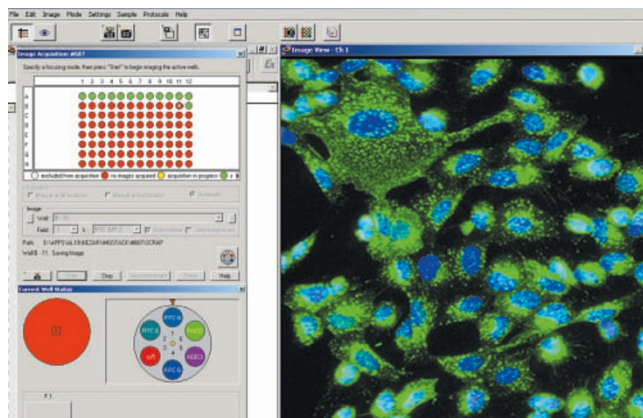


# The leading edge

Synthesizing and selecting the lead compounds that will become the drugs of the future are the heart of drug discovery. Tim Chapman goes back to basics.

Over the past decade, high-throughput screening (HTS) has revolutionized the process of drug discovery. Automation and miniaturization have allowed biotech companies to screen hundreds of thousands of compounds against disease targets, in the hope of identifying the precious few that can be developed into commercial, patentable medicines. To use a fashionable analogy, it's like speed dating for medicines.

But this approach of simply testing more and more compounds at faster and faster rates is not as productive as the industry had hoped. Because the earliest stages of the discovery process are relatively unselective, the vast majority of early hits fail to progress, either because of the attrition caused by medical or chemical failings discovered only in the later stages of development, or because the flow of potential leads is still too overwhelming. Drug-discovery researchers, and the companies that supply their technological needs, are now concentrating on improving both the quality of the compounds entering the first screening and the



Looking in: output from the IN Cell Analyzer 1000.

value of the assays being done. Rather than bigger and faster, the focus is on doing things earlier and smarter.

The demand for high-throughput screening began in the early 1990s, when drug companies began testing the landslide of compounds created by combinatorial chemistry against an increasing number of disease targets identified by genomic research. "What we then realized was that the bottleneck, rather than being in the set-up and running of the assay, moved to reading the assays themselves," says John Anson,

vice-president of development at GE Healthcare Bio-Sciences (the former Amersham Biosciences), based in Little Chalfont, UK.

Automated assays demanded the use of plates with arrays of wells, and early systems measured each well individually. As plate densities increased up to 1,536 wells per plate, complex camera systems became available that could read the whole plate simultaneously. GE Healthcare's high-throughput screening platform, LEADseeker, is now in its fourth iteration and offers a range of imaging modes within

the single instrument, allowing richer measurements.

"What's interesting about high-throughput screening is that there's no dominant way of measuring an assay," Anson notes. "Customers can use luminescence, radioactivity, or lots of different fluorescence modalities. There's not one dominant modality, so we had to focus on sensitivity and assay performance as well as speed." In fluorescence mode, LEADseeker reads a plate in about a minute, allowing throughputs of up to half-a-million assays per day. "It was that ability to

GE HEALTHCARE

## MINING THE PROTEOME

Most automated approaches to drug discovery are based on testing thousands of compounds against a single disease target. The platform developed by Serenex, a discovery company in Durham, North Carolina, takes the opposite approach by testing drug candidates against thousands of protein targets simultaneously. This technique, which the firm calls proteome mining, promises to deliver unique information about interactions between compounds and targets.

"In contrast to profiling compounds against a single target, what we're doing is profiling one compound against all targets in an entire sub-proteome," says Steven Hall, the company's senior vice-president of R&D. "We get extremely broad selectivity and information."

Serenex is focusing on the purine-binding proteins within the human proteome. This group of some 2,000 proteins includes protein kinases and many other clinically important proteins that use purine nucleotides such as adenosine triphosphate as co-substrates or co-factors. A potential

drug candidate can be tested against all 2,000 in a single step.

The approach depends on a patented affinity medium of small beads to which many molecules of a purine nucleotide such as ATP are tethered. When a cell or tissue extract is mixed with this medium, the tethered molecules ensnare its purine-binding proteins. Unbound proteins are then washed away, and a drug compound introduced. If the compound has an affinity for any of the bound proteins, it will displace them from the tethered purines. A selective drug will displace just one protein, whereas less-selective compounds will displace many different proteins. The displaced proteins can then be collected and identified by mass spectrometry and bioinformatics.

The aim is to help medicinal chemists make a more informed decision about which compounds to move forward to lead optimization. Larger-scale screening of large compound libraries against sub-proteome sets of proteins will provide a very useful and information-rich matrix of compound-target interactions, Hall notes.

"We feel that we will get to the point where we can develop some fundamental understandings of what kind of substructures will lead to effects on what type of target," he says. "It's only by screening large libraries of targets against large numbers of compounds that the informatics community will be able to get the connection between structure and protein affinity."

T.C.



Taking a molecular view.

TECAN

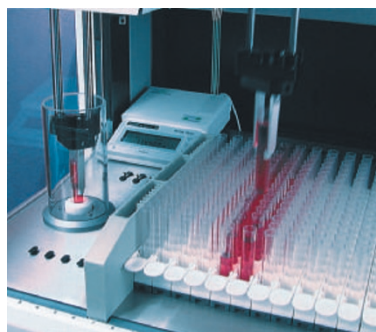
run things at this throughput that helped us realize where the next bottleneck would be, which is turning these hits into lead compounds," says Anson.

### Into the cell

The vast majority of compounds identified as hits in screening fail in the later stages of lead development, usually because they are biologically unsuitable for development as a medicine. It makes sense to find out more about the biological workings of a hit compound as early as possible, before you spend time and money on further development.

"One area where we are seeing continuing demand for products in the early stages of drug discovery involves cell-based assays," says Marc Feiglin, biopharma technology officer at life-sciences equipment supplier Tecan, based in Männedorf, Switzerland. Tecan's Cellerity system provides an automated system for high-throughput cell-culture production and assay preparation.

"One of the reasons people are using cell-based assays is they're starting to get earlier indications of problems that may occur later on," says Anson. GE Healthcare's IN Cell Analyzer platforms offer real-time high-resolution cellular imaging, allowing researchers to construct high-throughput intracellular assays that were previously impossible (see 'Proteins on the move', below).



Automated analytics.

Cell-based assays produce much more information on the effects of a potential drug compound than *in vitro* assays can, and this enables researchers to get more useful data as early in the discovery process as possible, says Mark Collins, senior product manager at Cellomics, a cell-assay platform developer

based in Pittsburgh, Pennsylvania. "You can screen millions of compounds very quickly, but the challenge for pharma is getting enough information out of that screening to make considered decisions to move forwards," he says.

Cellomics' high-content screening platforms such as the ArrayScan and KineticScan readers use fluorescence probes to obtain richer measurements of cellular activity. "The point is that you're measuring more than one thing that's going on, and analysing images in an automated fashion to get quantitative data," Collins says. "You can address biology you could never have understood before, things like morphology screening that you just can't do if you've smashed the cell up and put it in a test tube."

### Getting the knowledge

With both throughputs and the richness of assay measurements increasing, making sense of the vast amounts of data generated can be a bottleneck. Informatics or knowl-

edge-management systems are a necessity, and many platform providers offer integrated software packages.

"When you are generating multiple images and need to get a quantitative image out quickly, you're not going to do that by running millions of spreadsheets or looking at millions of images," says Collins. "What our informatics package does is automate that whole process of linking data together and linking images with numbers and storing that in a very robust database. The image is always there as a final answer, but once the instrument has done its job you're looking at numbers, which is a sight easier than looking at 1,000 images."

Informatics can also help reduce attrition by improving the selection of hits to take into further development. "By integrating our synthetic chemistry work with a knowledge-management approach, we've demonstrated that we can improve the interaction profile of a customer's project by increasing the number of lead series they have for each target, which gives them more shots at the goal," says Ed Hodgkin, vice-president for business development at discovery services group Tripos, based in St Louis, Missouri.

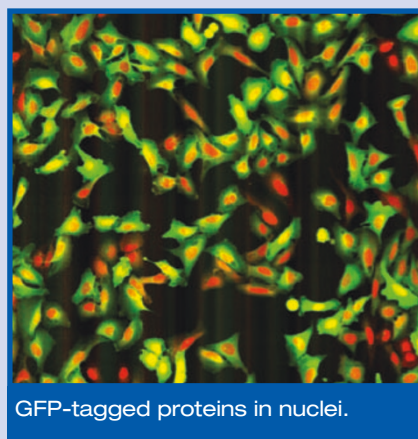
In a collaboration with New York-based pharmaceutical giant Pfizer, Tripos has developed the SARNavigator suite of tools for analysing high-throughput screening results. The package allows researchers to examine structure-activity relations (SAR) quantitatively to help determine which hits should go forward into the lead identification and optimization phase. "That combination of SAR-Navigator and the chemical compounds we

## PROTEINS ON THE MOVE

BIOMAGE

Intracellular signalling pathways are involved in virtually every disease process, and cell-surface receptor proteins that initiate these pathways are favoured drug targets (see *Nature* **428**, 225–233; 2004). In the intracellular pathways leading from these receptors, signalling proteins move within the cell — from the cytoplasm into the nucleus, for example. Testing the effects of potential drug compounds against these intracellular protein movements has, unsurprisingly, proved much harder than testing against proteins at the surface of the cell. The ability to do so promises to find drugs with new modes of action and greater selectivity.

A leader in this field is Biolumage in Copenhagen, Denmark, which focuses on green fluorescent proteins (GFP) as a tool in drug discovery. Biolumage's proprietary Redistribution technology uses genetically encoded GFP tags to track protein translocation in living cells, and so identify compounds that inhibit onward transmission of the signal. "It's pretty clear now that every signalling pathway has components that must move in the cell for the signal interaction to occur," says Len Pagliaro, vice-president for business development at Biolumage. "If you take these targets and



GFP-tagged proteins in nuclei.

track them, you can measure movement in response to potential stimulators and agonists, and have the potential to find compounds that inhibit that movement."

The firm deploys GE Healthcare's IN Cell Analyzer to carry out high-throughput imaging for screening and analysis of the effects of members of its own library of over 250,000 compounds. Such intracellular assays provide the opportunity to discover drugs that interact with target proteins in new ways — by targeting protein localization domains rather than catalytic domains, for instance. Studying living cells also creates more opportunities for serendipitous discoveries.

"Obviously, cell-based assays are inherently a somewhat dirty system biologically, and that has both benefits and liabilities," Pagliaro notes. "The

benefit is that in principle it's more like the pathway you're going to see *in vivo*. The liability is that when you get a hit, you don't know that the compound is acting on the target that you want to see it acting on. We like that aspect, because that can deliver very interesting progressible hits that we would not have found otherwise."

T.C.

provide is a very powerful combination in going from a hit to a lead," Hodgkin says.

Applying informatics along the discovery process can reduce the time taken in each phase, both by exploiting data mining to reduce the attrition of candidate compounds and by managing the logistics of the whole process. Benchmarking exercises by Tripos and Pfizer have found that the time taken to go from initial screening to a preclinical drug candidate can be reduced to 9–18 months from an industry standard of 2½–3 years.

Combining data-mining techniques with high-precision screening and library selection can offer dramatic improvements in productivity. Drug-discovery company Amphora, at Research Triangle Park, North Carolina, focuses on finding small-molecule therapeutics for big-pharma partners such as Strasbourg-based Aventis. "The quality of the data produced at Amphora has a direct effect on drug-discovery efficiency," says Jeff Riley, vice-president of corporate development. Structure–activity relation modelling allows the firm's chemists to move directly from primary screening to medicinal chemistry stage. "This enables Amphora to efficiently and repeatedly go from target to lead in six months," Riley notes. "This is true predictable drug discovery."

### Library science

Amphora also prides itself on the selectivity and purity of its compound library. "The old computer software adage, garbage in garbage out definitely applies to HTS," says Riley. "The higher the purity of the library and the more that is known about the quan-

titation, the higher the quality of the data."

Improving the quality of the compounds going into initial screening is perhaps the key factor in improving the usefulness of the hits coming out. Researchers are increasingly realizing the waste involved in simply screening vast numbers of compounds without sufficient preselection.

"I do not believe that you can solve the problem of finding drugs through a numbers game," says Stephen Hill, chief executive of biotech firm ArQule, based in Woburn, Massachusetts. "The problem with high-throughput screening is there's almost no end to it, because the number of compounds you could potentially use is  $10^{60}$ . If you have an ocean to choose from, it doesn't make much difference whether you sample a spoonful or a cupful—it's still not representative."

ArQule has developed a proprietary method for producing compounds with the best balance of drug-like properties, or optimal chemical entity (OCE) compounds, which it uses to offer high-quality library design and production to research partners such as Pfizer.

"The goal is to increase the likelihood that Pfizer will be able to generate from primary screening more interesting hits against a broad range of its targets," Hill says. "Because of the way we've designed those compounds, it's more likely the company will be able to optimize within that chemical space a compound with all the drug-like characteristics to make a medicine."

The OCE approach uses predictive models and absorption, distribution, metabolism and excretion (ADME) assays as early as pos-

sible, and looks for the best balance of drug-like characteristics at every step rather than simply prioritizing compounds with the highest affinity for the target. "What we try to do is think about all these parameters in parallel at every stage of the process, starting from library design and prediction through each iterative step," Hill says. "When you're trying to do a multiparameter optimization, trying to trade off up to a dozen chemistries, that's when you need the power of computational tools backed up by relatively quick and affordable experimental capability."

Producing high-quality, targeted compound libraries has become an industry in itself. One of the largest players is Chemical Diversity Labs (CDL), based in San Diego, California, which produces some 200,000 compounds in 500 new library systems per year. "There is a greater awareness of the problems associated with the failure of leads late in drug development, and the emphasis has shifted to their quality and survivability in the development process," says CDL president Nikolay Savchuk. "Customers want screening libraries with new chemistry, preferably beyond the scope of current patents and literature references, but also with a higher probability of hitting the target."

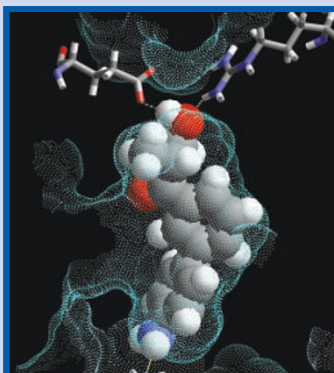
CDL's targeted libraries are derived from annotated databases that unify knowledge about targets, ligands and their relationships. Compounds are also ranked by factors such as novelty and ease of synthesis.

Entirely new compounds designed by molecular modelling (see 'Starting from scratch', below) create extra challenges for chemical synthesis, but the results can justify

## STARTING FROM SCRATCH

Computational methods are being applied across the drug-development process from the earliest stages of drug design. "We're seeing a lot of companies who don't want to spend time and money on screening existing compound libraries," says Bart Wuurman, chief executive officer at De Novo Pharmaceuticals, a leading *in silico* drug-design company in Cambridge, UK. "The traditional approach is looking for a needle in a haystack — what we do is create a needle."

Founded in 1999 as a spin-out from the University of Cambridge's drug-design group, De Novo last year refocused itself from in-house drug discovery to meeting the growing demand for its proprietary computational platforms. De Novo's core platform is a suite of structure-generating programs called SkelGen. The core algorithm generates new molecular structures by the random assembly of small molecular fragments. This stochastic approach rapidly searches the whole of the chemical space of potential drug compounds, explains chief scientific officer Philip Dean. The search is constrained by any information available about the structure of the protein target from X-ray crystallography. In cases where little is known about the target, information about the structure of a known ligand that affects the



SkelGen-modelled ligand.

target's activity can be exploited. The firm's Quasi2 program uses data on the known active ligands to create a negative image of the target site, which can then be input into the SkelGen algorithm to find new patentable compounds with a similar action.

"The unique technology is the ability to create novel chemotypes for any set of constraints," Dean says. "We think there's no one else out there who can do *de novo* design from just an active compound." The firm's researchers have also developed a method for accommodating the flexibility of the protein target, which structure-based molecular design has previously failed to do.

De Novo's platform takes into account the need to convert *in silico* designs into actual chemicals. "We have an algorithm that post-processes the SkelGen

structures to prioritize those structures that are readily synthesizable in the smallest number of steps," Dean says. "Not only are the compounds scored for the fit with the site but also for ease of synthesis." In collaborations with Roche, over 90% of output compounds have been relatively simple for chemists to synthesize. "Basically it's our intent to get the designed molecules into wet chemistry as quickly as possible," Wuurman says. **T.C.**



the effort. "There should be a good rationale for making molecules, and *de novo* design provides such a rationale," says Savchuk. "To be successful, *de novo* design needs to be innovative but cost-efficient. As with other types of synthesis, *de novo* design has a better chance of success when screening libraries instead of single compounds are produced."

Specifically targeted compound libraries will also allow the use of more complex assays earlier in the discovery process. "If you're using a much more focused library, you can address much more complex biology at the first screen," says Anson. "I'm quite excited about moving in this direction because lots of our cell-based assays will be very amenable to high-information content screens against these focused libraries."

### Back to nature

Researchers are also looking to broaden the sources of libraries for screening. Compounds created by combinatorial chemistry have dominated discovery programmes for the past decade, but attention is now returning to the wealth of compounds derived from natural products. Part of the reason is that combinatorial techniques may, paradoxically, be limiting the diversity of compound libraries.

"People traditionally built combinatorial libraries out of small molecules that they could easily synthesize, and there's a huge overlap with the libraries that other people have," says Dirk Ullmann, head of biology at discovery services group Evotec OAI in Hamburg, Germany. "Library purity and stability is also a concern."

Natural products offer unsurpassed chemical diversity and novelty, often with better affinity characteristics and evolutionary selection for biological activity. "The big advantage is you don't meet this kind of high number attrition you meet with small-molecule chemistry," Ullmann notes. "You end up much earlier with a good lead that might even be the natural product itself."

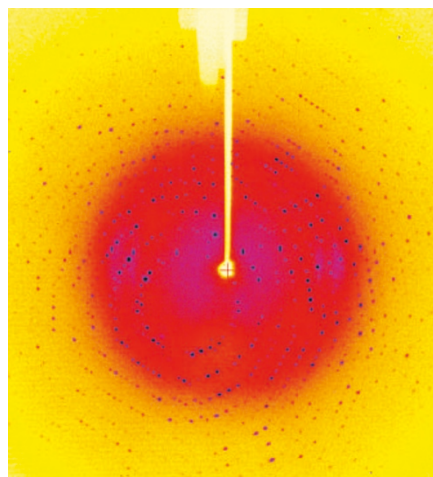
Testing natural products in a high-throughput environment is inevitably more challenging than dealing with synthetic libraries. Compounds from natural products have first to be isolated and fractionated, and creating suites of chemical analogues is more difficult.

Evotec OAI recently partnered with Biofrontera Pharmaceuticals in Leverkusen, Germany, to offer a high-throughput screening platform for one important group of natural products, those produced by bacteria and fungi, the source of many current antibiotics. The EVObioscreen platform accesses Biofrontera's largely unscreened library of fractions from some 50,000 strains of actinomycetes and fungi at ultrahigh throughputs. "It's much more unpredictable to work with plants or other organisms, so you really have

to work with microorganisms," says Ullmann.

An even more targeted supply of drug candidates may be provided by a new approach to drug discovery being piloted by a handful of companies in Europe and the United States. Fragment-based discovery uses very small molecules, termed fragments or scaffolds, as a basis for building designer compounds that may be able to tackle previously intractable disease targets.

"The whole concept of fragment-based drug discovery has gained quite a high profile in the past two or three years in the industry," says Harren Jhoti, chief scientific officer at Astex Technology in Cambridge, UK. "It's



Crystal clear: X-ray diffraction data from a protein-ligand complex.

fair to say that the hope that we all had for the high-throughput approach has perhaps not been met. One limitation has been the complexity of the compounds themselves."

Using very small molecular fragments as a foundation for building precisely targeted drug molecules allows you to keep the complexity to a minimum. The fragments in Astex's library, most of which are selected *in silico* on the basis of predicted binding efficiency, have relative molecular masses ( $M_r$ ) of 100–250 and are used as the basis of compounds of  $M_r$  up to 500.

### On the scaffold

Plexxikon, a scaffold-based drug-discovery company in Berkeley, California, has a library of some 20,000 scaffolds of  $M_r$  150–300. Again, these are used as the basis of drug compounds with the smallest possible molecular masses. "What we generally find is that these leads tend to be more soluble, have better metabolic stability, and better *in vivo* ability from the very beginning," says Mike Millburn, the company's senior vice-president of research. Plexxikon applies structural screening as early as possible in the discovery process, allowing the firm's chemists to identify the most optimizable compounds as quickly as possible.

Because fragments have an inherently low binding affinity, traditional bioassays are of limited use, although Plexxikon uses bioassays in its initial screening procedure. The key to fragment-based discovery lies with X-ray crystallography.

"The real unique aspect of using X-ray crystallography is that you can detect the binding of the fragment and very precisely define the binding orientation," says Jhoti. "In the past five or seven years, there's been a huge set of developments which has allowed X-ray crystal structures to be obtained in a much more efficient manner than before." (See *Nature* **421**, 661–666; 2003.) Astex's proprietary 'Pyramid' discovery platform includes high-throughput crystallography, data collection and analysis, and can generate up to 400 protein-ligand crystal structures from X-ray data in a month, comparable to the number achieved in a year by groups using more conventional crystallography tools.

Information about the co-crystal structure of the target protein and scaffolds allows chemists to select the ones that can be most efficiently optimized for synthesis. "Having that sort of information before you start the chemistry really allows you to explore many different chemotypes for a particular target," Millburn says. "That's in contrast to high-throughput screening where you may get two or three different chemotypes based on potency that may not be the most tractable in terms of synthesis chemistry."

### The bottom line

With such a diverse selection of innovative technologies and approaches being deployed at the earliest stages of the drug-discovery process, some will ultimately prove to be more productive than others. But most workers in the field agree that the current situation is not sustainable, and productivity must be improved.

"Productivity in industry is getting worse year on year," says Hill. "We know that today's drug-discovery productivity is really frighteningly bad, so continuing to do things in the way they've been done in the past is unlikely to get us where we need to be."

The good news for drug companies that have already invested heavily in high-throughput screening systems is that most of the new approaches can be added to existing systems. But to make the most of them, they will have to sweat their assets, says Anson. "It's about balancing the work, to run high-throughput factory-like operations with the recognition that we're still talking about a scientific discipline and there still has to be a certain degree of thought and planning to get a successful outcome," he says. "It's the companies that combine the ability to sweat their assets with really thoughtful design and processes that are really going to come out on top." ■

Tim Chapman is a freelance journalist based in Halifax, UK.