

Reprint from

Daily Biotech Updates... [www.genengnews.com](http://www.genengnews.com)

# GENETIC ENGINEERING

Vol. 21, No. 20, November 15, 2001

## NEWS

## Firms Creating More Focused Compound Libraries

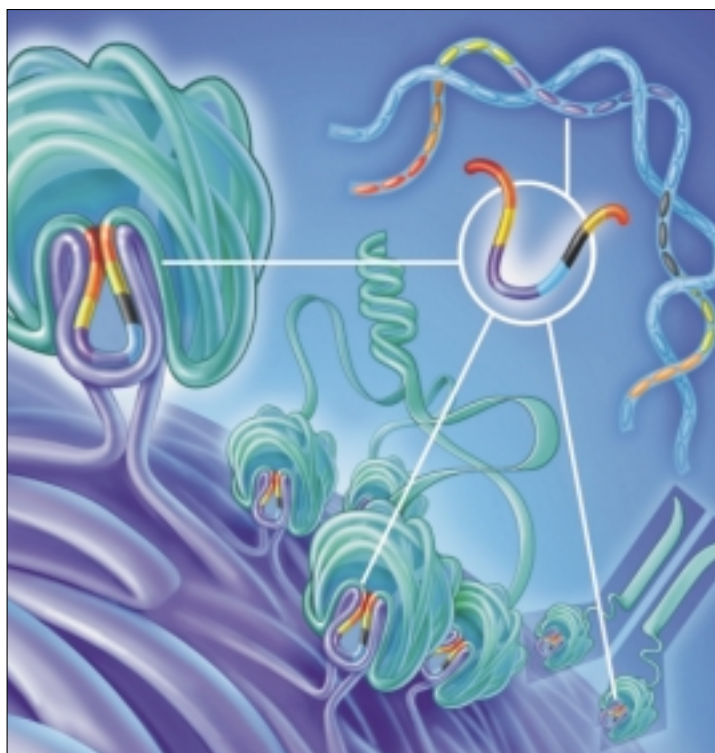
### Saving Time and Money in Drug Discovery Research

Nina Flanagan

A recent trend in drug discovery technology is creating more focused compound libraries rather than large, diverse ones. This is not only because of the obvious time saved but also the money savings. Several companies are developing methods to reach this same goal but are using different routes.

**Chemical Computing Group, Inc.** (Montreal) has developed software with a new approach to high-throughput drug design. The company's method uses high-throughput screening (HTS) experimental data to create a probabilistic QSAR (Quantitative Structure Activity Relationship) model, which is subsequently used to select building blocks in a virtual combinatorial library.

It is based on statistical estimation instead of the standard regression analysis. This is one application of five core applications the company has developed for its Molecular Operating Environment, or MOE.



MorphoSys' HuCAL<sup>®</sup> automated antibody-generation technology is designed to accelerate the identification of new disease-associated targets. MorphoSys uses HuCAL to generate fully human antibodies tailored to specific applications.

The Probabilistic Design Model has two applications: the QuaSAR-Binary<sup>™</sup> for HTS QSAR and the QuaSAR reagent for focused library design, or reagent scoring.

"The main difference here is with the QSAR models, normally linear regressions or regression trees are done. Instead, what we do is build up statistics of the occurrence of an active compound with 'descriptor values.' Then what we get is

the probabilities that the compound is active or inactive, as opposed to its activity as some sort of function of the descriptors," explains Chris Williams, Ph.D., director of scientific support.

The goal is to get a "bet" that the compound is going to be active. This is done with a probability-modeling formula that uses the facts that a compound is: active, "druggable" (that is, has good transport activities and is not toxic),

and has activity-assuming structure.

“Generally, we assume druggability will be independent of its activity against a certain target. That way, you can have probabilities balancing competing goals. What’s nice about these probabilities is that you can keep adding different things to it and just multiply the whole thing out to get your overall ‘bet.’ So you can add things without having to mess around with the model too much.”

### Binary QSAR Method

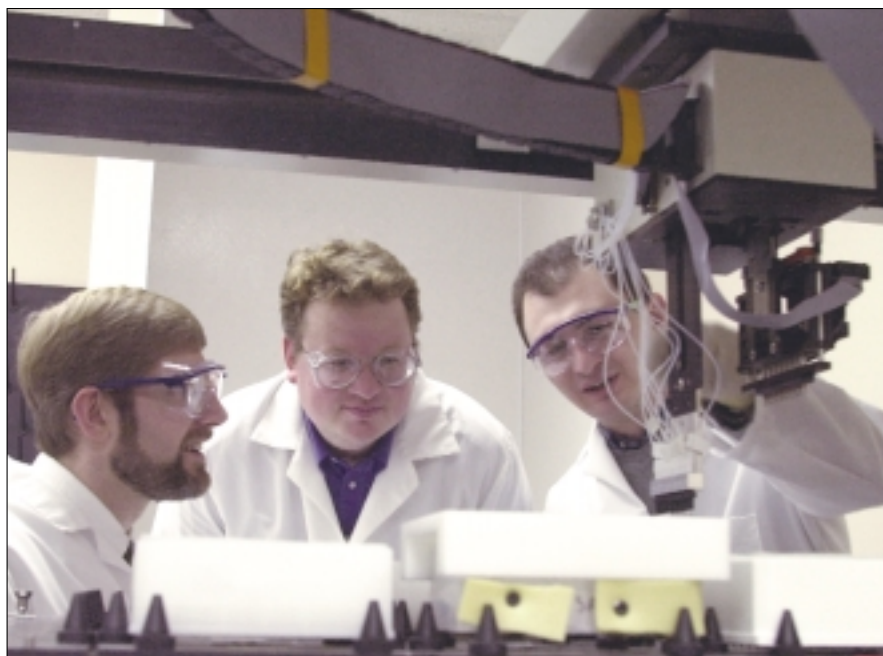
The company’s Binary QSAR method involves a mathematical formula that calculates the probability, given a molecular descriptor, that a molecule is active. Dr. Williams explains what makes this formula unique is that it includes data on inactive compounds as well as active ones.

“With our statistical model, we can include the inactive compounds, because that is useful information—statistics on what makes compounds inactive is very helpful.” The Binary QSAR method is used twice in the lead discovery process: once for the activity model and once for the druggability model.

The company discovered that atomic properties could be broken down into three characteristics: partial charge, molar refractivity (bonding interactions), and logP (lipophilicity of molecule). Each atom has a surface area in the molecule and it has these three properties associated with it.

“We find all the atoms with a partial charge in a certain range and add up all the surface areas. We use these VSA (Van der Waals Surface Area) descriptors to build the binary QSAR model,” says Dr. Williams.

“This is a simple, geometric formula that can be applied to large databases without having to do 3-D searching. You



Scientists at Albany Molecular Research check some of the robotic equipment used in the company's drug research and development programs.

can calculate surface-area descriptors of 5,000 compounds in a few minutes. This is intended for lead discovery and lead optimization in the early stages,” explains Dr. Williams. The binary models can be used to make activity models or ADMET models, and these can then be used to build a library.

The Binary QSAR can be used as a focusing agent when building a combinatorial library. The idea is to “score” R groups and only make compounds with the best R-group positions. Starting with R-group databases, scientists randomly sample the R groups, create the compounds, use the Binary QSAR model to predict activity of the compounds, and every time an R group appears in a compound, the score goes up.

This is the building-block scoring methodology—it estimates the probability that the building block is in an active compound. The R-group scores indicate how often it appears in an active compound. The top-scoring reagents are selected for pure combinatorial design.

Other companies have tools to make

QSAR libraries and to focus libraries, but Binary QSAR is proprietary to the Chemical Computing Group. Its cross-platform open architecture allows flexibility to be customized for whatever molecular computing it is intended to perform. Fully integrated, the software can be deployed throughout a pharmaceutical company to allow for sharing of information and data.

### Fully Human Antibodies

The demand for antibodies continues to grow via genome sequencing and the increasing number of new target molecules. Large numbers of antibodies are needed to determine functions of decoded genes and for therapeutic applications; it’s important that antibodies are 100% human to work successfully. In addition, antibodies must be customized to individual target molecules for optimum therapeutic effect.

**MorphoSys AG** (Munich, Germany) has developed technology to address these antibody requirements. HuCAL® (Human Combinatorial Antibody

Library) generates antibodies quickly and reliably, and its library currently has approximately 10 billion human antibodies. Its in vitro system develops antibodies synthetically, eliminating the need for experimental animals. The antibodies are of fully human composition.

“We looked at all the known human-antibody gene sequences and created antibody gene sequences that came as close as possible to that. Then we took all those gene sequences and reduced it down to 49 gene families, which represents the repertoire of human antibody genes and those are the ones in our library,” explains David Lemus, CFO.

He adds that MorphoSys is one of six companies worldwide that has technologies to create fully human antibodies. Lemus claims the company has “the ability to optimize our antibodies—meaning we can do screenings of different sub-libraries because our antibodies are built 100% synthetically. This allows us to optimize our antibodies to any specifications.”

### Molecular Cut and Paste

HuCAL's modular design enables the generation of custom human antibodies with drug qualities. This molecular “cut-and-paste” process means those portions of the antibody responsible for a particular property are easily accessible and can be readily changed.

“The modular system allows us to shuffle the genes of the complementary-determining regions (CDRs) of the antibodies (the gene regions that code for the area that hooks onto your target, i.e., binding properties). What we're able to do is to shuffle those genes in a modular way so we can use different modules at those CDR sites and create sublibraries easily.”

The company's proprietary TRIM (TRInucleotide-directed Mutagenesis) technology, exclusively licensed from Johns Hopkins, is a method for random mutagenesis of peptides or proteins that facilitates the targeted variation of CDRs.

Preassembled trinucleotides are used in chemical synthesis of the CDR sequences, ensuring the complete control over amino acid composition and avoidance of stop codons. This results in the CDR libraries being of substantially higher quality than is possible using conventional approaches, claims the company.

“If anyone shuffled these CDR regions to create an optimized antibody, it would just take too long—one round would take at least one to two years. But with HuCAL, the optimization process can be done in a matter of weeks. The modular libraries we create are difficult for other companies to do, so we offer a unique product in the antibody world,” states Lemus.

Another important advantage is high-throughput antibody generation capability, which plays an important role in the determination of gene function. AutoCAL™ automates antibody generation, allowing MorphoSys to meet the increasing demand for antibodies as research tools.

MorphoSys offers two types of collaborations. Target research may include a partner wanting access to their antibody library to validate potential targets. This may include Expressed Sequence Tags (ESTs), and the company would use its HuCAL-EST technology in combination with AutoCAL and expression profiling by immunohistochemistry using a proprietary human tissue collection.

Another type of target research deal would include out-licensing of their library for the company to use on-site for

their own target-validation work. Therapeutic product deals range from companies presenting one target molecule to many molecules.

A recent agreement with **Oridis Biomed GmbH** (Graz, Austria) will use MorphoSys' technology to validate drug targets while MorphoSys will have access to Oridis' human tissue bank. By the end of this year, the company will announce its plan for in-house target validation mechanisms for its own product development. The first HuCAL product candidate may enter clinical trials before the end of this year, however market entry is still several years away.

### Testing of Live Organisms

**Harvard Bioscience, Inc.** (Holliston, MA) recently acquired **Union Biometrica, Inc.** (Somerville, MA), and its high-throughput applications in drug discovery. Model organisms are becoming fundamental tools for functional genomic research and have potential for high-throughput drug screening. The company's technology, COPAS™ (Complex Object Parametric Analyzer and Sorter), allows rapid analysis and sorting of popular model organisms, including *C. elegans* (worm), *D. melanogaster* (fruit fly), and *D. rerio* (zebrafish).

These all have disease genes that are similar to human genes; analyzing them in relation to humans (comparative genomics) is becoming a powerful technique in drug discovery. This automated system provides fast, novel transgenic models of human disease and high-throughput screening of compound libraries in these living organisms.

“The main factor that differentiates COPAS technology is throughput. A million beads, each with a different compound, can be done in half a day. I'm not

aware of any company using either cell-based assays or plate-reader assays that can get anywhere close to that type of throughput.

“In the field of using model organisms, in drug discovery applications, the big difference in what we can do is that we automate otherwise manual processes. (The company states it can analyze one million *C. elegans* in a day.)

“So we’re talking about a massive change in productivity—picking out transgenic nematodes from a population of 10,000, you might have to pick just one. Doing that by hand is going to take a long time under a microscope, but we can do that in just a few minutes,” claims David Green, president of Harvard Bioscience.

### Fluid Design

The organisms are placed in a liquid suspension and enter an optically clear flow channel. Since these organisms mentioned above are optically transparent, COPAS uses a fluorescent marker technology to detect and quantify mRNA and protein expression. Because these live organisms are constantly moving, the company developed a novel fluid design.

A fluid stream orients the organisms and straightens them for about one millisecond, which is enough time for the laser optics to acquire data regarding the pattern of protein expression and to make a software decision to sort the organism for other analysis. The organisms are then sorted into single wells of microtiter plates.

“COPAS uses an air jet that pushes the fluid stream away so that when the detector finds something it likes, it shuts the air stream off and fluid drops down into a 96-well plate. The reason this is important is because if you try to sort nema-

todes or large beads with electrostatic sorting, the nematode would be killed, and if you are using beads, the chance of actually hitting a single well in a 96-well plate over a drop distance [over a meter] is almost zero,” explains Green.

Another feature of this system is the sheath fluid or the liquid that runs through the flow cytometer. The liquid is specially made to maintain laminar flow (smooth flow) so there is no turbulence. “This is essential to analyze the particles; if they are not flowing in a straight line, you won’t be able to analyze them properly,” adds Green.

The company says the system’s potential for ADMET screening and real-time gene expression has not yet been tapped, and these are significant markets. “I think you will see early ADMET screening move away from rat and mouse models and toward zebrafish, which are cheaper, easier to maintain, transparent, and breed quickly. We now have COPAS systems that can handle zebrafish embryos, which means you can do high-throughput animal research.

“I think this opens up a whole new area of drug discovery that’s previously only been done manually. This is the first mechanism to automate animal screening—it’s a breakthrough for that. Once the zebrafish genome has been sequenced, then I think you’ll see this really take off,” states Green.

The company recently announced record high-throughput screening results of one million beads sorted in half a day. Using the COPAS system, the Carlsberg Research Center (Copenhagen) developed special beads (PEGA beads) that are neutrally buoyant in water. The beads are about the same size as fruit fly larvae.

The reason for the large beads is that a reasonable amount of compound is

needed on the beads to perform post-facto NMR analysis, which avoids the need to do structural determination for the whole library; it can be done just from the “hit.”

The company plans to develop additional applications in combinatorial chemistry and model organism screening for COPAS. “We have strong technological positioning at the emergence of model organisms in the high-throughput drug discovery field,” Green says.

### Combinatorial Biocatalysis

**Albany Molecular Research Inc.** (AMRI; Albany, NY) has developed a drug discovery technology called combinatorial biocatalysis, which combines miniaturization, robotics, and high-throughput techniques with biocatalysts, such as isolated enzymes and microorganisms. Specializing in preclinical discovery and development services, the biocatalysis division offers scientific resources for enzymatic, microbial, and chemoenzymatic synthesis and scale-up.

The company’s proprietary platform, BIOACTIV™ (Biosynthesis and Identification of Active Compounds Through Iterative Variation) integrates iterative enzymatic and microbial reactions to create libraries of derivatives from lead compounds.

“The advantages of our system,” explains Peter Michels, Ph.D., senior director of the biocatalysis division, “is the fact that you can directly derivitize already active lead molecules to create variations that would be difficult to create using traditional synthetic chemical methods.”

Other advantages of the system include selectivity and chemo-flexibility of biocatalysis to work with these complex lead molecules. Also, the system

allows different transformations, like hydroxylation, to be performed, which is difficult to do with traditional chemistry.

“The number of derivatives is less than with traditional combi-chem. However, the quality of derivatives, in terms of purity and structural complexity, is much higher in the libraries we are producing. Also, since each derivative is from an already active biological molecule, it makes it much more likely that each derivative we make is going to have biological activity,” explains Dr. Michels.

The BIOACTIV system has two steps: high-throughput screening of biocatalytic systems and the creation of derivatives. The first involves identifying biocatalysts that can produce a transformation of the lead molecule using some type of mass spectrometry in a miniaturized format. Once the biocatalysts have been identified, different chemistries can be performed on the lead molecule.

The second step involves applying each identified biocatalyst to make a first generation of derivatives (approximately a dozen molecules). Then multiple transformations can be performed simultaneously on the lead molecule to make further generations of derivatives.

This two-step process was developed in-house. “This is an adaptation of HTS biological screening, but instead of looking for inhibition of enzymes, which indicates biological activity, we’re screening enzymes with the lead molecules and we’re not looking for inhibition but new reaction products,” states Dr. Michels.

Another important advantage of biocatalysis is that it works under mild, ambient conditions (similar to living cells), so a wide variety of chemistries can be done simultaneously and in parallel—enabling rapid screening and analysis. It takes about six to eight weeks to make

hundreds to thousands of derivatives, which is a much more focused library than traditional combi-chem.

The company is currently collaborating with drug and biotechnology companies to apply this process to their drug leads, including the optimization of lead molecules, the expansion of structural activity relationships, the broadening of patent coverage, and the making of unique derivatives to improve the chances for success of their pharmaceutical development program. AMRI is also identifying its own drug targets and has internal programs in oncology, immunosuppressants, and protein therapeutics.

### Focusing on Biased Arrays

**ArQule, Inc.** (Woburn, MA) has integrated technologies to perform high-throughput, automated production of chemical compounds and to deliver these compounds of known structure and high purity in sufficient quantities for lead optimization. Its AMAP™ (Automated Molecular Assembly Plant) performs high-throughput chemical syntheses for each phase of compound discovery.

It is the foundation of the company’s parallel synthesis approach to combinatorial chemistry and consists of an integrated series of automated workstations and proprietary software that controls and monitors the overall chemical production process.

AMAP also permits synthesis of compounds biased toward a particular target. Biased Array™ libraries streamline lead generation and the qualification process by allowing collaborators to focus on compounds most likely to be effective against the target. “We’ve developed pharmacophore models that we think define all of the structural requirements for activity of a particular ligand. So, with

the ion-channel biased-array libraries, we have already developed pharmacophores for several of these channels.

We’ve identified some ligands that work against the calcium-N channel as antagonists. We’ve compared all those ligands effective and selective for the calcium-N type and concluded there are some basic substructures that are necessary for that activity, so we’ve created a model that encompasses all of the necessary structural features for activity,” explains Norton Peet, Ph.D., vp of discovery alliances.

“Once we have the model, we can put it on our intranet site so the design chemists can come up with ideas. They can input structure to this model and get an answer back as to how well it fits into that pharmacophore model.”

The calcium-N-type ion channel is important in cerebral ischemia, or stroke. The company thinks if it can develop a ligand that is an antagonist of that ion channel, it would prevent ions from passing through, and this could be very useful in the treatment of stroke.

Dr. Peet explains that there is a lot of nonautomated work in developing the pharmacophore model. “We develop ligands from studying literature and finding ones with an affinity for ion channels. Once the model is developed, the automation begins because the proprietary software allows the new structure to be compared to the pharmacophore to see how well it fits.” The software provides answers fairly quickly—it can be instantaneous or it can take a few hours, up to a day. The software will also be used for the creation of additional biased arrays.

The company, says Dr. Peet, is moving away from the production of large, random arrays toward the production of biased arrays. “The focus is on small

libraries in this industry. We're not the only ones moving in this direction. One reason is cost. If we can work smarter by screening smaller sets of molecules then it costs less. Also, if you're screening against a specific target in a random library, your 'hit' rate is only about 0.1 percent.

"The goal we're striving for is to increase the hit rate for a given target in our biased libraries 10-fold—up to 1.0 percent." He added that the genomics effort would be identifying new ion channels, and that there are already several successful drugs on the market that are ion-channel drugs. **GEN**