change the practice of medicine by providing physicians with essential information to precisely prescribe the appropriate drugs according to patients' genetic make-up and will provide enormous health benefits and cost savings to the public.

References

1 Ross, J.S. and Ginsburg, G.S. (2002) Integrating diagnostics and therapeutics: revolutionizing drug discovery and patient care. *Drug Discov. Today* 7, 859–864

2 Roses, A.D. (2002) Genome-based pharmacogenetics and the pharmaceutical industry. *Nat. Rev. Drug Discov.* 1, 541–549

3 Durant, J. *et al.* (1999) Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. *Lancet* 353, 2195–2199

4 Chaix-Couturier, C. *et al.* (2000) HIV-1 drug resistance genotyping. A review of clinical and economic issues. *Pharmacoeconomics* 18, 425–433

5 Mallal, S. *et al.* (2002) Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* 359, 727–732

6 Hetherington, S. *et al.* (2002) Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* 359, 1121–1122

Michael M. Shi
Sequenom
3595 John Hopkins Court
San Diego, CA 92121, USA

Meeting the challenges in screening

Jack Elands* and Matt P. Botwood, IDBS, 2 Occam Court, Surrey Research Park, Guildford, GU2 7QB, UK; *tel: +44 1483 688 325, *e-mail: jelands@id-bs.com; tel: +44 1483 595 000, e-mail: mbotwood@id-bs.com

The theme of this year's Society for Biomolecular Screening Eighth Annual Conference, held in The Hague (The Netherlands; 22–26 September 2002) was *High Information Content Screening*. The conference was attended by approximately 2000 delegates.

Novel screening methods with high information content

Sheri Miraglia (Applied Biosystems; http://www.appliedbiosystems.com) chaired the *Novel Screening Methods With High Information Content* session. She introduced the theme with the analogy of travelling by plane or on foot: whereas aeroplanes do not enable anyone to observe fine details of the area over which they fly, huge amounts of detail can be observed when walking; however, this is not practical for significant distances. The answer might lie in specialized technologies such as spy satellites, which offer both speed and accuracy.

The speakers broadly covered two topics: high content screening (HCS) for the determination of multiple factors in a single assay system, and parallel, multiplexed assay systems.

Co-chair Len Pagliaro from BioImage (http://www.bioimage.com) likened HTS to sitting in an early plane, from which one could only see blue (sky) and green (land). Today, we try to derive a single value from all the instruments in a modern cockpit, which is unrealistic. Although many of the data are irrelevant provided they are within a certain specification, some data are extremely useful for characterizing the interaction between a compound and target.

Image based HCS

Kurt Scudder of BioImage described how HCS could be used to measure multiple parameters of a single, cell-based assay system. Primarily, this is used to measure a main event, such as the translocation of a protein. To assess the validity of the main result, other parameters are recorded (for example, instrument failure, fluorescent compounds, artifacts caused by toxicity and too-low cell count). The main result can be qualified using multivariate analysis and extra parameters (for example, principal component analysis). Manual analysis is incredibly valuable; to exploit this information fully, machine vision, imaging-based clinical diagnostics and remote sensing, or satellite imaging, are considered useful. Stefan Prechtl from Schering AG (http://www.schering.de) described a similar use of different labels to detect multiple subcellular components.

Yan Feng from the Institute of Chemistry and Cell Biology (Harvard Medical School; www.iccb.med.harvard.edu) took the concept of HCS a step further, showing examples of measuring incomplete nuclear separation during cell division, and discussed several translocation and protein distribution assays.

A different perspective on HCS

Screening using *Candida elegans* was described by Gerhard Weidner from EleGene – now at Protodyne (http://www.protodyne.intranets.com). This organism is well understood and models of several human diseases have been characterized. Several targets have been explored using a fully automated imaging system, and additional information, such as toxicity and drug availability, is inherently derived from this assay.

PII: S1359-6446(02)02516-3

Multiplexed assay systems

Robert Umek from Meso Scale Discovery (http://www.meso-scale.com) represented the multiplexed approach. He described an electro-chemiluminescence assay based on an array of electrodes in a 96-well plate. A range of targets, such as studying a complete signalling cascade, can be assayed because a different target is immobilized on each electrode. Oren Beske of Virtual Arrays (http://www.virtualarrays.com) described CellCards™, small slides on which cells are grown. Several tens of these slides can be dispensed in a well. In this way, multiple cell lines can be assayed in the same well.

A different multiplexing approach is based on the use of a repeat set of engineered zinc fingers. When fused with an activator, this enables the induction of the expression of proteins coded by the gene for which the zinc fingers were engineered (Michael Holmes, Sangamo BioSciences; http://www.sangamo.com). This is an excellent mechanism for testing different targets in a single system, especially because its applicability has been demonstrated with a broad range of targets.

Ralph Martel from High Throughput Genomics (http://www.htgenomics.com) showed another multiplexing system based on chemiluminescence. A large variety of different assays can be developed using an array of specific oligonucleotides printed in a microplate, as well as a set of linkers, which bind to the array oligos and to a target molecule. An mRNA assay and an SNP assay were discussed.

The definition of HCS

One could argue that HCS is still not uniquely defined and that it is used to describe a variety of technologies that are targeted to do more than just the single point-screening assay. This was illustrated by other technologies presented:

The use of multiple aqueous two-phase partitioning systems rapidly to create a physicochemical signature of a compound. Based on a set of reference compounds, a signature can be obtained from which specific–non-specific or agonist–antagonist activities can be differentiated (Arnon Chait, ANALIZA; http://www.analiza.com).

Lasers can be used to selectively kill or 'opto-inject' cells, or to bleach or photoactivate targets in cells (Manfred Koller, Cyntellect; http://www.cyntellect.com). Using imaging, cells can be selected and non-responders eliminated. Confocal imaging facilitates this by reducing bleaching and increasing the sharpness of the image.

Moving towards more information-rich assays is no longer just an upcoming trend, but a reality. The HCS sessions demonstrated that we are at the beginning of a rapidly expanding, new and potentially rich technology.

From data to knowledge: machine intelligence in informatics

Generating real knowledge from the mass of data in biological screening is one of the biggest challenges facing the industry. This session focused on the use of computers to aid this process. A key issue is the sheer volume of possibilities to be calculated and evaluated, hence primary goals are to reduce solution space and improve the quality and speed of predictions.

Douglas Kell's group at the University of Wales (http://www.wales.ac.uk) has a novel approach. Using metabolome data as the input, they turn the traditional hypothesis–data relationship on its head and start with data, which leads them to a hypothesis. Using a genomic computing technique they start with a program (denoted with an intuitive tree-like structure), evaluate it with a fitness function and then use experimental data to 'mutate' the program before re-evaluating its fitness. This approach can provide more realistic answers than traditional clustering methods.

Christoph Helma of the Albert-Ludwigs University, Freiburg (http://www.uni-freiburg.de) explored the use of data-mining techniques with inductive databases, and highlighted the importance of continuous data cleansing for knowledge discovery. Inductive databases (analogues to deductive databases, conceptually containing all facts derivable from the data and the rules, as well as the data itself) enable patterns and models to be queried instead of just data. Helma's example used an inductive database to mine molecular features (based on a SMILES/SMARTS format). He avoided all solutions that are either too specific or too general and concentrated on the most frequent solutions to reduce solution space.

Structure–property relationships

Christos Nicolaou from Bioreason (http://www.bioreason.com) focused on structure–property relationships, singling out a popular approach: the organization of data into structural classes and class-based reasoning. The main problem here is that the data is not fully used, resulting in under-representation of some structural classes and the generation of false positives and false negatives. Nicolaou's approach uses a new method to identify classes of structure. It is more intuitive, involves multi-domain classification and is not dependent on predefined fragments or activity. It defines structures by topology, not just by substructure, and, consequently, is more aware of the surroundings of the fragment, such as adjacent rings. The technique starts with standardization of the structures into a SMARTS based nomenclature. Once standardized, potential scaffolds are generated and the structures characterized using recursive partitioning. Both the molecular descriptor (structure) and physical properties are taken into account. This method appears to be good for rule extraction, new compound selection and noise detection (false negatives and positives). It is, however, susceptible to noise and imbalance in the different classes.

An automatic analytical pathway to study DNA motifs using freely available computational tools and traditional

techniques with novel algorithms was described by Kathleen Marchal (Catholic University Leuven; http://www.kuleuven.ac.be). Using a combination of sequence data and biological results (expression profiling), Marchal data-mined microarray results using a clustering technique and then data-mined the DNA sequences. Finally, comparative genomics and phylogenetic footprinting were deployed to reduce false positives.

Luc Dehaspe of PharmaDM (http://www.pharmadm.com) described a novel relational data-mining approach for multi database–multi discipline applications. Its main advantage is its ability to mine rich data sources. Although humans are good at mining rich data, current tools are not, mainly because data first needs to be reduced to a 'single table' format. Relational data mining aims to address this, leaving the data *in situ* and unconsolidated. By avoiding data reduction you have the potential to exploit more of your data for mining. One potential downside that was not mentioned is the likelihood of performance degradation using this approach compared with data warehousing techniques. However, the accuracy of predictions might be more important than speed in certain drug discovery applications.

Paul Blower at LeadScope (http://www.leadscope.com) discussed building predictive models from large screening sets using chemical characterization by common descriptors. Clustering techniques group the molecules, and, for each cluster, key features are identified that distinguish them from their nearest neighbours. The technique assumes that these features are important for their activity, which are then used to build a local prediction model. Some outliers, however, cannot be predicted. They are still working on the prediction algorithm, and some compounds have no prediction at all. Ultimately, they expect to use this technique to identify outliers and to aid in the selection of external libraries.

This was a stimulating session with a strong mix of presentations spanning a range of computational issues. Many of the ideas appear to have the potential to make a positive impact on the way we generate and make use of data in drug discovery. An analysis of other data-intensive sectors, covering how they tackle data issues using computational techniques, could be a most instructive area for further research.

Concluding remarks

Screening has come a long way since the SBS was founded. The much advertized 'brute force approach' has been replaced by more intelligent approaches. Screening could become the technology to validate in the lab what has been determined *in silico*. Although the initial SBS meetings were often describing the existing state of screening, more recently the focus has changed to reflect upon the future. The Eighth Annual Conference definitely continued and reinforced the forward-looking trend of the recent SBS meetings.