which to derive their HTS infrastructure, and in each of these areas, support teams were set up.

The group were able to report dramatic increases in throughput after one year, although there were areas where automation did not offer the complete solution. Early experience determined that the SPA-, ELISA- and photometry-based screens could readily be adapted to automated formats. Other assays, such as those based on electrophoresis or morphological change required manual operation, although this could be enhanced through the use of semi-automated instrumentation, such as 96-channel pipetters. The extent to which assays are now automated depends on assay type and ranges from zero to 100%, but overall improvements in productivity of up to fivefold have been observed, with a consequent increase in output of lead compounds. One problem, however, has been that screening biologists have been slow to 'buy in' to new systems and working practices.

Summary

The above is, by necessity, a very selective discussion of presentations; there were many interesting talks, lively panel sessions and workshops in the discovery field and also an equal number of (parallel) sessions for those interested in the automation of procedures associated with testing and analysis in pharmaceutical development. It was also possible to visit the headquarters of the organizers, Zymark, to view some examples of their own products in action. *ISLAR '97* will take place in Boston on 19–22 October; full details of the meeting and the Proceedings of *ISLAR '96* are available from Christine O'Neil, Zymark Corporation, Zymark Center, Hopkinton, MA 10748, USA, tel: +1 508 435 9500 (ext. 2224), fax: +1 508 435 3439, e-mail: islar@islar.com, web site: http://www.islar.com.

David Hughes

Advanced technologies for information extraction

IBC's second European forum for IT specialists and drug discovery scientists was held in London in November. Entitled *Advanced Technologies for Information Extraction: Analysis of Complex Data Sets to Aid Drug Discovery*, the two-day meeting was a largely stimulating attempt to approach the problem of extraction, handling and optimizing the use of information from three angles:

- development of computing solutions, specifically aimed at the pharmaceutical industry;
- existing techniques employed by the pharmaceutical companies themselves; and
- commercial and academic uses of advanced technologies outside the pharmaceutical industry, and their possible application in drug discovery.

Representatives from these three areas gave examples of the application of emerging techniques that combine computing methods, such as neuronetworking and fuzzy logic ('neurofuzzy' technology), with expert systems to quiz very large databases, achieving the object of obtaining useful and easily visualized information. These can provide a company with the competitive edge, be it in lead discovery, patenting strategy or process control.

It's easy if you try

Dr William Bains (Merlin Ventures, London, UK) opened the proceedings by expounding his view that the 'problem' faced by companies in making bioinformatics a high-value component of their drug discovery programmes is not one of too much data, or acquiring the means with which to visualize and analyse it, but essentially one of 'science and scientists'. Because of the increase in automation, the data generated by established drug discovery techniques is increasing at an exponential rate. Newer technologies, such as genomics and combinatorial chemistry, have created a new class of data with its own inherent problems, and emerging and future technologies, such as proteome, functional genetics and ultra-HTS, will generate gigabytes of data per week.

Engineering approach

Dr Bains estimates that, by the year 2016, companies will be searching public-domain databases of the order of 6.5 terabytes for information, but points out that data sets of this size, and larger, are already being dealt with productively by organizations such as the super/hypermarket chain J. Sainsbury's (transaction database) and the Hubble space telescope team (already a 25 terabyte data set), using robust data mining and analysis software. Dr Bains also argues that browsing and sifting this information should not be a problem; current and developing web-based tools offer sophisticated and interactive interfaces with such databases, and they are becoming increasingly interoperable and powerful. He proposes that, for data-rich subjects, biological research must adopt an engineering approach to bioinformatics, instead of remaining 'largely a cottage industry, with data generation being carried out in an intellectually idiosyncratic and geographically distributed mode' (see Ref. 1).

Pharmaceutical companies must look at analogous data problems outside their industry for robust, proven solutions, and apply these in a rational way to their bioinformatics programmes.

Let us do it for you

The acquisition earlier this year of Illustra Information Technologies by Informix (Menlo Park, CA, USA) has led the second largest database-provider in the world to develop just such an IT-based software package, aimed specifically at drug discovery research. Dr Tim Hampson (Informix, Ashford, UK) outlined their Universal Server project, a database expected to be available from December 1996, combining the best features of relational, and object-oriented massively-parallel technologies. Described as 'object relational', Illustra's database has been fused with Informix's fast and scaleable relational database 'Online', together with the 'Metacube' control system. Now in Beta test, Universal Server allows complex querying via an SQL-based logic layer, with all the navigational benefits of object-orientation, and scaleability from a laptop PC to the largest IBM SP2 massively-parallel processor in the world.

Java-based tools

Tripos Inc. (St Louis, MO, USA) have sold molecular modelling software since 1979, and state their goal as 'how to deploy applications for drug discovery, so that multi-functional and multi-site teams can access data and applications securely, easily, and with little or no additional training, so that users can focus on their work and not the computer'. Mr Paul Weber demonstrated their Java-based tools, all interfaced using thin (i.e. simple), 'Webtop' clients. These include molecular projectors and a sophisticated genetic-algorithm-based chemical library query system (GASP v1.0). Tripos have learnt to avoid 'bloatware' with over-rich function sets, and prefer to produce task-orientated applications, giving the user the option to increase functionality by downloading specific Java applets, as and when required. Their philosophy, in part, springs from an experiment: 'Sketch & Fetch', a query applet for chemical libraries, which was released on the WWW and received a favourable response. They have since developed this application to enable 2D to 3D building. Mr Weber is convinced that Java is here to stay, being specifically designed for network use, and providing the benefits of platform independence, ease of distribution and development, and lower costs. Interestingly, Tripos intend shortly to combine physical analysis with their computing expertise, and become a biotech company.

But we're doing it already

Mr Julian Abery (Xenova, Slough, UK) gave a comprehensive description of RAPIDD, a custom-built research aid to perform intelligent drug discovery with access to Xenova's proprietary natural products database via a host of analysis tools and on-line searching facilities. RAPIDD integrates data from various functions:

- culture collection, with automated organism tracking and plate generation, and an interface with the fermentation control system;
- screening, with lead development, validation, results analysis, data query and presentation; and
- chemistry, with sophisticated analytical tools to examine spectroscopic, organism and structural data.

The company seeks to integrate their automated processes of sample picking and plate generation with data capture and analysis in a way that will allow them to incorporate emerging technologies, such as advanced pattern analysis with neural networks. Mr Abery argued that their system provides Xenova with a competitive edge – they currently have 14 lead optimization programmes in progress.

Pharmacophore mapping

A pharmacophore is a hypothesis as to the 3D features recognized by a macromolecule of interest. Pharmacophore mapping seeks to discover the common 3D patterns in structurally diverse molecules that act at the same macromolecular binding site. This diversity can be crucial in drug discovery: it is relatively easy to obtain new patents for molecules that are diverse in 2D structure, yet possess a 3D similarity to known ligands. Dr Yvonne C. Martin (Abbott Laboratories, Abbott Park, IL, USA) gave examples of the methods used by Abbott in the pharmacophore mapping of dopamine D_1, D_2 and nicotinic agonists. Using DISCO, a programme that allows them to arrive at several different pharmacophore hypotheses, they generate the low energy conformations of all potent molecules selected. Points of superposition are then defined, and matching points and conformers are then quickly identified using a clique-detection (maximum common substructure) algorithm. The programme allows for flexibility of the macromolecular binding site (proteins often move when binding to a ligand), considers several conformers per molecule, and can be tailored by adding various constraints (for example, 'if the number of superposition points >3, then check chirality').

Dr Hans Briem (Boehringer Ingelheim KG, Ingelheim, Germany) described the Daylight (Daylight CIS Inc., Irvine, CA, USA) method for investigation of 2D molecular similarity in libraries. A binary fingerprint is generated for each molecule, based on all the possible bond paths through it, and then a Tanimoto similarity coefficient is calculated to describe similarity between any pair. 3D similarity can be assessed using DOCK 3.5, which allows the rigid docking of ligands from a 3D database into a given protein binding site. DOCK is freely available to academic institutions and is licensed with a charge to industrial centres (contact: Dr P. McCloskey, Molecular Design Institute, UCSF, CA, USA, e-mail: mcclosk@cgl.ucsf.edu). It works by filling a virtual representation of the receptor cavity with spheres, and matching atom distances in the ligands with the sphere centre distances, scoring the results in terms of contact points, as well as electrostatic and steric interactions, against a reference panel of proteins. Dr Briem says that they have combined the DOCK and Daylight approaches to 3D and 2D molecular similarity, respectively, generating two new lead compounds. Flexible docking strategies will further enhance the system.

We've been doing it for years

Mr Colin Bravington was involved in process control software and systems integration for a number of years before

re-entering academia to manage the Industrial Control Centre at Westminster University (London, UK). This group works closely with industry, providing services such as technical support and feasibility studies in diverse applications. They have applied two key techniques of data analysis and modelling/simulation to complex bioprocesses, such as fermentation control. By combining expert system-based process knowledge with pattern recognition and feature extraction techniques, neural networks can be iteratively trained to provide accurate, fast predictions at all the crucial stages in the fermentation process, facilitating more informed control decisions. The pattern recognition techniques of principal components analysis (PCA) and Kohonen self-organizing maps (SOM) have valuable applications in the analysis and visualization of data from very large data sets, and Mr Bravington showed how these have been applied in drug discovery by Xenova. The patterns of pharmacological activities in extracts prepared from 28 streptomycete organisms were compared with their classification using a probabilistic identification matrix, and the analysis of activity spectra enabled greater than 71% probability of correct classification. Pattern recognition techniques such as PCA and SOM have a considerable potential in natural product screening programmes.

Neurofuzzy systems

Neurofuzzy systems combine associative memory networks with fuzzy logic to produce transparent models, whose behaviour can be analysed mathematically and trained using simple, well-established linear optimization techniques. Professor Chris J. Harris (Image, Speech and Intelligent Systems [ISIS] Research Group, University of Southampton, UK) gave an introduction to this technology, with examples of how ISIS have applied it to such diverse applications as a car-parking problem, gas turbine control, navigation of an autonomous submarine, and HELIOS, a multi-sensor data fusion project that allows a helicopter to avoid obstacles in conditions of zero visibility. These 'grey box input' learning systems are already of relevance to drug discovery scientists – for instance, in allowing a degree of flexibility in querying a large database, and in the recognition of patterns in the data retrieved. Mr Bill Edisbury (NPower Ltd, Manchester, UK) demonstrated the power of hybrid artificial intelligence tools (combining fuzzy logic, genetic algorithms, case-based reasoning and expert systems) in mining large databases. A particularly stunning display of the technology was British Telecom's advanced system for visualizing call data.

A post-conference workshop provided some practical experience on the use of some of the advanced analysis techniques. Information on future IBC meetings can be obtained from: Jessica Robertson, IBC UK Conferences Ltd, Gilmoora House, 57–61 Mortimer Street, London, UK W1N 8JX. tel: +44 (0)171 637 4383, fax: +44 (0)171 631 3214.

Matthew Thorne
tel: +44 1206 230645
e-mail: thorne@masha.demon.co.uk

REFERENCE

1 Benton, D. (1996) *Trends Biotechnol.* 14, 273–279

In short...

Quintiles recently became the world's largest contract research and health care company by acquiring **Innovex**, a provider of global sales and marketing support for pharmaceutical companies. As a result, Quintiles will now be able to provide a full range of health care services, including product development, health care consulting, and sales and marketing expertise. A week before the Innovex deal, Quintiles completed the acquisition of BRI international, a leading medical device CRO.