

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."

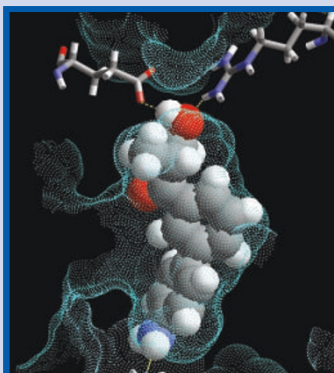
CDL's targeted libraries are derived from annotated databases that unify knowledge about targets, ligands and their relationships. Compounds are also ranked by factors such as novelty and ease of synthesis.

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.**