InfoTech Pharma '98: managing the knowledge

Whereas last year's InfoTech Pharma focused on the IT infrastructures that pharmaceutical companies require to be competitive, it was implicitly understood that most, if not all, of the companies represented this year had taken this on board and were looking at ways of deploying it. 'Knowledge management' was very much the buzzword at the InfoTech Pharma '98 conference and exhibition, held in January this year at the Mayfair Intercontinental Hotel in London.

Organized into six 'streams', as last year, the event allows delegates to follow their specific interests, even jumping between streams to catch sessions of relevance. The broad topic areas were: maximizing the effectiveness of IT in R&D; IT in sales and marketing; bioinformatics; IT in product registration; IT in drug discovery; and the strategic use of the Internet. There were also plenty of companies exhibiting software solutions, as well as evening workshops.

In one of the keynote addresses, Michael Heim (Eli Lilly, Indianapolis, IN, USA) outlined some challenges that a successful IT strategy must address to facilitate knowledge-management-driven R&D. In fact, 'Increasing the corporate IQ' by making knowledge easily available, responsive and well-curated was a theme of the conference generally. Heim maintains that 'in this business – driven by intellectual capital – you will know who successfully implemented knowledge management by attending InfoTech Pharma 2008; they will be the only ones around'.

'It's only a few Gigabytes' – Chris Jones (CERN) on the human genome

Another keynote speech came from Prof. Rolf Krebs (Boehringer Ingelheim, Germany). The mere presence of such a senior figure in the industry at the conference signifies the importance that senior management are attaching to what used to be just 'the IT function'. That function is now recognized as a key driver, and Prof. Krebs gave a broad outline of developments he expects to see, including third-party involvement ('we used to avoid research collaborations: not any more'), and sharing his vision that disease prevention, rather than treatment, will be the priority in the future.

As might be expected, there were common themes behind the IT infrastructures that two major pharmaceutical companies have set up to meet the challenges of the present and future, but there were also differences in their approach and implementation.

Zeneca

With the issue of integration very much in mind, Dr Neil Stutchbury (Zeneca, Alderley Edge, UK) described how Zeneca has now installed a standard Windows 95/Netscape interface, company-wide, that is customizable to meet the individual user's needs and regular uses. There was some interest as to the choice of Windows 95, rather than NT (cf. Novartis), and worries that setting up such a 'monoculture' might make the infrastructure prone to failure, much like a crop disease. Dr Stutchbury replied that the security offered by NT was not needed internally, and that the integration of some of Zeneca's old 16-bit legacy systems was more suited to the architecture of Windows 95, which also offers more to mobile users. They do not consider system failure caused by 'malicious data' (i.e. viruses) to be a problem – 'we can eradicate a macrovirus in a couple of days'. Surprisingly, the firewall is not virus-protected, for fears that this would slow down such a frantically busy gateway.

Bioinformatics at Zeneca has moved on. Nucleotide- and protein-sequence searching is now performed in a structured way (again all from the same Web-browser interface). In what seems to be standard practice now, they 'cull' the public-domain databases overnight; Zeneca's scientists then search these internally, with proactive alerting of new gene sequences, and without fear of prying eyes. In combination with their partnerships with commercial companies like Incyte, Zeneca's genomics capability is formidable.

'My group searches a thousand genes every day; good luck to whoever is listening in!' – Lee J. Beely, (SmithKline Beecham) on security problems with the Web

Presenting a user's, rather than an IT expert's point of view, Dr John Major gave a presentation outlining the impact of IT on Zeneca's HTS system architecture and processing technology. They have shortened the HTS phase of their drug discovery projects from an average of six to three months, partly because of the installation of a 'haystack' compound dispensary.

In the areas of document handling and prospective deals, Dr Stutchbury illustrated CONCERT, a TCP/IP WAN technology that uses 'good information management practice' to facilitate the sharing of resources between teams in Japan, the USA and the UK. All the key people had access to the same set of documents, created in Lotus Notes and viewed with a Web browser, and this resulted in 'better, more-informed decision making at all levels'.

It is evident that high-level commitment to the IT programme has transformed Zeneca. They have successfully implemented a complete reorganization of the business infrastructure, to include IT as a major component in many parts, not just R&D. It is also evident that the process is difficult, expensive and, also, unfinished.

1359-6446/98/$19.00. PII: S1359-6446(98)01183-0

Novartis

Dr Rene Ziegler (Novartis Pharma, Basel, Switzerland) outlined the critical intervention points for IT in Novartis' R&D. The 150 global discovery projects, 40 project teams in eight countries, researching at nine geographical sites and into seven different therapeutic areas presents quite a challenge. To meet it, the IT infrastructure is in place to 'integrate process, data and technology management'. Rather grandly titled, the 'Global Enterprise Workstation' gives everyone in R&D a standardized desktop client that is customized to their particular needs; it is still not clear whether everyone in Novartis discovery is actually using it.

'Physicists invented the Web, biologists are taking it over, and chemists have yet to discover it' – Andrew Lyall (Glaxo Wellcome)

Novartis uses firewalls to protect the corporate intranet, providing secure access to its 'extranet' for collaborators to exchange data, and Web access for those 'inside'. Identifying the many points at which IT intervenes in target identification, lead generation, optimization and early development, Dr Ziegler gave a few examples of innovative solutions, including WITCH (Windows input tool for chemistry), a Windows NT-based client using libraries from MDL that allows online compound registration by 300 users globally, with built-in quality and novelty checking and various sophisticated features.

IT funding at Novartis is from a global IT budget, divided and allocated per therapeutic area and spent locally, providing a stable finances structure for IT development that aims to provide ready-to-use tools for drug discovery, as well as medium-term development for data management and maintenance.

Genomics

The return from genomics investment has been slow in coming, and although its potential value in target identification and drug design in the discovery process is often mooted, a direct impact on commercial success is hard to demonstrate. Dr Lee J. Beeley (SmithKline Beecham, Harlow, UK) gave a fascinating account of the successful use of database mining for the identification of novel targets in the area of obesity and diabetes.

Using BLAST searches of the human dbEST (expressed sequence tag) databases, a novel uncoupling protein, similar to that found in the brown fat cells of mammals but specific to human muscle tissue, was identified. A similar BLAST search, use of the structure-prediction programme PHD and exon/intron analysis enabled SmithKline Beecham to patent an assay for proteins that mimic the action of the four-helix bundle cytokine leptin (a 'satiety factor'), and to clone and express its receptor. Further analyses determined that the human *OB R* gene produces another, related leptin receptor with good potential as a target for regulation of nutritional intake. Dr Beeley stressed the value of combining dry (analysis *in silico*) with wet (cloning *in vitro*) cycles in this successful project, and demonstrated a significant economy of effort gained by the electronic work done.

Limitations and challenges

Some of the limitations and challenges faced by bioinformatics were illustrated by Dr Freiderich Rippmann (Merck, Darmstadt, Germany). In regard to infrastructure, there is a 'massive' shortage of highly trained bioinformatics people and few available third-party software solutions (most of which were represented at the conference, see below). Storage of the vast amounts of data that bioinformatics requires ('two million basepairs a day') is not a problem; its curation and retrieval is. LIMB, the Listing of Molecular Biology Databases, currently has ~300 entries: a number that implies that maintaining up-to-date versions of all of them in-house is 'clearly impossible' (?), and proper guidelines to their restricted use need to be in place.

Currently, there are ~400 protein targets for drugs. Well-founded estimates suggest that there should be 3,000–10,000 targets available in the human genome [Drews, J. (1996) *Nat. Biotechnol.* 14, 1516–1518]. The problem for genomics is the selection of suitable targets from this abundance. Merck have largely funded the University of Washington's sequencing of more than 300,000 ESTs from various human tissues, the data from which has been deposited in the public domain. But the quality of the data is a problem: by quizzing EST databases you will not find every member of a well-characterized protein receptor family (such as the serotonin receptors), or new members of it. 'So, if you feel your depression research will benefit from a 15th or 16th serotonin receptor, do not look for it in EST data banks'.

'Everything is beta quality' – G. Baehr (Sun Microsystems), complaining about software not keeping up

A technology that is surely in its infancy, but growing fast, is the use of metabolism databases, which link a genome to the organism's metabolic pathway components. Dr Rippmann argued that what is urgently needed are databases that do the same for regulatory and signal-transduction pathways.

High-throughput bioinformatics

Undoubtedly impressive, if a little scary, Incyte Pharmaceuticals' genomics programme goes from strength to strength. Scott C. Clarke (Incyte, Palo Alto, CA, USA) described advances in sequencing technology and the computational analysis and annotation used to provide services to their 'genomics partners', including the 14 new major pharmaceutical companies signed up since the beginning of 1996. Over three million ESTs have been processed to date for their LifeSeq database, updates for which are distributed monthly to 58 customer sites. PathoSeq now contains 22 microbial and two fungal genomes, and Incyte

makes '30+ cDNA libraries per day'. The technology used to achieve this involves 68 ABI Prism 377 (96-lane) DNA sequencers running around the clock and, not surprisingly, massive computing power fron Sun, SGI and DEC. They also release the data into the public domain, providing annotation and helping to curate the GenBank EST database. Incyte's vision of the 'ultimate genomics database' was presented as a cube, made up of layers containing the genomes of increasingly complex organisms, the top (human) surface of which contained tissue- and disease-specific expression libraries. This 'cDNA closure' is cheerfully expected to be achieved 'by 2008'. By then, Incyte will probably sequence your grandmother's DNA for £9.99 + tax.

High-throughput screening and combinatorial chemistry

Dr Dora Schnur (Pharmacopeia, Princeton, NJ, USA) described the use of computer-aided drug design and diversity analysis of large (some of them very large) combinatorial libraries. Advanced data-visualization techniques are used in conjunction with sophisticated diversity metrics for quizzing libraries of between 10,000 and 150,000 compounds, with successes in lead identification for Pharmacopeia.

Neurogen's 'Accelerated Intelligent Drug Design' (AIDD) system was presented to delegates by Dr Charles Manly (Neurogen, Branford, CT, USA). It consists of a 'tight integration of automated robotic combinatorial synthesis, automated high-throughput screening, computational chemistry (modelling and virtual screening) and research informatics with medicinal chemistry', and has enabled Neurogen to find 200 dopamine D_4 antagonists with >65% inhibition at 100 nM in a project with a turnaround-time of three weeks. Using entirely proprietary software, AIDD is implemented iteratively in a 'limited similarity' virtual screen of an archive of compounds. Deconvolution of the (real) samples of interest only takes place at the final stage, and the sheer speed in cycle times gives the technology a leading edge.

'I've got repetitive strain injury in this [the reboot] finger' – Andrew Payne (Cyberdynamics) discussing Microsoft's software vs CORBA and Java

Helios also uses virtual screening techniques in conjunction with real chemistry. Being a new company, it 'doesn't have a problem with legacy systems' says Dr Herschel Weintraub (Helios, Louisville, KY, USA), showing how their Combinatorial Chemistry Optimization Assistant (or COCOA, for short – written in Visual Basic and an intentional pun about Java Beans) is used to control robotic synthesis. With it, they have reduced the optimization step in combinatorial library synthesis to 'a couple of days'.

Software solutions

There were numerous exhibitors at InfoTech Pharma '98 offering software solutions that are becoming more finely tuned (and, therefore, attractive) to pharmaceutical companies' needs. Space precludes a comprehensive listing here, but new developments include the acquisition of the specialty chemistry company Receptor Research, UK by the discovery research organization Tripos (St Louis, MO, USA). Receptor Research's solid-state chemistry 'will enable Tripos to directly apply its molecular informatics, design and analysis expertise in new compound discovery collaborations'. Another development is Elsevier Science's purchase of MDL Information Systems (San Leandro, CA, USA), which now has over 800 customers worldwide.

It's all very well having a Petabyte of data (1,000,000 Gigabytes), but how will you look at it? Dr Chris Jones (CERN, Geneva, Switzerland) gave the delegates food for thought from outside the pharmaceutical industry with some of the mind-boggling statistics found in high-energy physics. Asserting that data storage was 'no longer a problem' (one of the experiments at CERN this year will produce data at the rate equivalent to a pile of CDs 1 km high, per second), making sensible use of the data was key – 'in the process of drug discovery the most important component will be really clever people'.

IBC is planning an InfoTech Pharma '99: those interested should contact Sarah Duncan (tel: +44 171 453 2701, e-mail: sarah.duncan@ibc.co.uk).

Matthew Thorne
tel: +44 1223 836 396
e-mail: thorne@masha.demon.co.uk

In short…

Work has commenced on **Pfizer's** new £109 million R&D facility in Sandwich, UK. The new investment will contribute to 1,000 new jobs over the next few years. The new 50,000 m^2, three-storey block will house a broad range of research groups with differing operational requirements. Total investment in the Sandwich site during the 1990s is expected to reach £830 million. Pfizer was flagged as a 'World's Most Admired Company' in the pharmaceutical sector by *Fortune* magazine last year.

The 21st annual **Bristol-Myers Squibb** Award for Distinguished Achievement in Cancer Research has been awarded to Dr Michael Sporn for his work in chemoprevention of cancer. Dr Sporn developed the first widely recognized class of chemopreventives – derivatives of vitamin A, which he christened 'retinoids'. He elucidated the relationship between structure and activity for these compounds. He and others have shown that retinoids can reverse or inhibit the development of cancers at most epithelial sites. Dr Sporn is Oscar M. Cohn '34 Professor of Pharmacology and Toxicology and of Medicine at Dartmouth Medical School in Hanover, New Hampshire.