CAMBRIDGE HEALTHTECH MEDIA GROUP
Bio-IT World
Indispensable Technologies Driving Discovery, Development, and Clinical Trials

FORWARD THIS ISSUE TO A COLLEAGUE

SUBSCRIBE TO *BIO-IT WORLD* MAGAZINE

Welcome to your digital edition of *Bio-IT World* magazine!

We hope you are enjoying your subscription! As a member of the Bio-IT World community, you also have complete access to all online Archives, Buyer's Guides, Events, and Career Centers. Be sure to bookmark WWW.BIO-ITWORLD.COM for daily delivery of fresh insights and critical resources – or place Bio-IT World headlines on your desktop (or your web site) with our XML news feeds.

In addition, you can select from a wide variety of FREE enewsletters covering such targeted topics as Genomic Medicine, Drug Discovery, IT Trends, Regulatory Compliance, Technology for Healthcare and Clinical Trials. For a complete list of enewsletter titles, please go to http://www.bio-itworld.com/newsletters/sign-up.html

And as always, we welcome any and all comments or suggestions-just send us an email with DIGITAL SUBSCRIBER as the subject line.

>> CLICK HERE for technical assistance with your digital magazine
>> CLICK HERE for all other subscription customer service requests.

Bio-ITWorld reports the following digital circulation figures as verified by the Qmags Self-Audit procedures:

February 2008 Issue	
Total qualified digital subscribers:	6425
Biopharma/Chemicals	1919
Biotech/Bioinformatics	1811
Acad/Gov, Clinical Trial, CRO	1607
Hosp Lab/Healthcare, Other	1088

Dec/Jan 2008 Issue	
Total emails sent:	6390
Undeliverable :	100
Total delivered:	6271
Removed:	19

NAVIGATION KEY

- Previous/Next Page: mouse click (or keyboard) on left/right arrows to turn page
- Contents: takes you to the Table of Contents page
- Zoom In: top will zoom in to the top of page & bottom will zoom in to the bottom of page
- Zoom Out:takes you from single to double page view
- Front Cover: takes you to the Cover Page of magazine
- Search Issue: allows full search by keyword or phase (can also search archives!)

Alternative Thinking About The Size of Blade Systems:

CAREFUL, DON'T TRIP OVER THE INFRASTRUCTURE.

Strength and power, now in a smaller, more convenient size. Introducing the HP BladeSystem c3000. All the technology of our larger BladeSystem in an efficient, compact, affordable package. Careful, watch your step.

Technology for better business outcomes.

Powered by the Quad-Core Intel® Xeon® Processor[1]

Get down to the smallest details. Visit hp.com/go/nocompromise69
1-866-625-1019

1. Intel, the Intel logo, Xeon, and Xeon Inside are trademarks or registered trademarks of Intel Corporation in the U.S. and other countries.

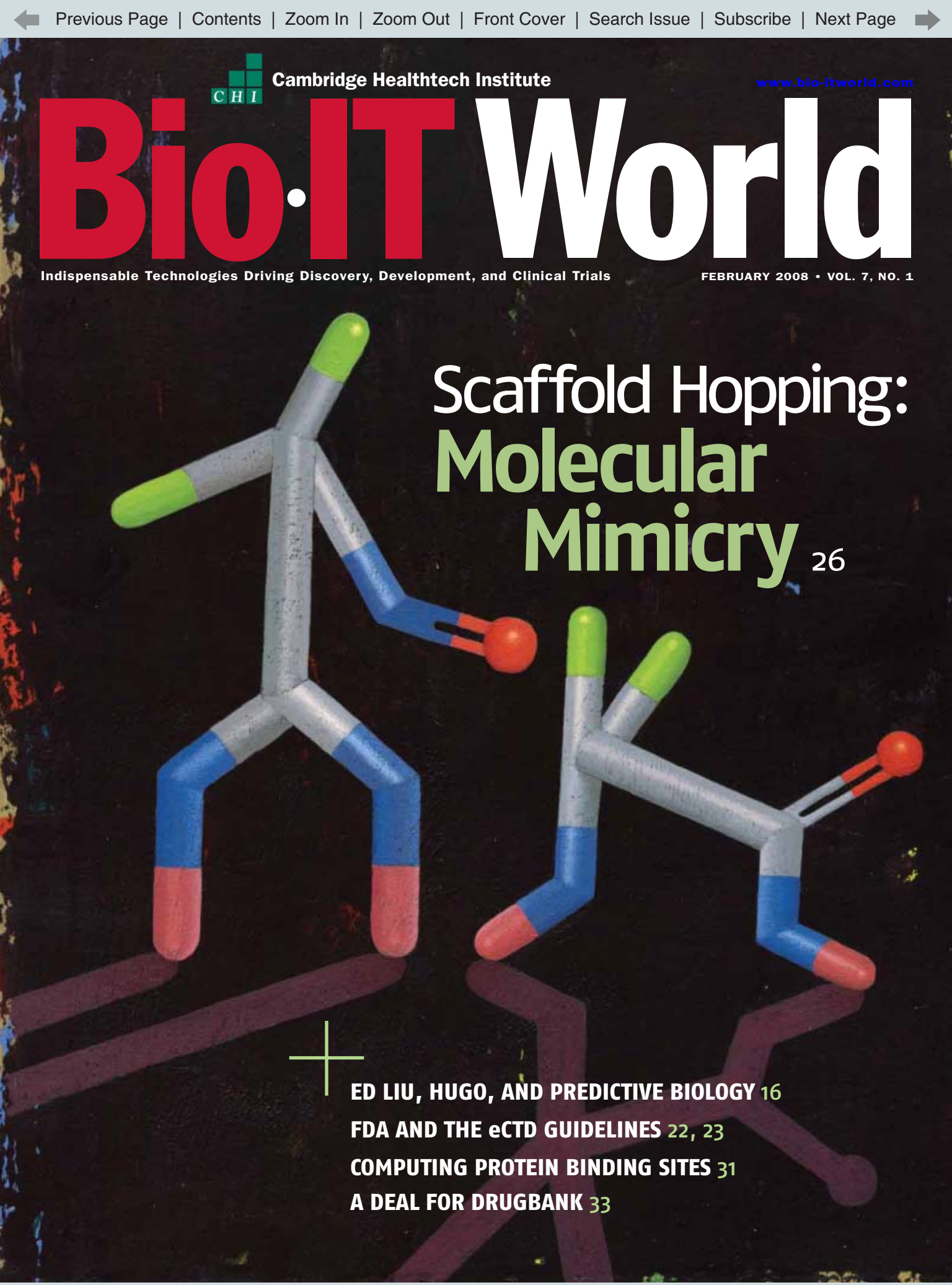
Cambridge Healthtech Institute

www.bio-itworld.com

Bio·IT World

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

FEBRUARY 2008 • VOL. 7, NO. 1



Scaffold Hopping: Molecular Mimicry 26

Visit the Thermo Scientific Exhibits at Pittcon 2008 at **Booths #1735 and 1741**

Count on Us.

Over 1500 customers worldwide already count on Thermo Scientific informatics solutions to manage one of their most important assets – laboratory data. These industry leaders have turned to Thermo Scientific LIMS and CDS software in their efforts to make the world healthier, cleaner and safer.

If you are looking for industry experience, domain expertise, innovative technology and world class after-sale support, you can count on us to meet your needs now, and well into the future.

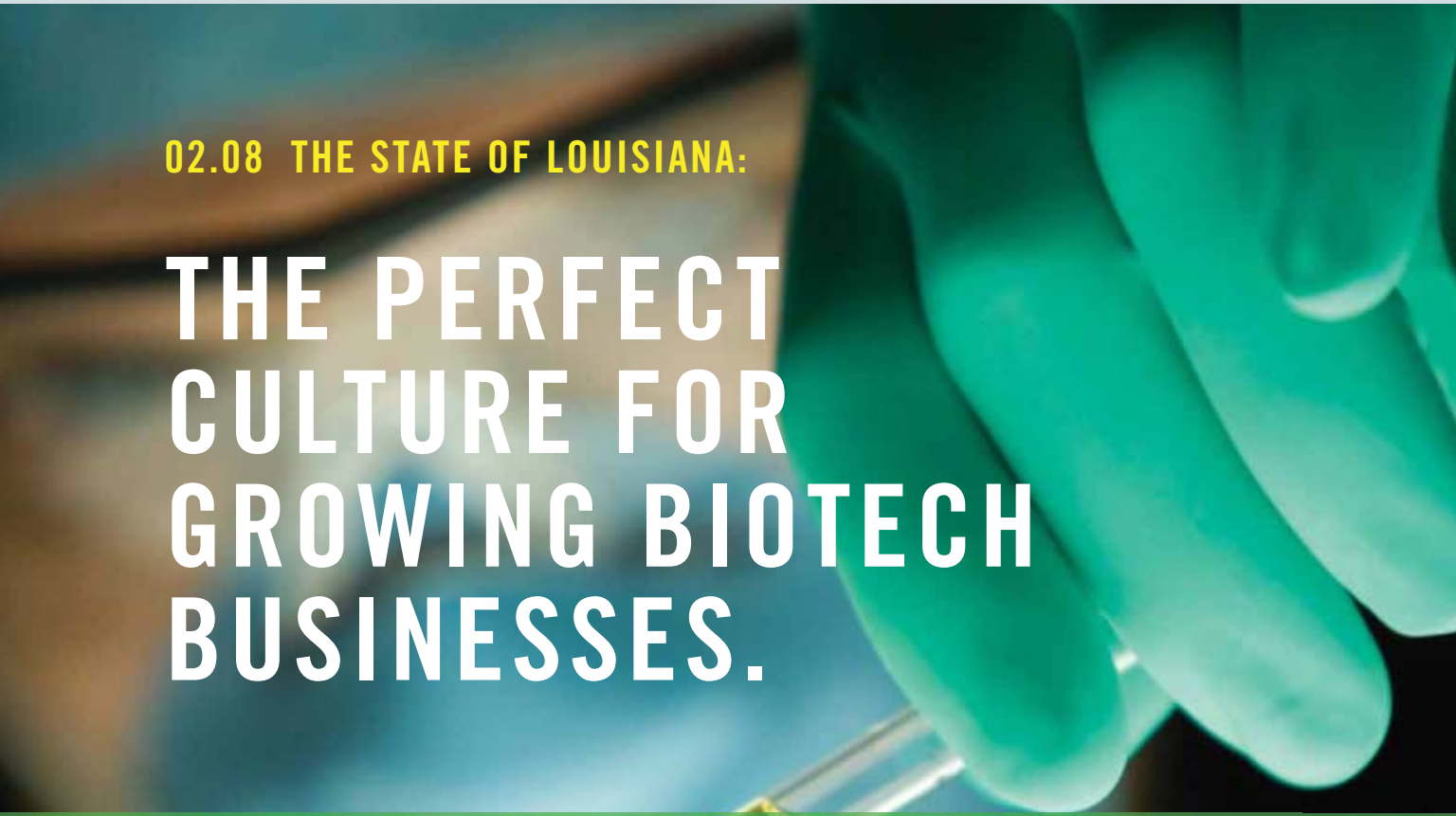
Find out which Thermo Scientific informatics solution is right for you. Visit **www.thermo.com/solutionfinder**.

+1 866-463-6522 • marketing.informatics@thermofisher.com

We have a solution for you, no matter what industry you are in:
pharmaceuticals, food and beverage, industrial chemicals, petrochemicals, metals and mining, water and wastewater and many many more.

Part of Thermo Fisher Scientific

Thermo SCIENTIFIC



02.08 THE STATE OF LOUISIANA:

THE PERFECT CULTURE FOR GROWING BIOTECH BUSINESSES.

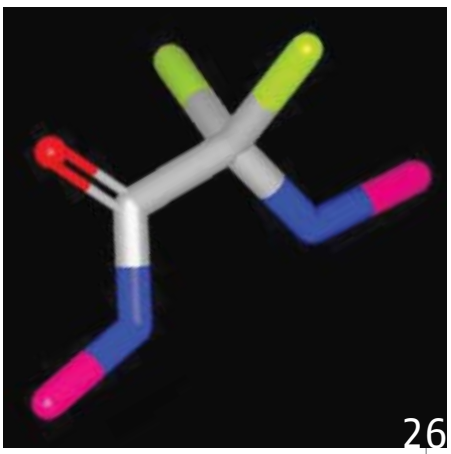
▶▶ Pennington Biomedical Research Center, a world leader in nutrition and disease prevention research, is investigating genetic factors that may contribute to eating behaviors, body weight and binge eating ▶ NuPotential, located at the Louisiana Emerging Technology Center, has developed a revolutionary approach to reprogramming somatic cells into pluripotent stem cells ▶ Esperance Pharmaceuticals is developing a unique targeted anticancer drug that selectively kills cancer cells on contact ▶ Embera NeuroTherapeutics, based at InterTech Science Park, has developed patent-pending treatments for addiction and a wide array of neuropsychiatric disorders ▶ BioFluidica Microtechnologies is developing on-site DNA testing instruments for use in forensics, medical diagnostics and pathogen detection ▶ Louisiana's Angel Investor Tax Credit allows refundable Louisiana income or franchise tax credits of up to 50% of investments ▶ The Tulane Center for Gene Therapy is the first research facility in the nation dedicated solely to using a patient's own donated stem cells

LOUISIANA ECONOMIC DEVELOPMENT

To learn more or to find out about valuable Gulf Opportunity Zone Incentives, call Bob Fudickar at 225.342.6816 or visit LouisianaForward.com/Bio-ITWorld

Contents [02·08]


16


26


10

Cover Story

The Search for Unusual Suspects

Scaffold hopping expands range of core molecular shapes for lead generation. By Vicki Glaser 26

Cover illustration by Martin Jarrie

TABLET TECHNOLOGY YOU CAN CLICK AND TOUCH

RESPONSIVENESS AND ACCURACY YOU CAN DEPEND ON.

The new Dell Latitude™ XT notebook lets you capture information easily by both pen and touch. All while the advanced palm-rejection technology helps eliminate frustrating interference. And it all starts at less than 4 pounds*.

SIMPLIFY YOUR IT AT
DELL.COM/Responsiveness or call 1.866.209.3355

DELL

Company Index

Advertiser Index

This index is provided as an additional service. The publisher does not assume any liability for errors or omissions.

VOLUME 7, NO. 1

Editorial, Advertising, and Business Offices: 250 First Avenue, Suite 300, Needham, MA 02494; (781) 972-5400

Bio-IT World (ISSN 1538-5728) is published monthly except July/August and December/January by Cambridge Bio Collaborative, 250 First Avenue, Suite 300, Needham, MA 02494. *Bio-IT World* is free to qualified life science professionals. Periodicals postage paid at Boston, MA, and at additional post offices. The one-year subscription rate is $199 in the U.S., $240 in Canada, and $320 in all other countries (payable in U.S. funds on a U.S. bank only).

POSTMASTER: Send change of address to Bio-IT World, P.O. Box 3414, Northbrook, IL 60065. Canadian Publications Agreement Number 41318023. CANADIAN POSTMASTER: Please return undeliverables to Station A, P.O. Box 12, Windsor, ON N9A 6J5.

Subscriptions: Address inquires to Bio-IT World, P.O. Box 3414, Northbrook, IL 60065 (888) 835-7302 or e-mail biw@omeda.com.

Bio·IT World®

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

EDITOR-IN-CHIEF
Kevin Davies (781) 972-1341
kevin_davies@bio-itworld.com

EXECUTIVE EDITOR
John Russell (781) 972-1342
john_russell@bio-itworld.com

ASSOCIATE MANAGING EDITOR
Allison Proffitt (781) 972-1345
aproffitt@healthtech.com

ART DIRECTOR
Mark Gabrenya (781) 972-1349
mark_gabrenya@bio-itworld.com

WEB EDITOR
Catherine Varmazis (781) 972-1344
catherine_varmazis@bio-itworld.com

DIRECTOR, BUSINESS DEVELOPMENT
Angela Parsons (781) 972-5467
aparsons@healthtech.com

VP SALES – CA, WESTERN US, CANADA, PACIFIC RIM
Alan El Faye (213) 300-3886
alan_elfaye@bio-itworld.com

REGIONAL SALES MANAGER – NEW ENGLAND, NORTH EASTERN US
Kay O. Christopher (860) 693-2991
kchristopher@healthtech.com

REGIONAL SALES MANAGER – MIDWEST, SOUTH EASTERN US, EUROPE
Tom Loughran (781) 972-5482
tloughran@healthtech.com

SENIOR DIRECTOR OF MARKETING & OPERATIONS, PUBLICATIONS
Joan A. Chambers (781) 972-5446
jchambers@healthtech.com

CIRCULATION & MARKETING MANAGER
JoAnne Sperry (781) 972-1350
jsperry@healthtech.com

PROJECT/MARKETING MANAGER
Lynn Cloonan (781) 972-1352
lcloonan@healthtech.com

ADVERTISING OPERATIONS COORDINATOR
Kathrene B. Tiffany (781) 972-5440
ktiffany@healthtech.com

GRAPHICS AND PRODUCTION MANAGER
Ann-Marie Handy (781) 972-5493
ahandy@healthtech.com

Contributing Editors
Michael Goldman, Karen Hopkin, Deborah Janssen, Pauline Parry, Salvatore Salamone, Tracy Smith Schmidt, Mark D. Uehling

Advisory Board
Jeffrey Augen, Mark Boguski, Steve Dickman, Kenneth Getz, Jim Golden, Andrew Hopkins, Caroline Kovac, Mark Murcko, John Reynders, Bernard P. Wess Jr.

Cambridge Healthtech Institute
PRESIDENT
Phillips Kuhl

Contact Information
editor@healthtech.com

250 First Avenue, Suite 300
Needham, MA 02494

You Stand in Front of Our Instruments All Day... We Stand Behind Them...24x7

High Value, Substantially Expanded Services from Applied Biosystems

With over 25 years of experience in the development, manufacture, and service of innovative instruments, and with over 1000 highly trained, dedicated service professionals, AB Global Services is uniquely qualified to deliver a full suite of real-world lab services. From Remote Services to On-Site Application Consulting to Qualification and Professional Services and everything in between, AB Global Services is your value-added partner to help boost your productivity and maximize the return on your technology investment.

- Instrument Repair and Maintenance
- Smart Services
- Qualification Services
- Professional Services
- On-Site Application Consulting
- Training

To learn more go to http://info.appliedbiosystems.com/wwservice or contact your local Applied Biosystems sales representative.

Applied Biosystems

First Base

2008: Change is Coming

KEVIN DAVIES

You know the pharmaceutical industry is in trouble when the CEO of one of the top ten big pharmas decides to abandon his 44th-floor midtown view of Manhattan for cheaper digs on a lower floor. But that's what Bristol-Myers Squibb (BMS) boss James Cornelius says he will be doing this year.

Some pharma CEOs may think that a bit drastic, but 2007 was another mediocre year for the industry. Just 19 new drugs were approved by the FDA, continuing a bleak run of productivity with few signs of an imminent upswing (See, Can Academics Save Pharma? *Bio•IT World*, Nov. 2007; Inside Andy Grove's World of Bio-IT, *Bio•IT World*, Dec. 2007). But amid the gloom, there are — as there usually are — signs of genuine hope and innovation. From our standpoint, they exist in the fruits of basic research and in the output of dynamic new drug companies, as well as signs that big pharma is tackling its many problems head on.

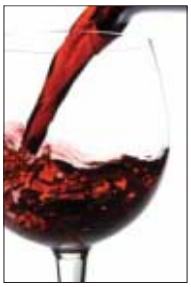
So as we kick into 2008, here are a few silver linings that have caught our eye, trends that suggest there is more reason for hope in the bio-IT arena than ever before.

Complex Genetics: 2007 was a watershed year for the study of complex, multigenic diseases. To recap the list of gene mapping breakthroughs for common diseases that affect most Americans would fill this column. And 2008 was barely a week old before further exciting revelations of a genetic glitch associated with autism. As researchers continue to tease out mutations and copy number variants associated with common diseases, look for better insights into clinically relevant pathways and ultimately new therapeutics.

Biologics Bonanza: 2007 was "the year pharma fell in love with biologics," according to The Motley Fool. And when AstraZeneca forks over a whopping $15 billion for a leading biologics company, MedImmune, one can only agree. It's not just the favorable patent situation surrounding biotech drugs, however. As the recent successes of Genentech show, biologics can be very potent weapons against cancer and other diseases. BMS, Pfizer, Novartis and others are all expanding in this area.

Structure Prediction: One of *Science* magazine's "Breakthroughs of the Year" was the long-awaited crystal structure of the beta-adrenergic receptor. For the family of G-protein-coupled receptors, that makes two structures down, a mere 998 to go! Hence the sharp interest in protein folding predictions. 2007 saw a spectacular paper from David Baker's group (See, Improving Structure Prediction, *Bio•IT World*, November 2007), pointing to major progress in this field. (We will spotlight a stunning drug design application of similar methods in next month's issue.)

Red, Red Wine: As someone who is partial to a spicy Shiraz every now and again, it's good to know that the occasional indulgence might have health benefits. *60 Minutes* said as much a few years ago with a report on the Mediterranean diet. Now Sirtris Pharmaceuticals, which went public last year, is demonstrating remarkable properties among derivatives of the active ingredient in red wine, a compound called resveratrol. It's a reminder of the remarkable innovation at the heart of the biopharma industry.

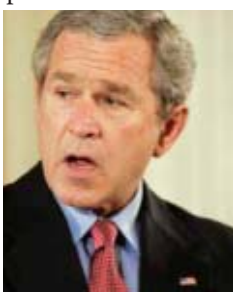


Far Out FDA: As reported elsewhere in this issue (See p. 22), the FDA's new mandate regarding electronic clinical documentation went into effect January 1st. While this might cause short-term pain for some biopharmas, it can only help the industry in the long term. This and other initiatives, including genetic testing information on drug labels (See, Predicting Dose Response, *Bio•IT World*, September 2007), show some welcome far-sightedness at the FDA.

Stem Cells: In a science story that truly deserved to be front-page news, Japanese and American researchers reported success in reprogramming normal human cells to become stem cells, simply by introducing four genes. This does not obviate the need for continued embryonic stem cell research or excuse the White House's smug reaction to the news, but it ensures that stem cell research has new, exciting avenues to explore.

Predictive Technologies: The FDA's growing interest in predictive technologies is evidenced by R&D agreements with informatics and *in silico* powerhouses Ingenuity and Entelos. These items, together with a growing body of papers in the literature, suggest both growing interest in and near-term payoff from predictive approaches.

Consumer Genomics: From 23andMe "spit parties" to the launch of the first full-genome sequencing services, 2007 was the year that personal/recreational/consumer genomics became a real entity. How will the public respond to these new services? Or physicians? How will the industry be regulated? And how much closer will we be to the $1000 genome by the end of 2008? Stay tuned...



Picking 44: And then there's the small matter of this year's U.S. presidential election. No endorsements here, but the prospect of a more tolerant and enlightened view of science and medical research cannot come soon enough.

WINE: DREAMSTIME.COM; BUSH: BRENDAN SMIALOWSKI/AFP/GETTY IMAGES

Cambridge Healthtech Institute's Seventh Annual...

World Pharmaceutical Congress

Assays for Efficient and Safe Drug Development

Conferences & Exhibits May 12-14, 2008 • Loews Philadelphia Hotel • Philadelphia, PA

MAY 12-13

Cell-Based Assays for HTS

Cardiotoxicity and Drug Safety

Targeting Pain with Novel Therapeutics

MAY 13-14

Biochemical-Based Assays for HTS

Hepatotoxicity and Drug Safety

Novel Approaches to Lead Optimization

Corporate Sponsor:

SCHRÖDINGER

Lead Sponsoring Publications:

Bio-IT World

Make sure to mention key code U02 when registering!

Cambridge Healthtech Institute
250 First Avenue, Suite 300, Needham, Massachusetts 02494
Telephone: 781-972-5400 or toll-free in the U.S. 888-999-6288
Fax: 781-972-5425 • www.healthtech.com

WorldPharmaCongress.com

GATC Biotech Offers Increased Platforms, Productivity

German sequence contractor deploys three rival next-gen systems.

BY KEVIN DAVIES

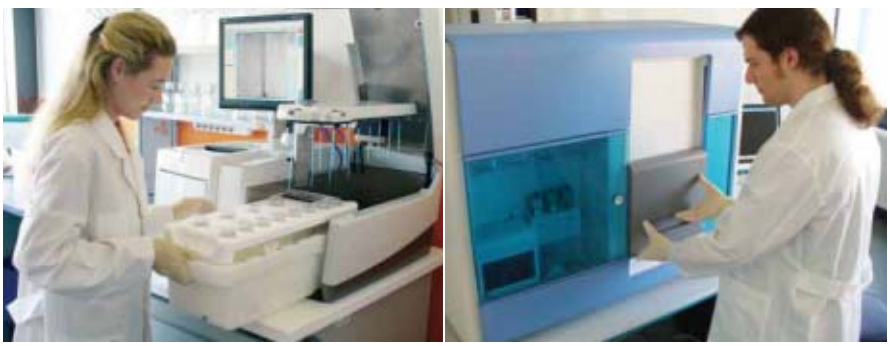
Many executives would envy GATC Biotech CEO Peter Pohl. His company, based in Constance, Germany, near the Swiss border, has enjoyed a 200-fold increase in productivity in the past 12 months, courtesy of the boom in next-generation sequencing systems.

Pohl co-founded the high-throughput sequencing firm in 1990, with his two brothers and father, a former professor of molecular biology at the University of Constance, who originally conceived the idea. The Pohls launched the company with a mere $12-13,000 in starting capital. "It's possible!" says Pohl.

During the past 15 months, Pohl says GATC has installed "all three validated sequencing systems" — Roche, Illumina, and Applied Biosystems' (AB) SOLiD — along with what he calls the "gold-standard" 3730 instruments from AB. "We have all four leading technologies under one roof," says Pohl. "We are the only commercial sequencing provider to my knowledge that can offer all these."

GATC began life as a service company for research groups and industry. The firm's first contract was a fragment analysis project for a big pharma for almost $37,000. Today, Pohl says that same work would be around $588! The next contract was for the European Union, and GATC has been profitable ever since. Much of its work has centered on microbial genome sequencing, having worked on about 100 projects. Another major focus is the potato genome project, still using standard shotgun cloning and Sanger sequencing (with four AB 3730 instruments).

"Speed is taking off," says GATC founder and CEO Peter Pohl.



GATC Biotech has added the SOLiD system to compliment instruments from 454 (left) and Illumina.

By embracing new sequencing technologies, GATC has boosted its annual sequence output to some 350 gigabases. The company offered its first 454 sequencing in November 2006, installing the machine in January 2007, followed by an Illumina 1G Analyzer in early 2007, and an AB SOLiD machine last November. "We're still evaluating the different technologies," says Pohl, adding that it's unlikely GATC will focus on one single system. "We'll use the advantages of all the technologies," he says (see "Data Analysis").

Looking ahead, he says, "The whole era of single molecules sequencing technology will be very interesting. Speed is really taking off." Purchasing a HeliScope is a possibility. "We wouldn't be one of the leadings services providers in the world if we weren't considering it," says Pohl.

This year, Pohl expects GATC to work on a pair of disease-related genomics projects sequencing up to ten full human genomes.

Pohl says priorities for the instrument makers must be to reduce costs in chemistry and increase capacity. "It comes down to price per base pair," says Pohl. "Then we have to see which technology is progressing with the most professional team in order to get capacity up, costs per base down, and the best technical support."

Back in 1990, Pohl says cost of sequencing a base pair was $25. "Today, the cost per base pair is definitely less than 0.1 cent/base pair. This is 30,000 times less than just 16 years ago! If we look forward another 16 years, we come to a price for onefold coverage of a human genome of [about $99]." Within a decade, Pohl firmly predicts that whole genome sequencing for medical applications will be less than $1000. But ensuring sufficient data quality to use as a diagnostic tool is a key question. •

Data Analysis

Christopher Bauser, GATC's head of bioinformatics, sees first hand the strengths and weaknesses of the new sequencing platforms.

Christopher Bauser

"The 454 system has relatively long read lengths, so it's an advantage for *de novo* sequencing," says Bauser. "Coupled with paired end reads, it's a very powerful system for assembly. But the number of reads or bases is significantly less than the SOLiD or Illumina systems per run. The cost of a single run on each of the systems is about the same, but a [454] GS FLX run is only 7–10 hours, whereas the Illumina GA is 3 days. SOLiD is also 3–4 days. The costs/base for SOLiD and Illumina are probably lower in that regard, but the reads are relatively short."

Meshing data output from different platforms "is an issue, but it's not a problem," says Bauser. "Most of the analysis methods we're using work just fine with text files or FASTA formats, it's a trivial little program that we use to transform sequencing output into these text files, which we can analyze with standard software."

Bauser says the method of analysis hinges on the underlying technology. "If most of the data we're analyzing for a specific project is in color space, then we use programs best suited to color space, and just translate any Sanger sequences we have into color space/FASTA format." In such cases, it often makes sense to use one of the AB software packages, "and translate other technologies into color space." Roche and Illumina likewise provide analysis software that is suited to their specific technology, but also compatible with standard data formats.

Nor is handling the data output proving too difficult — at least for now. Storing all the image files would be "a huge amount of data, but the information in the pictures, as soon as they've been analyzed, is superfluous. So eight hours after the run is finished, you can dispose of about 90% of the data you've generated. The sequences and quality scores need to be saved, and that's going to be a problem eventually. But at the moment, it's reasonable to keep these and just get rid of anything else."

Some of GATC's 60 employees are heavily involved in developing software tools to help render the information into a form that scientists can analyze quickly. "There's an entire zoo of programs that Ph.D. students and postdocs have been writing to analyze this information. The next big thing is going to be assembling all of these into a small set of analysis programs that will be suitable to do the different types of sequencing project — *de novo* sequencing, resequencing, transcriptome research, and so on." — KD

PREVIEW: Bio-IT World Conference & Expo 2008

Bio-IT World's seventh annual conference takes place on April 28-30, in Boston. Here we select just a few of the potential highlights from this year's expanded and keenly anticipated program.

Entry to the keynote presentations and exhibit hall is free with advance registration. Conference registration and full program details available at: www.bio-itworldexpo.com.

The Billion Dollar Man

Joshua Boger, founder and CEO of Vertex Pharmaceuticals, is known to many as one of the central characters in Barry Werth's seminal account of the drug discovery industry, *The Billion Dollar Molecule*. For nearly two decades, Boger has led a thriving biopharmaceutical enterprise amidst a maelstrom of shifting technologies and regulations.

Keynote Lecture | April 29

Consumer Genomics ... Discuss

With at least three companies launching full personal genome scans for $1,000 or thereabouts, the era of consumer or recreational genomics is here. In this plenary panel, guests including George Church (Harvard), Dietrich Stephan (Navigenics), Fred Ledley (Bentley College), and Jeff Drazen (New England Journal of Medicine) will discuss the latest technologies, the early response from the public, the limitations of genotyping, and the medical and ethical challenges of bringing personal genomics into the medical mainstream.

Plenary Panel | April 30

Embracing Semantics

In this opening workshop, Alan Ruttenberg (principal scientist, Science Commons) discusses ways to embrace the Semantic Web in your organization. Ruttenberg will discuss basic components of the Semantic Web, present case studies, discuss tools and partner organizations, and concrete advice on specific challenges that delegates may have.

Pre-Con Workshop | April 28

EDC Under Examination

The centerpiece of this year's Clinical and Medical Informatics track is a panel discussion dissecting the current and untapped uses of electronic data capture. Moderated by Ron Waife, this exchange will highlight recommendations for companies to shorten development timelines and accelerate decision-making — opportunities that Waife says are not being fully utilized in the industry.

Conference Track #5 | April 29

Benjamin Franklin Award 2008

For the fourth consecutive year, Bio-IT World Expo serves as the host of the Bioinformatics Organization (Bioinformatics.Org) annual Benjamin Franklin Award. The winner, who will be chosen from a shortlist of six finalists (See p. 19), will be presented with this prestigious award and deliver a guest lecture. Past winners include Sean Eddy (HHMI/2007), Michael Ashburner (Cambridge/2006), Ewan Birney (EBI/2005), Lincoln Stein (Cold Spring Harbor/2004), James Kent (UCSC/2003), and Michael Eisen (Berkeley/2002).

Franklin Award | April 29

23 and Linda

Few companies attracted as much media attention last year as 23andMe, the consumer genomics start-up with ties to Google. Co-founder Linda Avey conceived the outline of a personal genomics company during her years at Affymetrix and Perlegen, before collaborating with Anne Wojcicki to launch the company. Finally out of stealth mode, 23andMe launched a $999 whole genome scan test last November. In one of her first public lectures, Avey will discuss the genesis of 23andMe and the personal genomics revolution.

Keynote Lecture | April 30

Systems Biology in Practice

Systems biology's evolution from theory to practice is spotlighted in two sessions that promise to be fascinating and important. In *Model-Simulated Design of Cancer Therapies*, Merrimack Pharmaceuticals R&D chief Ulrik Nielsen will present the company's use of detailed mechanistic models to advance cancer therapeutics. Merri-

Nielsen

Briefs

mack is one of the few pharma companies formed specifically to leverage systems biology approaches and now has candidates in early trials. In *Systems Approach to Drug Development: Bio-simulation and Bio-mathematics,* panelists will explore how mathematical approaches are being applied in drug discovery and personalized medicine. Panelists include pioneers from the systems biology technology community and early adopters from big pharma, including: Dinesh DeAlwis (Lilly & Co.), Zvia Agur (Optimata), David de Graaf (Pfizer), and M. Vidyasagar (Tata Consultant Services).
Conference Track #3 | April 29

Best Practices 2008 Awards
After a two-year break, *Bio•IT World*'s Best Practices Awards are back! The winners will be announced at a special gala reception and dinner held on the second evening of the event. Hear from our special guest speakers, meet the editors, and come congratulate the winners of this prestigious event. (Tickets are available for purchase to anyone registering for the conference at www.bio-itworldexpo.com)
Best Practices | April 29

Next-Gen Data Management
How many times last year did excited scientists plug in a new instrument from Illumina or 454, only to be confounded by the volume of data pouring out? Well, help is at hand. In this must-attend pre-conference workshop on next-generation sequencing management, members of the BioTeam will share their insights from numerous engagements over the past 12 months in industry and academia to create solutions to the torrents of next-gen sequencing data. Guest speakers include George Church (Harvard Medical School/Knome), Dick McCombie (Cold Spring Harbor), as well as several vendor representatives.
Pre-Con Workshop | April 28


Church

John, Johnson & Johnson
Following high profile IT and informatics positions at Celera and Eli Lilly, John Reynders has recently assumed the position of CIO for Johnson & Johnson's

Pharma R&D division. Last month Reynders discussed with *Bio•IT World* his new responsibilities (See, Reynders Takes the CIO Reins, January 2008). In this opening keynote speech, he will analyze the critical technological and organizational issues facing not just big pharma but the bio-IT community at large.
Keynote Lecture | April 28

Chemical Biology and Therapeutics
Although best known for its work in genome analysis, the Broad Institute has a dynamic chemical biology program, performing high-throughput screening on chemical libraries built using diversity-oriented synthesis in the search for potential lead compounds. Much of the data are publicly available through the group's website, Chembank. David DeCaprio
Conference Track 4 | April 29

User Groups
Bio-IT World Expo once again plays host to some prestigious user group meetings that can only enhance the overall value of attendees' experience. As in previous years, Oracle will host a pre-conference life sciences and health care user group meeting (its 9th annual), kicking off on April 27th. This year, we are also delighted to partner with the Symyx Software Symposium — that meeting is co-located with the Bio-IT World conference, and begins on April 30.
Oracle Life Sciences User Group | April 27–28
Symyx Software User Group | April 30–May 2

BIOMARKER VACUUM
Researchers at the **George Mason University Center for Applied Proteomics and Molecular Medicine** have announced a nanotechnology tool that may make protein biomarkers easier to identify. The smart hydrogel nanoparticles mix with a patient's blood sample to "suck up" protein biomarkers from blood and preserve, protect, and stabilize the molecule. Scientists are currently applying the technology to cancer and other diseases.

FINANCING SUCCESS
Genizon BioSciences has closed its Series E financing to the tune of approximately $31 million. The new investment will be partially used to fund additional genome-wide association studies in the four major diseases associated with metabolic syndrome: obesity, type II diabetes, dyslipidemia, and hypertension. The lead investor is BTF B.V. of Haarlem, The Netherlands.

GENETIC LINK FOR AUTISM
The Autism Consortium used the new **Affymetrix** 5.0 chip to scan the genomes of more than 3,000 individuals, about half of which have been diagnosed with an autism spectrum disorder (ASD). The scan resulted in compelling evidence that a region on chromosome 16 appears to play an important role in ASD susceptibility. The findings were verified by independent observations at **Children's Hospital Boston** and **deCODE Genetics**.

LAWSUIT SETTLED
Illumina has settled all ongoing lawsuits with a $90 million payment to **Affymetrix**. Both firms have agreed not to sue for use of current and future products. The litigation surrounded Affy's claim that Illumina's products infringed on patents, and were originally filed in the U.S., U.K., and Germany in 2004.

OSINT and the Pharmaceutical Enterprise III

JAMES B. GOLDEN

In my previous two columns, I discussed the application of Open Source Intelligence (OSINT) to improving pharmaceutical IT and business analysis, and outlined some tools and techniques for building a pharmaceutical intelligence capability. In the final column in this series, I discuss how to make intelligence actionable, with a real example.

Useful intelligence needs to be actionable. What's the point of searching for something if you don't do anything with it? It's not about search, it's about find — and doing something once you find it. Useful intelligence needs to be relayed to decision makers who can use it to improve the business, make decisions about a marker or compound, or solve a regulatory question.

Once you create a repeatable system for answering your own or management's questions, you may be asked to create all manner of regular reports. These reports could include information organized by category including:

- **Indications and Warnings** identifying potential actions with the goal of providing sufficient warning to preempt or counter their outcome (a bad press release, a competitor's announcement, the cancelation of a clinical trial, a warning letter from the FDA, etc.).
- **Current Intelligence** including the integration, evaluation, analysis and interpretation of information regarding persons, products, organizations, competitors, and regulatory bodies.
- **Targeted Intelligence** including analysis of specific identified groups or persons, company-related assets (physician's groups, principle investigators, manufacturing facilities, drug delivery systems, etc.), or vulnerabilities for exploitation. Identifying and monitoring trends in off-label prescriptions is an excellent example.
- **Scientific and Technical Intelligence** that directly effect projects in your pipeline. These intelligence products cover technical capability; scientific knowledge related to drug creation, license, and use of patents and intellectual property; engineering specifications; medical and biological methods and platforms; etc.

One of the most interesting uses of our OSINT system has been around gauging "sentiment." Traditionally, OSINT analysts have produced intelligence to help the Department of Defense (DoD) understand sentiment — how local language news sources and government officials regard the United States and its Armed Forces. From a life sciences perspective, this approach can be used for the creation and calculation of scientist's, thought leader's or organization's feelings or emotional response as manifested by descriptions in open sources and changes in rhetoric or opinion.

For example: does the totality of the biomedical literature really think your mechanism of action is a good candidate for drug discovery? Has sentiment for a particular biomarker's utility changed over time? Is it still a good bet for investment? How do I quantify and report it to management in an understandable and actionable way? We often produce trend reports for sentiment over time that can be viewed through a dashboard portal. Our partners get access to the underlying trend data and quantitative metrics, plus we put some additional analysis on top.

A Pharmaceutical OSINT Tool Kit

So, now we have a pharmaceutical OSINT tool kit. This workbench includes a solid grounding in intelligence philosophy about how to answer questions, and a set of tools and techniques that could include:

1. OSINT search and content extraction tools (RSS aggregators, Web crawling, and scraping);
2. Content classification and tagging tools;
3. Categorization and clustering tools;
4. Entity and relationship extraction tools;
5. Taxonomy/Ontology creation and management tools;
6. Presentation software (including visualization, network descriptions);
7. Desktop search engines (personal knowledge refinement by individual analysts);
8. Analytics tools (increased chatter, number of mentions, sentiment, opinion).

OSINT is about finding information to answer questions in public sources.

Let's imagine a real-world problem to test this approach. Imagine you're a member of a product development team around imaging biomarkers. Your company, Global MegaPharm, wants to identify imaging biomarkers that will help validate a therapeutic product in the company's pipeline and could also be a product in and of itself. Among the burning intelligence questions you might want answered include:

1. What are the subject-specific keywords for internal searching to see what's been done to date within our own company?
2. What are the relevant journal abstracts and papers regarding the putative biomarker?

3. What are some of the trends around this research, e.g. volume of papers, ranking of papers by citations, author status, gauging of "sentiment?"
4. Can we identify the source(s) of cutting-edge research from the authors' affiliations?
5. Can we rank those institutions and universities by their relevance to our project?
6. Can we identify thought leaders and departments within those universities?
7. How can we identify upcoming seminar topics and speakers?
8. Can we spot trends and predict future developments around our putative biomarker?
9. With whom are those thought leaders working?

If done right, your OSINT tool kit could help you assemble a workflow for answering these questions quickly and efficiently. Using our pharmaceutical OSINT approach you could:

1. Identify relevant (if not seminal) academic papers and reviews for understanding the imaging biomarker;
2. Use automated entity and relationship extraction software to create a keyword list;
3. Use this keyword list to identify PubMed abstracts and full-text articles for further investigation;
4. Run entity and relationship extraction software on this new corpus of information;
5. Review keyword lists for accuracy and completeness;
6. Use revised keyword lists to pull information from open sources utilizing deep web content mining (journal articles, news sources, meeting abstracts, blogs, government grants);
7. Expand keyword lists into a taxonomy or ontology for improving enterprise-wide search. Your computational biology group should be overjoyed to receive a well constructed biomarker ontology from the product teams!
8. Calculate trends including number of papers, thought leaders, top institutions, and key concepts to identify sentiment;
9. Cluster people, institutions and concepts based on trend numbers and relationships to each other;
10. Create relationship maps between imaging biomarker concepts, people and places;
11. Identify additional intelligence and drill deeper into individual parts of the process;
12. Deliver intelligence to your team on a regular basis that includes updated keywords, OSINT sources, trend metrics, leading industry and academic researchers.

Intelligence is about answering questions. OSINT is about finding information needed to answer those questions in public sources. While search is an important component, it's not the only component for creating actionable intelligence that will help you improve the business and drive your organization forward.

Jim Golden is a CTO at SAIC. He can be reached at goldenj@saic.com.

Computational Biology

From Genomes to Systems

Edison Liu talks about Singapore, HUGO, and predictive biology.

Edison Liu is Executive Director of the Genome Institute of Singapore, in many ways the flagship institute of the country. Liu manages the scientific direction, recruitment, and the budget for the entire institute. The institute employs some 260 people full time. Fifty percent are Singaporeans and 50% are foreign. This year, Liu adds president of HUGO to his many responsibilities.

Bio•IT World*'s Allison Proffitt visited Liu in Singapore and asked him about the big picture of research in his institute, his country, and his new organization, HUGO.*

BART NAGEL PHOTOGRAPHY

On Singapore and the Genome Institute of Singapore

BITW: What is your focus moving forward for GIS?

Liu: Our focus has been and will continue to be genome to systems, using transcriptional regulation and the transcriptome as the fundamental organizing principles. From there we seek a deep integration of technologies with biology that allows us to address very important biological questions that span from fundamental laboratory discoveries all the way to population studies. So that has been our leitmotif from the beginning and it's continuing. The beauty, though, is that we have concrete examples of how that has worked extraordinarily well and that this integration has given us greater insights in biology and genomics than we normally would if people were working alone.

How is GIS different from other institutes?

First, we're selective for people who want to do integrative science. Second, people are recruited for very disparate skills, whether you are technology or informatics or biology. You hope to think in biological endpoints, but you come from very different orientations. The challenge is that you don't always have the same assumptions when you get into relationships or you get into a project. There we've set systems whereby individuals have degrees of freedom as individuals. But their advance is dependent on joining others. We have a process whereby if you want to do something new with new resources, you need to present that idea to the collective, your community. Everybody is welcome to come and listen. They comment and ultimately decisions are made in terms of whether a project is a go or no-go. And if it's a go, what are the milestones that we want to monitor over time? We seek people and projects that drive convergence. That's where the power is, when we kind of converge on the same question and arrive at the same outcome but coming from different directions. It's really extremely powerful…

The fundamental attributes we reward are dogged adherence to excellence and, number two, a deep sense of collegiality and communal support. Now, those are actually opposing attributes. Okay, you have camaraderie and conflict. It's something that, again it's not easy to get at the balance.

On the Human Genome Organization

What is HUGO's relevance and purpose?

That's a very good question. HUGO was constructed and established to help governments know how to do the genome sequence. It's been done. The challenge right now is really twofold. One is genomic medicine. All the things we're doing right now are sufficiently mature that they'll have an impact on how we diagnose, how we treat, how we develop drugs.

The systems biology concept at Lilly was completely a genome-to-systems approach. It was based on transcriptional responses. They're using it as a very intriguing way to organize their drug development framework.

To make personalized genomics real, we have to account for the interactive effect of groups of genes. That's only one aspect. The genetic association studies show the multitude of genes or genetic loci that have small but significant impact on common diseases. It is the composite effect of many genes that will make up the population diversity contributing to a

single disease.

The speed in which this approach and genomic technologies are moving into medicine poses a tremendous challenge for the genomicist, the ethicist, the drug companies, and the patient. And it has to do with scale. When you're dealing with small scale, there's time, there's all sorts of stuff that you can stop and go, but now it's this rush that is 5, 6, 7 orders of magnitude faster and more comprehensive, and you're flooded with information. How do you then dissect that all out? How we manage that, how we regulate that which always requires a computer intermediate is another challenge. How do we deal with a situation where I can know everything possible about you genetically? How do we safeguard it? What kind of legislative process?

Those are really important issues that an organization like HUGO can really contribute to... HUGO is the only [genetics organization] that has no geographic or national limitations.

And what's the second part?

What we're seeing since 1980 has been pretty remarkable — but in the last ten years has been dramatic — is a complete shift in world order. Not in terms of who's on top and who isn't, but in parity. The emerging economies of the world, from Korea to India, China to the southern shores of Australia, South Africa to parts of the Middle East, and Eastern Europe to all of Latin America, there's really a dramatic change in both scientific capabilities, and governments' expectations...

Genomics and genetics are the fastest way for a country to get into high-end biology. Why? Very straightforward. Genomics is technically modular, you can buy powerful machines, and if you have an interesting genetic question you can get it done. The technologies are applicable to plant biology, energy, environmental remediation — it's secular when it comes to that. It's a powerful technology that you can use for all sorts of projects to propel an emerging country to rise in the biological world.

Think about it, if you have a unique population like Iceland. You can beat out everybody. 280,000 people who had no infrastructure in biology suddenly became one of the leaders in genetics. It's an overnight thing. Whereas you would need to build a lot of infrastructure to get a mouse facility in place and a lot of time for breeding, you can set up a bunch of sequencing machines and smart people can do informatics and you can get started right away. The beauty of that is there's no shortage of smart people who are computationally enabled in many of these countries, and they get really turned on that their mathematical skills can have an impact in biology. Well HUGO's role is to help these countries and these people rise to the occasion. It's very exciting.

Japan and Australia started the biological drive in Asia. When we talk about emerging countries, we now include Korea, India, Singapore, Taiwan. South

(CONTINUED ON PAGE 18)

Edison Liu

(CONTINUED FROM PAGE 17)

Africa is doing some interesting things, and in Latin America there are Brazil and Mexico starting significant genomics programs. So, I think we're really talking about some powerful changes in the scientific social order.

Will this change HUGO's identity?

The people who need it the most right now are what I would call the emerging economies... They need a helping hand to figure out how to solve their problems, how to get the systems in place, how to frame the scientific question, how to enhance the critical mass. A big question is: What does it mean to really have genomics work for you?

... In the past, investments in health within emerging economies have primarily been viewed as a social responsibility and a resource consumer. In the last 10-15 years, health has become potentially a revenue generator because of medical tourism, because of biotech, because of pharma. And so we have seen a transition point, where governments of emerging economies began seeing their investments in biomedical sciences also as an investment in economic growth. The dual benefits have not been lost to the politicians: ultimately not only will health get better, but our economy will benefit too because it'll spin off into knowledge creation. HUGO cannot only assist that process, because I think you need economic development, but influence the discussion in such a way that the health component is also augmented. Can we do that as a professional organization? Who knows, but I'd like to try.

On the rise of technology

What are the limiting factors to sequencing technology?

First of all, the $1000 genome is important because of cost considerations. It's the same issue of connectivity. The lower the costs, the more people will use the technology, and the more useful that technology will get. The information is not just additive, it's exponentially more valuable. The web is a classic example. At a specific saturation point, everybody has to use it. It's not an option.

The technical limitation, really, is going to be informatics. We're already having a hard time storing this information, let alone analyzing it. It's only going to get worse. Now luckily, the physicists and the climatologist have worked out a lot of the issues of high performance computing. We're still far behind them in terms of the utilization. But to catch up, we're going to need some technologic advances to assist our computational capabilities.

Given the technical advances, do you think the days of arrays are numbered?

No. I don't think so. But I think what it will do, is that your profit margins in arrays will fall to rock bottom — it'll shake out. Only a few companies will be able to make them on a cost effective basis... I think the smartest thing these chip companies could have done is merge with sequencing companies. Whereas Illumina [acquired Solexa] and 454/Roche bought out Nimblegen. It's the consolidation that mature technologies always undergo... In fact, I think it'll be better for us in that we'll be able to use 100 chips for the price that we use two now.

But the limiting factor is the informatics. I have a feeling we're going to solve those issues, but it's going to lag behind the technologies, unless there's enough money put into it. This is where governments have to invest. Private companies aren't going to put in millions — hundreds of millions — of dollars into supercomputing capabilities.

On Big and predictive biology

Is the Cancer Genome Atlas money well spent?

Yes! I'm biased. Having said that, could it be done better? Probably. How could it be done better? I don't know, and we won't know that until after the fact. Should we just sequence a bunch of tumors and in what manner, etc.? I'm much less for a rigid extraction process as opposed to a structured discovery. There is a difference. The former demands lock-step protocols and study designs for all tumors, whereas the latter takes into account the clinical nuances of each tumor type and the state of molecular/genomic knowledge. I think we're still at a phase where we need to do a little structured discovery.

You published an excellent commentary in *Cell* a couple of years ago on Systems and Predictive Biology. What exactly do you mean by those terms?

The term predictive biology means to identify all the molecular components of a biological process, and to understand how the components interact so that outcome can be predicted in a *de novo* fashion. A lot of the stuff that we're doing right now in transcriptional regulation in human cell lines is fundamentally to understand the complex regulatory networks at the most precise systems that I know — transcription factors and binding sites. And then to learn enough about how to manage the information and know their dynamics so I can predict outcome given a pharmacologic challenge. In a way, this is a physiologist's rendition of synthetic biology and I'm very keen on that.

Does that require a change in how you think about your work?

Yes, it has to. For example, gene discovery. What I used to do is do some kind of screen that would take a year or two to set up. Then painfully pull out 20 candidates. And then pray that one or two of them will be sufficiently interesting that I can spend the next ten years investigating. Now I am less concerned about the individual genes and focus on groups of genes working in concert.

For example, we worked out that the major prognostic expression profiles of breast cancers or most cancers is the proliferation index. And it's reflected in about 1000 to 2000 genes that are consistently changing with proliferation. Any subset of those 1000 or 2000 can make a prediction of outcome in patients. Instead of focusing on prognosis only, I am asking now what are the biochemical and genetic connections for these transcriptional networks, and I can use this information to identify pathways to target in the form of new treatments. To accomplish this, we are finding both experimental and algorithmic ways to filter out, not the noise, but the concomitant roars, that come out of any process. •

Bioinformatics.Org Unveils 2008 Franklin Award Finalists

Finalists promote open access of materials and data.

BY KEVIN DAVIES

The Bioinformatics Organization (bioinformatics.org) has announced the names of the six nominees for the 2008 Benjamin Franklin Award. The winner will be presented with the award and deliver a lecture at the Bio-IT World Conference & Expo, on Tuesday, April 29.

The 2007 winner, Sean Eddy (HHMI Janelia Farm), joined a distinguished group of bioinformatics researchers recognized by Bioinformatics.Org and its some 24,000 members, dedicated to the open access of materials and data, including Michael Ashburnerz (Cambridge/2006), Ewan Birney (EBI/2005), Lincoln Stein (Cold Spring Harbor/2004), James Kent (UC Santa Cruz/2003) and Michael Eisen (UC Berkeley/2002). The six finalists for the 2008 award are:

Philip E. Bourne (Co-Director, Protein Data Bank, University of California San Diego) — Bourne is the founding Editor-in-Chief of PLoS Computational Biology, and co-director of the Protein Data Bank. He continues to develop widely used software tools including SciVee, a free scientific video delivery site (www.scivee.tv).

James L. Edwards (Encyclopedia of Life, Smithsonian Institute) — Edwards has been an advocate for sharing of biodiversity data since the early 1980s. His latest project is the construction of an open-access Encyclopedia of Life (www.eol.org).

Robert Gentleman (Fred Hutchinson Cancer Research Center) — Gentleman, one of the minds behind statistical tool R (www.r-project.org), co-founded and developed BioConductor (www.bioconductor.org), an open-source/open-development software project for the genomic data analysis.

Michael Hucka (California Institute of Technology) — Hucka is the head of the Systems Biology Markup Language team (sbml.org) and the coordinator of the development of SBML, one of the first XML languages widely used in biosciences. .

Francis Ouellette (Ontario Institute for Cancer Research) — Ouellette was an early supporter of the PLoS community (www.plos.org/downloads/plos_ouellette.pdf) and has been a proponent of open access for work derived from public funding, e.g. Genome Canada (genomecanada.ca).

Steven Salzberg (University of Maryland, College Park) — Salzberg produced several popular open-source bioinformatics tools (mummer, glimmer, TransTerm, Jigsaw, etc.), and helped start the Influenza Genome Sequencing project. •

Checks and Balances in Synthetic Biology

Synthetic biology matures amid IP changes.

JAISREE MOORTHY & DENNIS FERNANDEZ

In recent years, biotechnology has undergone a major paradigm shift, the focus moving from understanding nature to engineering nature. This new synthetic biology approach seeks to go further than biomimetics and transcend nature's limitations by rewiring signaling and metabolic pathways in cells, potentially creating an entire organism from scratch.

These modifications coax living cells to function as factories for producing drugs, as scavengers to breakdown waste, as fuel cells or as computational and electronic units. Due to its diverse applicability, synthetic biology will impact many industries such as pharmacology, bio-energy, agriculture, and electronics.

The intellectual property (IP) portfolio — an essential component of any business including patents, copyrights, and trademarks — plays a key role in synthetic biology, as it addresses issues in biological systems as well as those of computers and electronic circuits and other forms of micro-machinery. Various biological parts, such as oligonucleotides, gene products, functions of the engineered system, simulation software and biomedical devices may be patented. Genomic-based patents straddle the thin line between definitions of natural and man-made; the argument in favor of synthetic biology is that the genes are entirely man-made.

The U.S. Supreme Court ruling in favor of patenting engineered organisms in Chakrabarty's case in the early 1980's set precedence for organism ownership. LS9 and Amyris Biotechnologies are some of the companies that have modified the genome of organisms to function as biofactories.

A recent challenge that arose was the patenting of engineered biological systems such as the patent filed by J. Craig Venter on the "minimal bacterial genome" (20070122826), which describes trimming the genome of the smallest known bacteria (*Mycoplasma genitalium*). This particular patent is in the limelight and under sharp scrutiny by the GeneWatch UK and Canada-based ETC Groups because the claims are very broad. A danger in any nascent field is that broad claims could be issued which would thwart further progress in the area because the patent owner now acts as a 'gatekeeper' and can monopolize the market.

Synthetic biology is now at the crossroads of IP-protected and open biology — analogous to the open source model in software engineering (e.g. Linux). Some of the drawbacks of IP protection in this field are high license costs, making it difficult for a start-up to thrive, and monopolization of information by a few companies. For example, Venter's minimal genome patent application was described in *Science* last year as "the start of a high-stakes commercial race to synthesize and privatize synthetic life forms." In addition, Venter's plan to develop an operating system for biologically based software adds to the controversies in this area.

Synthetic biology is now at the crossroads of IP-protected and open biology.

Shared Resources

In the open biology approach, scientists share engineered genetic codes, referred to as biobricks, through public domains (including the Registry of Standard Biological Parts hosted by MIT). The argument for this path is that biological tools and parts can be developed quickly without competition among the researchers and companies.

A concern against the open model is that there are no economic returns for developing expensive parts that involve long-term research resulting in quick fixes becoming the norm and the overall products suffering from quality assurance. A combination of IP-protected and the open model could address the issues in economic growth in synthetic biology.

One idea, proposed by J. Henkel and S. M. Maurer (TUM Business School and Goldman School of Public Policy at the University of California, Berkeley), is to reduce the duration of the patent so that the inventor is awarded the expenses incurred, after which, the inventor may deposit the biological part into the public domain and share with other researchers.

Recently, the Patent Reform Act of 2007 proposed changes that could adversely affect innovation and economy of this industry. For instance, the reform is not supportive of licensing the patent, especially when the patent holder is not involved in product development. This change will greatly affect universities that license the technology to industries and use the revenue to fund further research in this area.

Another drawback is that this reform act proposes to change to "first-to-file" policy rather than "first-to-invent," which deters independent inventors from entering the market.

Synthetic biology promises novel ways to solve global health, food, and energy problems. But as with any new technology, synthetic biology presents potential dangers such as epidemics caused by engineered organisms and uncontrollable changes to the ecosystem.

Granting patents in ignorance of the above concerns could pave the way for an industry that could step over ethical barriers, both in context of society and in its relation to the environment. Synthetic biology is still in its infancy and therefore will require rules, regulations, and moral guidelines to help it mature into a full-fledged industry. •

Jaisree Moorthy, is at the Bioengineering Department, University of Pennsylvania & Dennis Fernandez is the senior managing partner of Fernandez & Associates LLP.

Clinical Research

TGen Targets Cancer with Molecular Directed Therapies

Researchers hope to go after molecular targets, not cancers.

BY DEBORAH BORFITZ

Do cancer patients respond better to an onco-genomic guided therapy than a doctor's "educated guess?" Results of a landmark molecular profiling study will provide the answer and, if it's yes, will likely spur interest in the concept as a way to increase treatment success while lowering development time and expense, says Stephen Anthony, director of TGen Clinical Research Services at Scottsdale Healthcare in Arizona.

The $6 million project is sponsored by TGen (Translational Genomics Research Institute) and Scottsdale Healthcare, the company's clinical unit, and funded through an endowment from the Stardust Foundation. The molecular profiling study began in September 2006 and patient accrual will be completed by the first quarter of 2008, says Anthony.

Based on genetic information from tumor tissue samples and immunohistochemistry stains (for localizing proteins in cancer cells), investigators will select a molecular directed therapy, says Anthony. "Potentially, that will guide us to use a leukemia drug to treat patients with breast cancer. It will force us to think outside of the box because we're going after a molecular target and not a type of cancer."

Anthony, a hematologist and medical oncologist, says all of the patients he sees at Scottsdale Healthcare are "highly refractory to conventional treatment. Patients have to be qualified to go into clinical trials because that's all we offer here."

The focus of TGen is to develop targeted therapies and it has "strong genomic capabilities to carry out the mission," says Anthony. Already, more than a quarter of the 400 oncology drugs under development are on the "radar screen" of TGen, one of the few non-academic centers working to accelerate drug development for cancer patients.

Five-year-old TGen does all the basic science work at a facility in downtown Phoenix (opened in December 2004) and all Phase I clinical trial work gets done at Scottsdale Healthcare, says Anthony. TGen is particularly adept at "enrichment" of patient populations for clinical trials, by homing in on patients likely to have the target a compound is designed to attack, says Anthony. Herceptin, for instance, is given differentially to breast cancer patients who have too many copies of the HER2 gene and express the HER2 receptor.

If an oncology drug gets through a first-in-man trial and looks promising, TGen partners with other large cooperative oncology groups, such as U.S. Oncology and Southwestern Oncology Group, to participate in later-stage trials. In response to growing interest in targeted therapies among drug developers and the general public, TGen is considering adding a second clinical site outside the state of Arizona in the second half of 2008, he says.

Protein Tracking

A majority of trials done by TGen involve measuring the level of protein in a cancer cell, not looking at all 22,000 genes in fresh tumors, as is the case with the landmark molecular profiling study now underway. "If you see a clear expression at the gene level, and...the expression of that gene at the protein level, then you're looking at a protein specific-driven cancer cell, and it's appropriate to use a therapy directed at that protein or molecular target," says Anthony.

Cancer is a heterogeneous disease, says Anthony. "On a good day, 30 to 40 percent of [existing] therapies work." Better results come from treatments based on patients' unique genetic signatures. Cambridge, UK-based KuDOS Pharmaceuticals (a wholly owned subsidiary of AstraZeneca), for instance, has shown that women with faulty genes BRCA1 and BRCA2 have a "dramatic response" to treatment with a poly (ADP-ribose) polymerase inhibitor.

"Within the last year, in particular, there has been a progressive swing toward molecular targeted therapies in oncology drug development," says Anthony.

> Better results come from treat-ments based on patients' unique genetic signatures.

Still, not everyone has bought into the concept.

Some companies prefer to start with conventional drug development, since the FDA threshold for approving oncology drugs is only a 10 percent positive response rate. They may then test a molecular targeted therapy to tease out why the drug worked so well on that minority, says Anthony.

In the field of molecular medicine, the biggest technological need is in the arena of computational biology because of the sheer number of factors that can impact cancer-causing genetic mutations, says Anthony. "The need to find definitive therapies that accurately target and eliminate disease while not harming the patient will require a critical understanding of the myriad networks of communication that cancer cells possess. With greater understanding of these pathways, we can offer patients what they have always wanted, and that is more quality time with their family and friends." •

Resolutions Around Regulatory Compliance

Emerging biopharmas should update eCTD compliance standards in 2008.

BY WARREN PERRY

The New Year is a good time for emerging biopharma companies to look at what their larger counterparts are doing in terms of regulatory compliance, particularly with regard to global electronic submissions.

For various reasons, many emerging drug discovery companies believe they do not have to adhere to the same standards as large pharma companies. Perhaps they are unaware of what is required, or simply unfamiliar with the Food and Drug Administration (FDA). But these companies can only fly under the radar for only so long. Lack of preparation can cost them lots of time, resources, and money. Some understand this, but others delay putting systems in place for as long as possible.

Companies can't do this half way — this is an electronic content world and paper submissions are quickly falling by the wayside. Effective January 2008, the FDA's Center for Drug Evaluation and Research (CDER) mandated that all investigational (and/or marketing) applications be submitted in an electronic Common Technical Document (eCTD) format. In that light, there are three major areas that emerging life sciences companies should keep top of mind with electronic submissions to meet the new guidelines:

1. Audit Trails

This is the single most important aspect to have in place with a system in order to validate data and ensure they are 21 CFR Part 11 compliant. A submission system should be able to easily tell a company and the FDA who did what and when in system. An audit trail is focused on metadata and content within a system — it is not only about documents that are in the system. In general, despite best intentions, it is nearly impossible to anticipate future regulatory requirements and corresponding inquiries. Companies need to focus on things that are auditable now but must also be mindful of the future. A content management system that can easily absorb and adapt to future requirements is a huge plus.

2. Control of Content

The implementation, validation, training, and operation of a System of Record is as important as the content itself. The system must be installed and maintained to meet CFR21 Part 11 standards. Standard operating procedures (SOPs) must be set up to control operation of the system itself in order to be in FDA compliance. These can cover basic things such as rebooting the system, backup, and who has access to what data. These need to be under the control of the IT department to ensure proper operation of the system. A content management system that will be used to generate eCTD's is mission-critical for emerging life sciences companies, and is typically their first such system. Thus compliance and regulatory affairs professionals don't have access to much of the limited IT resources. Plus, IT is often not familiar with these types of systems. Companies may need to lean on vendors that can provide services and guidance around disaster recovery, virtual system access, and security measures.

> Lack of preparation can cost emerging drug discovery companies lots of time, resources, and money.

3. Consistency Across Documents

The FDA can refuse to accept a company's submission, even if it is electronic, for seemingly trivial things such as incorrect fonts or improper margins. Authoring templates are another important change for an emerging biopharma company when it comes to FDA submissions. While it does not necessarily sound rational (What does when it comes to the Federal government?), the FDA is a government agency comprised of rules and regulations. The FDA's nitpicky document requirements go back to when all submissions were submitted in a paper format. Some of the text might not be legible if too close to the binding. Also, companies would often use smaller fonts to cram more information onto less pages and thus reduce costs. Companies need to have templates around content consistency in regards to medical writing, SOPs, and manufacturing. Each needs to clearly define the style and layout of different documents, and provide guidance as to the expected content. All documents need to follow the same format, typically through using templates. The best way to enforce consistent, current templates is through an electronic data management system (EDMS.)

For emerging biopharma companies focusing on these three areas of compliance, the first priority is to conduct an audit of what types of documents and data will need to be entered into an EDMS. Keep in mind the document types, sources, and who will need to review and potentially approve them. You must then determine what type of system best suits your needs, for example by attending a workshop with the Drug Information Association (www.diahome.org) or the Regulatory Affairs Professional Society (www.raps.org). Both organizations can point to technologies that will help you morph from an emerging pharmaceutical company to the next Wall Street success story. •

Warren Perry is a compliance advisor with QUMAS. Email: wperry@qumas.com

ISI Expands Range of eCTD Options

eSubmissions software aids companies without IT, regulatory resources.

BY ANN NEUER

Many large pharmaceutical companies were sitting pretty on the January 1, 2008, deadline for required use of the electronic common technical document (eCTD) for electronic submissions to FDA. But the rest of the market is in a very different place, says Lisa Meyer, director of marketing for Image Solutions Inc. (ISI), a provider of electronic submissions software and services.

"Big pharma has been looking at the eCTD as a strategic trigger to examine how they conduct business and how they need to re-design processes to comply with changing regulatory requirements. Smaller players don't have the IT or regulatory resources to go through the process internally so they are turning to outsourcing," Meyer explains. Even bigger companies are opting to outsource rather than risk a "Refuse to File" decision.

"Customers, whether big or small, want choices," Meyers says, and to meet that demand, ISI offers a range of software solutions that enable in-house operations coupled with outsourced services. As needs change, customers can modify their packages, add to them incrementally, or outsource the entire process.

ISI has processed more than 600 eCTDs and holds a 40 percent eCTD market share among the top 50 pharmaceutical companies with its eCTDXPress product. To handle the anticipated growth in outsource volume, ISI has opened operations in Tianjin, China, to handle back office functions such as the processing of case report forms (CRFs).

According to Meyer, ISI is finding that a few emerging companies have never heard of the eCTD. Others may have heard of it, but a focus group held this summer revealed that many are convinced that FDA will not enforce the eCTD standard.

Meyer says, "The reality is that eCTD is here. The challenge for ISI is to bring the right expertise to the table for the companies that don't know what questions to ask or understand how implementation of eCTD will impact their business processes." •

This article first appeared in Bio-IT World's eCliniqua newsletter.

Register by March 14 and Save up to $200!

Cambridge Healthtech Institute's Sixth Annual

Bio·IT World Conference & Expo

April 28 – 30, 2008 • World Trade Center • Boston, MA

Keynote Presentations by:

Linda Avey
co-Founder
23andMe, Inc.

Joshua Boger, Ph.D.
President & Chief Executive Officer
Vertex Pharmaceuticals, Inc.

John Reynders, Ph.D.
Vice President & Chief Information Officer
Johnson & Johnson, Pharma R&D

2008 Benjamin Franklin Award presented by Bioinformatics.org

Concurrent Tracks:

- Informatics and IT Infrastructure & Operations
- Informatics for Genome Analysis, Biomarkers, and Target Discovery
- Predictive and *in silico* Science
- eChemistry Solutions
- Clinical and Medical Informatics

Keynote Panel:

The Future of Personal Genomics: A special plenary panel discussion featuring

George Church, Ph.D., Harvard Medical School
Dietrich Stephan, Ph.D., Navigenics
Fred D. Ledley, MD, Bentley College
…and more

Platinum Sponsors: BLUE ARC, GenomeQuest, Microsoft, TATA CONSULTANCY SERVICES

Corporate Sponsors: Agilent Technologies, BIOTEAM, SCHRÖDINGER

Official Publication: Bio·IT World

Lead Sponsoring Publications: Digital HealthCare & Productivity, eCliniqua, GEN Genetic Engineering & biotechnology News, nature, Science AAAS, TheScientist

Event Features:

- **Access** All Five Tracks for One Price
- **Network** with 1,500+ Delegates
- **Hear** 85+ Technology and Scientific Presentations
- **Choose** from Four Pre-conference Workshops, the Oracle Life Sciences and Healthcare User Group Meeting, and Symyx Software Symposium 2008
- **Attend** Bio-IT World's Best Practices Awards **NEW**
- **Connect** with Attendees Using CHI's Intro-Net
- **Participate** in the Poster Competition
- **See** the Winners of the following 2008 Awards:
  Benjamin Franklin
  Best of Show
  Best Practices
- **View** Who's Who & What's New in the Exhibit Hall
- And Much More!

Organized & Managed by:

Cambridge Healthtech Institute
250 First Avenue, Suite 300, Needham, MA 02494
Phone: 781-972-5400 • Fax: 781-972-5425
Toll-free in the U.S. 888-999-6288

www.Bio-ITWorldExpo.com

Pre-conference Workshops:

Monday, April 28

Harnessing the Semantic Web for Your Organization (8:00 am-12:15pm)
This workshop will demonstrate the maturing state technologies that focus on deployment strategies for successfully embracing Semantic Web technologies in your organization.

Data Visualization for Effective Drug Discovery Decisions (1:30-4:00pm)
Participants will understand the role of visualization in bioinformatics and medical informatics as applied to the drug discovery process.

Next-Generation Sequencing Data Management (8:00am-4:00pm)
Organized by BioTeam, this workshop presents the analysis, assessment, design, implementation, testing, and support needed to bring a research organization from technology adoption to publication in the next-gen world. Special guest speakers include Cold Spring Harbor Laboratory, Broad Institute, Cornell University Biotechnology Resource Center, Harvard-Lipper Center for Computational Genetics, Naval Medical Research Center, and Intel, as well company representatives from Applied Biosystems, Helicos, Illumina, and Roche.

Electronic Lab Notebooks and Collaborative Knowledge Management for Life Sciences R&D and Manufacturing (8:00am-4:00pm)
Learn how to apply ELNs to create a comprehensive scientific informatics strategy for Collaborative Knowledge Management in R&D and manufacturing enterprises.

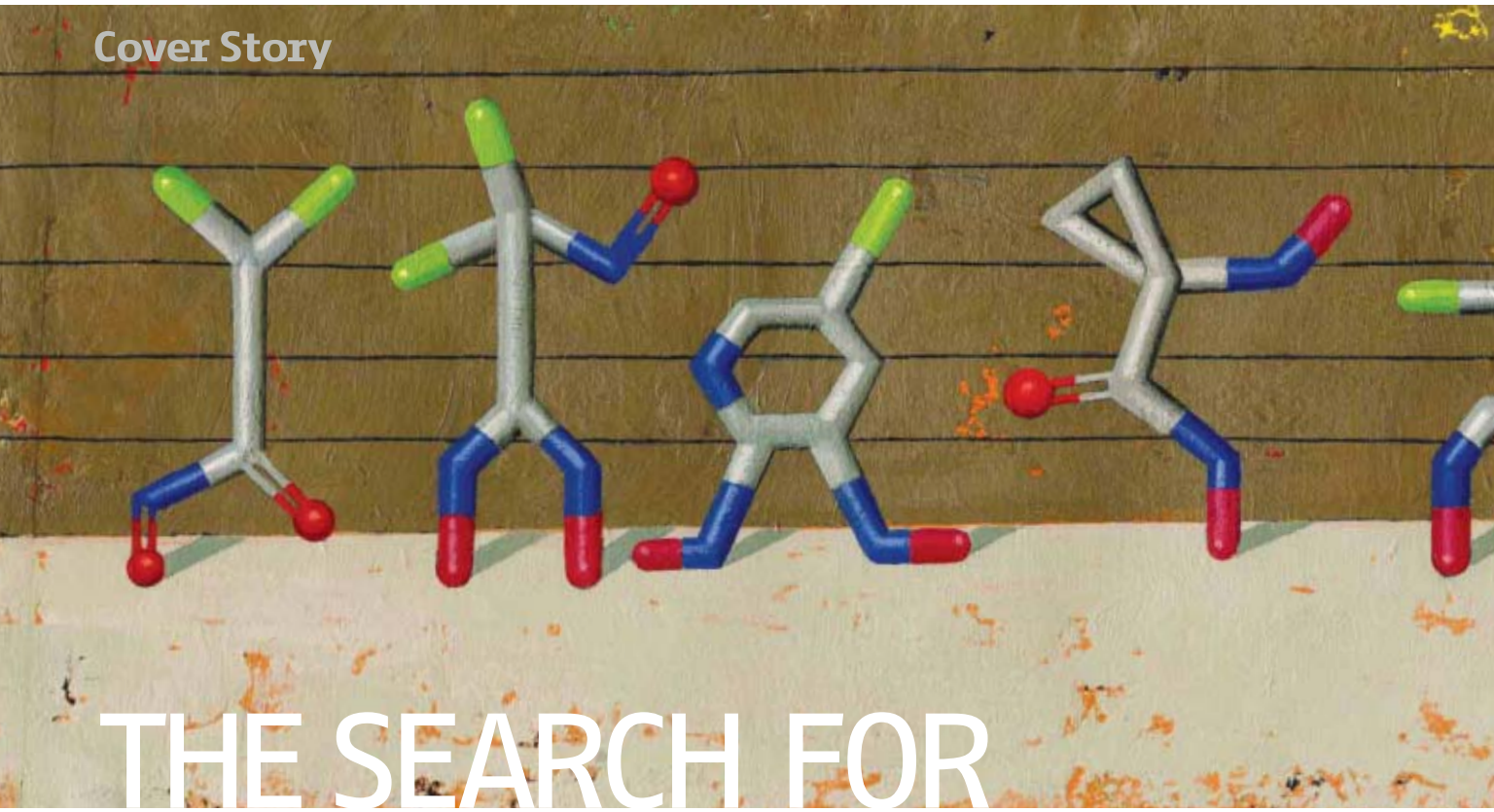
Presentations from:

Zvia Agur, Optimata
Patrick Aloy, Ph.D., Anaxomics Biotech
Samuel (Sandy) Aronson, Harvard Partners Center for Genetics and Genomics
Linda Avey, 23andMe, Inc.
Andreas Bender, Leiden/Amsterdam Center for Drug Research and Novartis Institutes for BioMedical Research, Inc.
Alex Birch, Sirna Therapeutics (a wholly owned subsidiary of Merck & Co., Inc.)
Joshua Boger, Ph.D., Vertex Pharmaceuticals, Inc.
Keith Boyce, Immune Tolerance Network / UCSF
Kenneth Buetow, Ph.D., National Cancer Institute
Barry Bunin, Ph.D., Collaborative Drug Discovery, Inc.
Belinda Chen, Ph.D., Institute for Information Industry
David Chiang, Sage-N Research, Inc.
George Church, Ph.D., Harvard Medical School
Richard Cramer, Ph.D., Tripos
Chris Dagdigian, BioTeam, Inc.
Souripriya Das, Ph.D., Oracle
David DeCaprio, Broad Institute/MIT
José S Duca, Ph.D., Schering-Plough Research Institute
Chris Dwan, BioTeam, Inc.
Kevin Davies, Ph.D., Bio-IT World
Dinesh DeAlwis, Ph.D., Lilly & Co.
Robert Deans, Athersys, Inc.
David de Graaf, Ph.D., Pfizer
Khaled El Emam, Ph.D., TrialStat Corporation & University of Ottawa
Robert Feinstein, Ph.D., Kelaroo, Inc.
Richard Gill, Ph.D., ProNAi Therapeutics, Inc.
Georges Grinstein, Ph.D., UMass Lowell
Kamalakar Gulukota, GenomeQuest, Inc.
Peter V. Henstock, Ph.D., Pfizer Research Technology Center
Alan Higgins, Ph.D., Viamet Pharmaceuticals
Edward P. Jaeger, Ph.D., Johnson & Johnson Pharmaceutical Research
Kevin A. Janes, Ph.D., Harvard Medical School
Marti Jett, Ph.D., Walter Reed Army Institute of Research
Jason Johnson, Ph.D., Rosetta Inpharmatics/Merck & Co., Inc.
David Jordan, Lineberger Comprehensive Cancer Center
Les Jordan, Microsoft
Vipul Kashyap, Ph.D., Partners Healthcare System
Linda King, Eli Lilly and Company
Saul Kravitz, Ph.D., J. Craig Venter Institute
Phillips Kuhl, Cambridge Healthtech Institute
Dmitry Kuznetsov, Ph.D., Vital-IT / Swiss Institute of Bioinformatics
Fred D. Ledley, MD, Bentley College
Towia Libermann, Ph.D., Beth Israel Deaconess Medical Center
Michael N. Liebman, Ph.D., Windber Research Institute
Craig Lipset, Pfizer
Richard Lysakowski, Ph.D., The Collaborative Electronic Notebook Systems Association (CENSA)
Jomol Mathew, Ph.D., Dana-Farber Cancer Institute
Rakesh Nagarajan, Ph.D., Washington University School of Medicine
Ulrik Nielsen, Ph.D., Merrimack Pharmaceuticals
Jean Peccoud, Ph.D., Virginia Tech Bioinformatics Institute
Anton Petrov, Ph.D., infoQuant, Ltd.
Aditya Phatak, Persistent Systems Inc.
Rudy Potenzone, Ph.D., Microsoft Corp.
John Quackenbush, Ph.D., Dana-Farber Cancer Institute
Marco Ramoni, Ph.D., Harvard Medical School
James Reaney, BlueArc Corporation
George Redmond, National Cancer Institute
Ron Ranauro, GenomeQuest, Inc.
Christian Reich, Genstruct, Inc.
John Reynders, Ph.D., Johnson & Johnson, Pharma R&D
John Russell, Bio-IT World
Alan Ruttenberg, Science Commons
Vadim Sapiro, J. Craig Venter Institute
Josef Scheiber, Ph.D., Novartis Institutes for Biomedical Research
David M. Sedlock, Ph.D., Millennium Pharmaceuticals Inc.
B. Woody Sherman, Ph.D., Schrödinger, Inc.
Suresh Singh, Vitae Pharmaceuticals
Philippe Stas, AlgoNomics
Dietrich Stephan, Ph.D., Navigenics
Pablo Tamayo, Ph.D., Broad Institute
William Van Etten, Ph.D., BioTeam, Inc.
M. Vidyasagar, Ph.D., Tata Consultant Services
Ronald Waife, Waife & Associates, Inc.
Chris Waller, Ph.D., Pfizer Global Research and Development
Bernard Wess, ProtonCare and PERSEID Software
David Wishart, Ph.D., University of Alberta
Knut Wittkowski, Ph.D., D.Sc., The Rockefeller University
Olive Yuan, Ph.D., Regeneron Pharmaceuticals, Inc.
Yingyao Zhou, Ph.D., Genomics Institute of the Novartis Research Foundation
Tatyana Zhukov, Ph.D., H. Lee Moffitt Cancer Center & Research Institute
and more!!!!!

Make sure to mention keycode U02 when registering!

www.Bio-ITWorldExpo.com

Cover Story



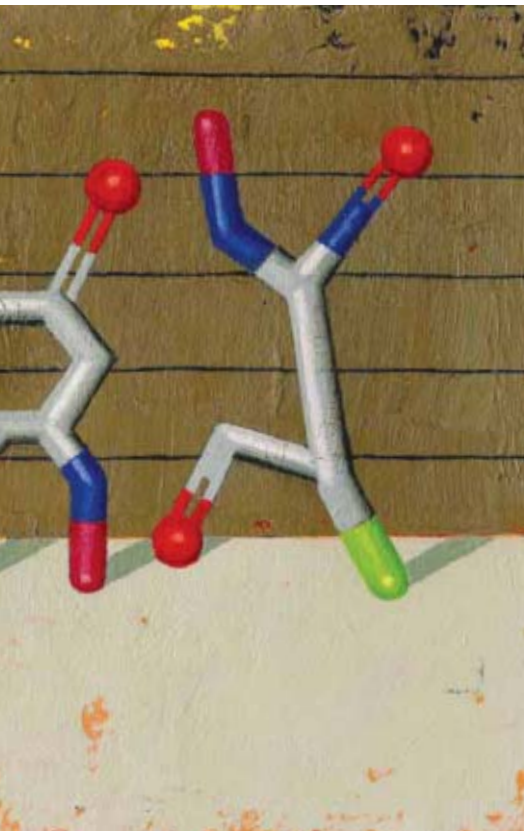
THE SEARCH FOR Unusual Suspects

Scaffold hopping expands the range of core molecular shapes for lead generation. By Vicki Glaser

In the search for novel intellectual property (IP) that can maximize product life, medicinal chemists are embracing fragment-based design strategies. These can overcome some of the limitations of high-throughput screening (HTS) and "leverage information in the literature," says Kenneth Foreman, a computational chemist at OSI Pharmaceuticals.

If a compound has a desired activity, ideally with *in vivo* proof of concept, and "if we can take that compound and make something that does essentially the same thing that we own, we can leapfrog into a competitive position," says Foreman.

Although medicinal chemists have utilized such strategies for years, decisions were often based on assumptions about the core structure. Today, the literature reveals a trend toward smaller, more focused libraries, and fragments rather than full-fledged compounds as the starting point for lead generation. Increasingly there is a focus on the design of ligands



MARTIN JARRIE

that have a desired biological activity, using an approach called scaffold hopping — or, alternatively, ligand or core hopping. Ironically, by starting small, researchers have been able to expand the searchable chemical space and are delivering a broader variety of novel chemical structures that can serve as the basis for large-scale *in silico* screens for both lead discovery and lead optimization.

Supporting the predictions of core structure is one way in which current computational tools for scaffold hopping can have an impact — using the scientific literature to give medicinal chemists a head start. "The tools are helpful by being unbiased toward which fragments are important and which are not," says Foreman. "The more ligand information you have, the more likely that you will be able to pick out the key pharmacophoric elements."

Additionally, "the lack of bias in these computational tools can aid in converting [HTS] hits that have certain undesirable core features into new, IP-able cores," says Foreman.

Software tools can help identify the "key generic interacting elements—the rings, donors, and acceptors"—and use this information to match compounds from either an internal database or an external vendor collection to the reference compound, he explains.

What has increased the feasibility of scaffold hopping in a realistic timeframe is the combination of ever-increasing computer processing speed and a more mature knowledge base regarding structure- and ligand-based design.

"Running these approaches in parallel—for example, running a shape-matching, ligand-based approach at the same time as a docking, structure-based approach—and looking for the commonalities," can identify fragments and compounds that meet multiple criteria, says Mehran Jalaie, a computational chemist at Pfizer. The combination of 2-D strategies and shape-based, 3-D approaches to explore the same chemical space allows chemists to pick up molecules or fragments that look fundamentally different but share pharmacophoric features.

Furthermore, today's faster computer speeds are enabling parallel searches based on shape-matching features and electrostatic properties. "This is allowing us to look at molecules in a generic way and identify new molecules that fit the desired paradigm," says Jalaie. As processing speeds increase, the software tools will soon be able to search a larger chemical space resulting in an almost interactive process.

Varied Architecture

The subject of lead generation and structure-based small molecule drug design traditionally conjures images of a lock-and-key configuration, with the goal of generating a physical or *in silico* model of a drug target and a custom-designed molecule that can bind within its active site. However, over the years, computer-assisted modeling strategies driven by docking strategies have met with limited success in boosting drug discovery productivity.

By the 1990s, when combinatorial library synthesis surged to the forefront, the emphasis shifted from target structure-based design to ligand design and the production of large compound libraries intended to probe a more diverse sample of chemical space. This approach also proved less successful than anticipated, but along the way medicinal chemists started thinking about molecules as fragments — as a collection of sub-structures.

The term "scaffold hopping" was coined by former Hoffmann-La Roche researcher Gisbert Schneider. "It defines the techniques used to identify isofunctional molecules — molecules that have the same bioactivity but different architecture — in other words, different chemotypes," says Schneider, the Beilstein Endowed Chair for Cheminformatics at Johann Wolfgang Goethe-University in Frankfurt, Germany.

As an example, consider ß-lactam, a patented scaffold found at the core of many current antibiotics. Scaffold hopping offers a route to avoid IP conflicts as well as perhaps the side effects associ-

> "[Scaffold hopping] defines the techniques used to identify isofunctional molecules — molecules that have the same bioactivity but different architecture — in other words, different chemotypes."
>
> **Gisbert Schneider**, Johann Wolfgang Goethe-University

ated with ß-lactams to find bioisosteric replacements — compounds that mimic the primary activity of ß-lactams but have a different chemical structure. The goal is describing molecules "not on an atom-and-bond level, but on a more abstract, conceptual level that ignores chemical structure," says Schneider.

Conventional medicinal chemistry suffers from a relatively narrow focus on a comfort zone of familiar synthetic chemistry — narrow at least in comparison to the whole of chemical space. A few years ago, an analysis in the *Journal of Medicinal Chemistry* of the molecular frameworks of known drugs revealed limited diversity in terms of compound shape. Half of the 5,120 drug compounds evaluated shared only 32 different frameworks. Schneider and Kristina Grabowski, writing last year in *Current Chemical Biology,* advocate exploring unique molecular frameworks that "complement 'drug space'" based on their computer-based analysis and comparison of the architecture of drugs and natural products. Their analysis identified more than 1,000 scaffolds not found in a survey of drug compound libraries.

One challenge in scaffold hopping is that chemists are unlikely to venture too far outside the chemistry space they know best. Products such as Tripos's AllChem and BioSolveIT's FeatureTrees help remove that obstacle.

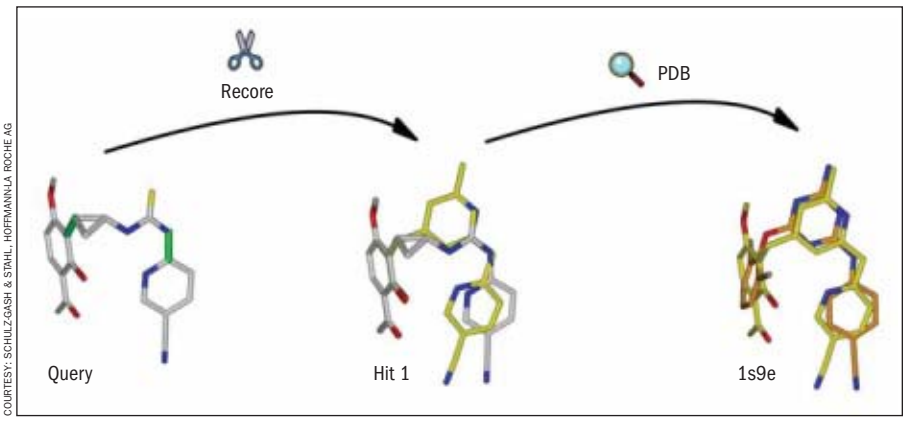
BioSolveIT developed the FTrees tool for "fuzzy similarity searching" to facilitate virtual HTS. Feature Tree, its underlying topological descriptor, captures connectivity and physico-chemical properties of functional groups. An FTrees alignment defines the optimal similarity of two descriptors, enabling SAR detection. According to the company, FTrees can search a catalog of 60,000 compounds in 15 seconds on a standard PC. For *de novo* design via fragment space searching, the software can process 10^{18} compounds in about five minutes. Pfizer has employed this technology to screen its virtual combichem collection of about 3 billion compounds using FTrees Fragment Space (FTreesFS) to perform similarity searches.

Courting Chemists

Some chemists may be reluctant to embrace core hopping due to the nature of its output. Scaffold searching can yield molecules that look quite different from those chemists typically synthesize, raising questions about the wisdom of investing time into producing unfamiliar compounds. This reluctance may be reasonable if based on concerns of synthetic feasibility, as scaffold hopping and *in silico de novo* design strategies typically do not take synthetic tractability into account. Schneider's group alleviates this concern by constructing new molecules based on building blocks derived from existing drugs or natural products. "We produce chimeras of these fragments on the computer, and because these scaffolds are drug-like they are better accepted by the chemists," he says.

"There is an inherent conflict between what IT says to make versus the cost and time" a medicinal chemist must invest to produce a compound, says Tripos CSO Richard Cramer. Tripos intends for AllChem, a product under development, to improve productivity at this juncture. It incorporates a chemistry engine that generates the most accessible and pharmaceutically acceptable scaffolds and R-groups based on established sets of feasible reactions and easily obtainable building blocks. "The resulting database enables researchers to search a chemical space of at least 10^{20} structures, searching for novelty while at the same time staying grounded in synthetic reality and providing chemists with ready-to-use, computed synthesis routes," says Cramer.

Patent space is becoming increasingly crowded, notes Christian Lemmen, CEO of BioSolveIT. "If a lead dies, for whatever reason, then [often] a whole series dies altogether, which generates the need for sufficient diversity in the pipeline," says Lemmen.



COURTESY: SCHULZ-GASH & STAHL, HOFFMANN-LA ROCHE AG

BioSolveIT's ReCore tool identifies new scaffolds based on the molecular core of the query.

Activity Focus

Unlike traditional medicinal chemistry, which tends to generate analogues of active compounds, core hopping searches for novel scaffolds while preserving the activity of a potential lead. It utilizes computational tools that focus on properties relevant for binding rather than on chemical structure.

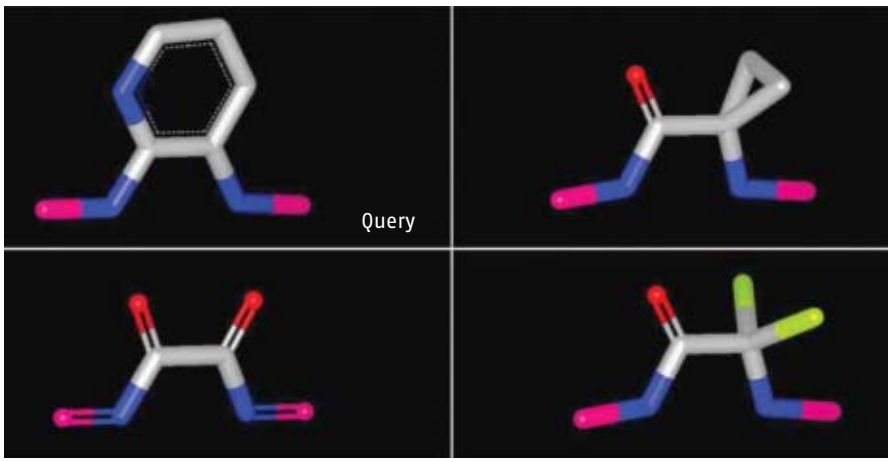
Fragment-based design strategies can be employed to enable a scaffold hop. One of the main challenges for *in silico* ligand design is "the accurate prediction of binding affinities and the ability to detect low-affinity binders, which is often the case for fragments," says Lemmen. BioSolveIT's ReCore is a ligand-based design tool that "abstracts from the ligand structure by taking only the vectors connecting the core piece to its R-groups," explains Lemmen.

As its name infers, ReCore targets the core of a molecule for removal, searches 3D-fragment libraries for a suitable replacement, and generates a new scaffold, maintaining the surrounding components to create a chemically distinct query compound. ReCore was developed by Patrick Maass, at the Center for Bioinformatics, University of Hamburg, in collaboration with Hoffmann-La Roche. The developers recently described its use for scaffold hopping based on small-molecule crystal structure conformations (*J. Chem. Inf. Model.* 2007; doi: 10.1021/ci060094h).

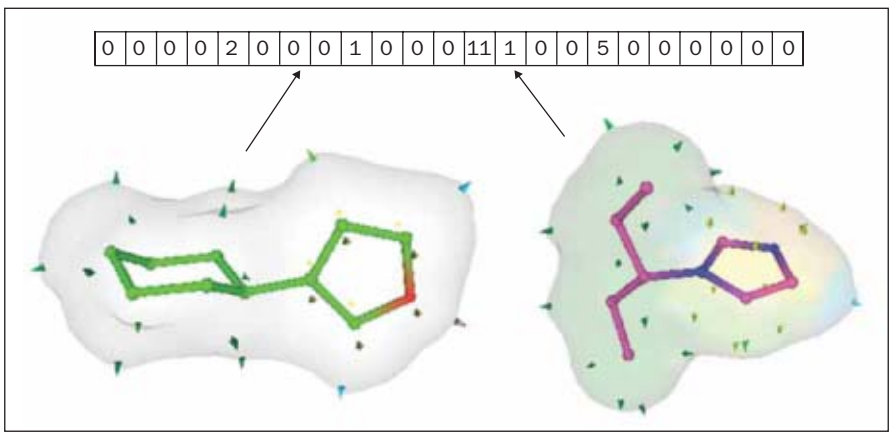
Tripos's topomer technology began in the company's ChemSpace virtual library design software. Cramer describes a topomer as a molecular fragment that is aligned according to a set of rules. Topomer searching treats molecules as a collection of fragments. Developed by Cramer and previously used as an in-house product development tool, Tripos is now releasing two products — Topomer Search and Topomer CoMFA.

Topomer Search aids users in searching corporate databases and scaffold hopping to find compounds with a similar shape as an identified lead. Cramer describes in-house studies comparing topomer searching with docking strategies that showed ligand similarity searching to be more effective for predicting desired bioactivity and of value for generating patentable chemical entities. While compounds identified by lead hopping from off-the-shelf compounds may look similar to a target protein, they "are structurally dissimilar enough to look different to the patent office and also differ based on 2D fingerprints," says Cramer.

Topomer CoMFA, which is in beta testing, takes the topomer concept further and can help distinguish between hits of similar shape by developing a 3D quantitative structure-activity relationship (QSAR). When used for lead optimization, the technology exploits the corporate database as a source of molecular fragments or scaffolds for use in virtual screening to rank compounds based on their predicted potency.



BROOD output: The 2,3 diaminopyridine fragment is the query (top left), the three other fragments are hits with high shape and chemical similarity to the query.



LASSO (from SimBioSys) uses a "fuzzy" descriptor based on the surface properties of ligands. The molecules shown have different skeletons, however share the same descriptor, which is ideally suited to scaffold hopping applications.

Molecular Alignments

Although their names may not be entirely descriptive, ROCS, EON, and BROOD represent a family of software tools from OpenEye Scientific Software designed for lead hopping. ROCS, the company's flagship shape-matching product performs shape and chemical functionality comparisons of a database of compounds to a query molecule. This approach has proven successful in large part because similarity in three dimensions is not necessarily related to similarity in two dimensions.

"People tend to look at molecules as 2D structures, whether on paper or on a computer screen," says Paul Hawkins, senior applications scientist at OpenEye. "And they tend to tinker with the molecule at the 2D level, making small, incremental changes." The result is typically a set of analogues of an existing compound that explore a limited region of chemical space.

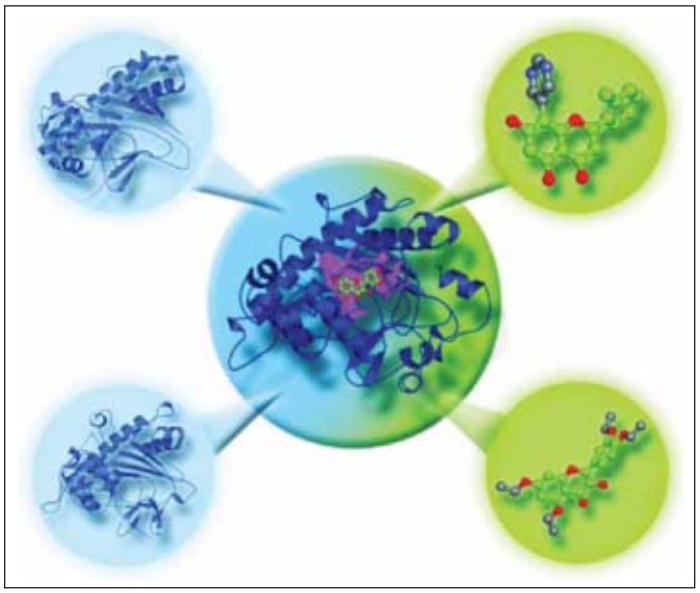
"To find truly different molecules that have similar activity, or lead-hops, requires thinking about similarity in 3D," Hawkins says. "Similarity searching based on shape and electrostatic properties goes beyond atomic compositions and how those atoms are joined together," explains Hawkins. "Molecules with very similar shapes might look very different when depicted in two dimensions."

EON and BROOD are evolutions of the company's 3D property matching technology. Based on a shape alignment of two molecules, EON calculates and compares their electrostatic potentials, while BROOD uses the same principles to compare molecular fragments rather than whole molecules. EON identifies molecules predicted to have similar activity, and BROOD returns a set of suggested R-group replacements and information on possible starting points for exploring chemical space. Future versions will build in estimations of properties such as solubility and other ADME qualities for rank ordering.

"The ultimate arbiter of whether a compound 'works' should be how well it matches the 3D structure of the protein's active site," emphasizes Hawkins, keeping in mind that the structure of neither the target nor the ligand is static. Both have

intrinsic flexibility and it is not possible from a 2D model to predict how the protein might adapt to accommodate the shape of a ligand.

"Structure-based approaches have traditionally been viewed as superior in many applications in drug discovery," observes Hawkins. However, a weakness of traditional structure-based methods and docking engines is the approximations that must be made to get to a solution in a reasonable timeframe; "these approximations compromise the quality of the output," he says. "Since we can treat an active ligand in a rigorous way in terms of its shape and electrostatic properties, we can do at least as good a job as with protein-ligand docking for identifying interesting compounds for screening."



The BioFocus DPI stARLITe database allows the user to navigate through compound, assay, activity, and target relationships in a chemogenomics fashion.

Filling the Toolbox

Several companies have developed computational tools that can facilitate lead hopping and devised innovative ways of generating new scaffolds and lead compounds. SimBioSys designed eHiTS LASSO (electronic High Throughput Screening Ligand Activity by Surface Similarity Order) as a tool for virtual ligand screening. It generates an interacting surface point type molecular descriptor from the 3D structure of a ligand that is conformation independent and, together with a neural network machine learning technique, screens molecular databases at a rate of 1 million structures in less than 1 minute.

Chemical Computing Group's MOE pharmacophore modeling methodology comprises tools for scaffold replacement that identify substitution points on a scaffold molecule and the location of potential R-group substituents, as well as tools for pharmacophore elucidation and searching, and the Pharmacophore Consensus application that suggests pharmacophore queries based on a set of aligned active compounds.

Discovery Studio 2.0, from Accelrys, includes algorithms for fragment-based design, activity profiling, and flexible docking. The Catalyst software suite can be used for 3D pharmacophore modeling, pharmacophore-based alignment of molecules, and generation of pharmacophore hypotheses based on SAR data.

Recently, Schneider and colleagues described an approach that combines a 3D alignment-free pharmacophore descriptor with artificial neural networks (ANNs). The ANNs were trained for receptor selectivity using known antagonists with IC_{50} < 1μm as reference structures. Self-organizing maps (SOM) helped select structurally diverse compounds for bioactivity testing via virtual screening. Several hits were identified, the most potent with functional IC_{50} values ranging from 9 to 21 μm. Various chemotypes were identified that did not overlap with the scaffolds of reference compounds based on atom types, bond order, carbon scaffold, and/or ring size.

Defining a Bioactivity Fingerprint

The StARLITe (structure activity relationships from the literature) database of bioactive molecules developed by BioFocus DPI contains medicinal chemistry information abstracted from the scientific literature. It includes some 400,000 compound records representing approximately 48,000 chemical series and 3,440 distinct molecular targets, and covering more than 1.4 million assay data points. Experimental data linked to biological activities include sequence data mapped to assay results and links to synthetic chemical routes and assay protocols.

Based on either sequence data or an activity pattern, the user can search active ligands for privileged scaffolds or search targets for closely related receptors. The software ranks the scaffolds based on fragment specificity (compared to all known scaffolds in the StARLITe database) and fragment elegance (synthetic ease). It identifies a subset of scaffolds that are synthetically accessible for use in activity-focused library design.

StARLITe's compound-activity mapping capability can be used to profile hits from HTS to identify alternative activities and to identify scaffolds that share similar patterns of bioactivity. The software queries bioequivalent scaffolds and determines a similarity score, which can then be used to develop a bioactivity fingerprint.

Edith Chan, a research fellow at BioFocus, extracts compounds from the literature, breaks them down into scaffolds and R-groups, and uses the scaffolds to search StARLITe and generate a focused library for a protein and its homologues based on the target's amino acid sequence and SAR data linked to individual compounds and scaffolds. She describes this approach as "knowledge-based hopping," in contrast to *de novo* ligand design strategies.

Where can scaffold hopping lead? If done properly, and with a little bit of luck, to previously unexplored chemical space, new and perhaps "funny looking" molecular structures, and a series of bioactive compounds that could bring new life to drug targets in pharmaceutical pipelines for which no good lead candidates have been found. •

From PlayStation to Protein Surfaces

Mercury Computer Systems subsidiary SolMap brings compute muscle to *in silico* drug design.

BY KEVIN DAVIES

In late 2005, Mercury Computer Systems conducted a successful partnership with the lab of Boston University biomedical engineering professor Sandor Vajda. It liked the results so much it decided to invest in to Vajda's company, SolMap Pharmaceuticals.

"Mercury is in the business of solving very, very computationally intense — and intensive — problems," says Mirza Cifric, general manager of SolMap. For two decades, Mercury was best known for providing 3D reconstruction engines for computed tomography applications in medical imaging, CT and MRI to companies such as Philips, GE, and Siemens.

In recent years, Mercury has sought to expand its life sciences business from what Cifric terms, "a hardware computer business into a more application-focused business." That shift was prompted by the advent of more and more powerful Intel processors. Says Cifric: "The company had to grow vertically in terms of the value add. And that came by acquiring algorithms and bringing value to the application-level capabilities."

Cifric, a native of Bosnia who moved to the United States in 1995, was tasked with developing Mercury's business in preclinical discovery. Mercury already had some customers, notably Pfizer, for which it developed (originally for Agouron, which Pfizer acquired in 2000) a computational chemistry platform. Pfizer took the visualization software that Mercury had developed for Agouron and incorporated various algorithms and components into the computational chemistry framework, not unlike a Pipeline Pilot. "We learned lots about preclinical discovery and chemistry from that relationship and we were looking for ways to expand what we could contribute [to] pharmaceutical industry," says Cifric.

One way to impact the drug discovery process was to apply its computational prowess to combinatorial problems in genomics, systems biology, and other fields. Cifric roamed university campuses from looking for intellectual property and companies with such capabilities. Vajda's company proved attractive "because SolMap's technology was heavily bottlenecked by computational limits," says Cifric.

> One way to impact drug discovery is to apply computational prowess to combinatorial problems.

An expert in computational approaches to protein structure analysis, Vajda and his grad students had created technology around binding-site identification of protein targets. But even on an IBM Blue Gene supercomputer at Boston University, the simulations typically lasted 24–48 hours.

Mercury saw an opportunity: it was the first company to commercially deploy STI (Sony/Toshiba/IBM) Cell (cell broadband engine architecture) processor, the heart of Sony's PlayStation 3. "With some engineering expertise, which is really the bread-and-butter of Mercury, we were able to take that 48-hour simulation down to literally minutes," says Cifric. Today, those protein surface analysis calculations take 5 to 10 minutes for the whole simulation.

SolMap Solutions

Despite the potential of SolMap's technology, the question remains: how to make a successful business? Cifric explored options including providing hardware, software, and/or services to pharma companies. The reality however, he says, is that "unless you're impacting the drug discovery pipeline in a very meaningful way, there really isn't a significant value add to be captured in the pharmaceutical space."

In order to turn SolMap (currently 15 staff) into a drug company capable of taking drugs from *de novo* design into the clinic, Cifric hired Frank Guarnieri, the founder of Locus Pharmaceuticals (See, Locus Focus, *Bio•IT World*, December 2002). "Guarnieri has developed the small molecule design capability on this technology," says Cifric. In addition, "we have married experimental capabilities, specifically structural biology with our computational tools. So we're using NMR and X-ray crystallography, integrated with the computational analysis, to do fragment-based drug design." Cifric says the addition of those two wet-lab capabilities has brought about "an absolute metamorphosis [compared to] what this company was."

In addition to collaborations with a range of clients to enhance their drug discovery capabilities, SolMap is pushing three in-house programs. One program is in hypertension (a renin inhibitor program), another in COPD (chronic obstructive pulmonary disease), which Cifric hopes to speed into late preclinical stages. He notes that a unique structural insight in the hypertension program, initiated less than 18 months ago, is already in lead optimization, because of the platform's capabilities.

The third is in the antibacterial space, developing an adjuvant to existing antibiotics that prevents drug resistance. Cifric says SolMap's target is well validated. "If we're able to inhibit this particular target, we will make all antibiotics that have been on the market for 20-plus years, which

are completely unefficacious at this point, relevant once again... effectively resurrecting them!"

Competitive Advantage

"The discontinuity between knowing everything about the binding site and designing a small molecule drug is bigger than Grand Canyon," says Cifric, who claims that SolMap's technology affords unique insight into protein structural analysis. "We have very unique insight into the bind and sub-binding pockets and the hot spots relative to the binding," he says. For a completely *de novo* design, SolMap can pursue a fragment-based approach, exhaustively exploring the binding area. In other cases, SolMap might identify known binding sites as well as sub-pockets of those sites. The company then generates a rank order of hot residues most relevant to drug design. "It allows us to derive a known inhibitor to the additional sub-pockets or hot spots, which are often a unique finding of our platform. And then sometimes alleviate some of the problems that an inhibitor might have, such as potency or toxicity."

SolMap's fragment-based approach, which involves testing *in silico* fragment libraries, relies on a density function rather than binding specific fragments and "docking" them. The information obtained from the binding analysis produces a cloud of all possible chemistries that sit in the structural space. Then, Cifric says, using the simulation, a medicinal chemist can make fragment substitutions on a set of scaffolds, and make fragment substitutions on those scaffolds to obtain full coverage of the binding sites by generating a small focused medicinal chemistry library that is target specific yet extensible.

A former colleague of Guarnieri's at Locus, Bruce Dorsey at Cephalon, was the first medicinal chemist to design an HIV protease inhibitor at Merck. "He didn't go after potency, he went after extensibility and he was able to alleviate the problems further down in the biologically relevant tests," says Cifric. "We've implemented that methodology in our *in silico* design, but then we rapidly go to high field NMR structure, which gives us the direct confirmation of the hypotheses." SolMap typically synthesizes a library of 50 to 60 small molecules that are then examined using eNMR, selectively searching the target surface space.

Despite encouraging progress, SolMap has no intentions of launching clinical trials itself, says Cifric. The research collaboration model would deploy SolMap's core capabilities effectively as a preferred partner in structure-based drug discovery programs. "It's a lot more short-term oriented, but very cost effective and risk-sharing for the partners... Nowadays, if you don't have intellectual property creation of your own, the valuation of your business will be dramatically impacted if you're just a service business." •

DrugBank Database in Commercial Partnership

Deal with GenomeQuest exchanges access privileges, clinical data for funding.

BY KEVIN DAVIES

The Canadian developers of a popular drug database have partnered with a software company in an unusual win-win relationship. The deal exchanges short-term funding that might ensure the long-term survival of the database in return for exclusive data access to the commercial partner, GenomeQuest.

DrugBank — a database of drugs, their properties, targets, and metabolizing enzymes — was originally developed in 2005 by David Wishart's group at the University of Alberta. Wishart's team also produced the Human Metabolome Database (See, The Human Metabolome Project, *Bio•IT World*, April 2007).

Wishart's group recently unveiled DrugBank version 2.0 — a major improvement on the original version (See, "DrugBank Data") — in the 2008 database issue of *Nucleic Acids Research*. But GenomeQuest's subscribers have enjoyed a preview of the 2.0 release for the past three months. This month, GenomeQuest launches DrugBank Pro, an enhanced version of 2.0 that includes investigational drug data not publicly available.

Michael McManus, GenomeQuest vice president and general manager, says "DrugBank Pro is an important part of our offering. It wasn't just us that noticed this was a cool database, our customers also noticed. It adds a lot of chemistry richness." Wishart's team was receptive when first approached by GenomeQuest. A series of meetings in Canada and Boston eventually clinched the agreement.

"We saw the [original 2006] DrugBank paper and instantly recognized it was very important for our users," says GenomeQuest's senior director, content development, Kamalakar Gulukota. "DrugBank may be a small database, but nevertheless it's separating the grain from the chaff... Ultimately, it's what all the pharma and biotech companies are looking for."

The deal benefits both parties. In return for funding the Wishart group's efforts (support comes also from Genome Canada), the Westborough, Mass., company gains a six-month preview of the DrugBank data such that newly approved drugs will appear in GenomeQuest before they appear on the web. Given the current pace of new drug approvals, that discrepancy won't affect too many compounds or inconvenience users of 2.0.

More significantly, DrugBank Pro offers exclusive access to information on investigational drugs in clinical trials, which again would only appear in 2.0 once they are approved. While DrugBank 2.0 carries information on some 3,000 experimental drugs, these are ligands to proteins in the Protein Database that are not in clinical trials, and may never get there because of toxicity or other issues.

Gulukota stresses that the collaboration helps academic and public users by helping to ensure DrugBank's survival. "DrugBank was developed as an academic project," he says. "David is aware that unless there's continuous interest — someone driving it — it will fall by the wayside. Each year, *Nucleic Acid Research* publishes hundreds of databases, but most databases fall into disuse."

Wishart agrees: "I think the interaction with GenomeQuest has been mutually beneficial and it certainly was key to helping get the latest release of DrugBank in tip-top shape," he tells *Bio•IT World*.

Silo Mentality

The great value of DrugBank 2.0 stems from its aggregation of data that are traditionally siloed, or as Wishart terms it, "an effort to bridge the 'depth versus breadth' gap between clinically oriented drug resources and chemically oriented drug database."

"The data are available in multiple sources, but you need to connect it up," Gulukota says. "Almost all targets and drug metabolizing enzymes have multiple entries in GenBank and other databases, including patent databases. The main problem is, when you're looking at a target, do you get the chemistry information? Conversely, when you're looking at a chemistry database, do you get the targets?"

Gulukota continues: "There's been this silo effect in pharma companies, where

(CONTINUED ON PAGE 34)

DrugBank Data

- DrugBank version 2.0 debuted in January 2008, and includes entries on "all (or almost all) drugs... approved in North America, Europe, and Asia."
- DrugBank 2.0 has information on 4,900 drugs, including more than 3,000 experimental compounds, and 60 percent more FDA-approved drugs. It also contains 1,565 drug target sequences.
- Each "DrugCard" entry holds more than 100 data fields, including chemical, pharmacological, molecular biological data, and ADME and drug-drug interactions.
- Wishart's group developed much of its own software, including BioSpider and other webcrawlers that pull information from various sources, patents, the Orangebook, company press releases, etc. Text-mining software (PolySearch) collects information on drug targets.
- Most drug "fact boxes" in Wikipedia are generated from DrugBank tables.
- DrugBank offers text, BLAST, and chemical structure query tools. The ChemQuery tool now includes ChemAxon's MarvinSketch structure drawing tool.

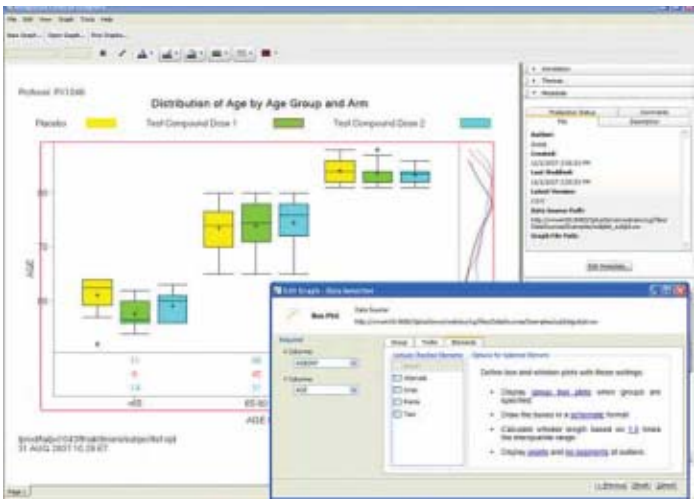
Computational Development

Insightful Makes Sharing Graphs Easier

Software solution makes statistical graphics accessible across drug discovery.

BY JOHN RUSSELL

Sharing graphs of clinical trial and preclinical data among diverse decision-makers is often challenging. Insightful Corp. has tackled this problem with the launch of Insightful Clinical Graphics (iCG), a software solution for defining and repurposing statistical graphics across all functional areas in drug discovery, development, and marketing.



iCG makes statistical data accessible across the organization.

Key obstacles for sharing such graphs in the past have included: the unfamiliarity of non-statisticians stakeholders with necessary software and data tagging and manipulation techniques; a more generalized lack of ease-of-use characterizing those tools; and often cumbersome methods for converting graphs and data into different formats.

Insightful attempts to sidestep these issues by leveraging Web 2.0 social networking technology, user-friendly point-and-click interfaces, and automating much of the behind-the-scenes drudgery such as retaining necessary meta-tags.

The end result, "reduces the time and resources needed for graphical review and submission reporting of pre-clinical and clinical data, while providing consistent information flow between functional areas and regulatory agencies," says Michael O'Connell, director of life sciences for Insightful. iCG should also reduce the demand on scarce statisticians who were sometimes called upon to facilitate data sharing rather than focus on critical data interpretation and trial design.

Productivity Focus

Insightful is targeting the top 50 pharma companies and says it has three early adopters using the product though he declined to name them. The graph sharing framework is available now, and being sold on a per seat basis and priced "not much differently" than market norms, according to O'Connell, who says only SAS has a somewhat similar offering.

The financial pressures prompting biopharma to cut costs and personnel has put a premium on productivity tools such as iCG. "If you go into a company that's just laid off 10 percent of their workforce, they are eager to get projects in place that drive productivity or drive early detection of any safety issues and so funding for projects [that leverage products such as iCG] has continued, though the sales cycle hasn't changed," says O'Connell.

He says there's been considerable restructuring in Insightful's account. "Our core user has been the statistical community. One trend we've seen over the last year or two is statisticians, rather than being in a central group, are being dispersed to therapeutic areas and are reporting into the business unit."

Insightful also designed the new solution to handle regulatory requirement, so if you are in a compliant environment and inserting a graph into regulatory filings, it's fairly easy for the software to grab necessary audit files. •

DrugBank Database

(CONTINUED FROM PAGE 33)

chemistry and biology don't necessarily speak the same language, let alone talk to each other. Often you have chemists who don't know anything about the target, [and] they don't necessarily look at splice variants, or other receptors the chemical might bind. Many questions that could nip non-productive research in the bud are never asked because there isn't that mindset. Pharmaceutical executives have needed to connect chemistry and biology for a long time."

Last December, GenomeQuest offered a "sneak peek" of DrugBank Pro to its users, and hopes it will attract new users within its client base. Users will be able to ask many key questions earlier, says Gulukota: "Are there splice variants of this gene I need to be worried about? If one of those variants is in the heart, and you're exploring neurology indications, you might want to look for cardiovascular side effects or cardiovascular indications."

McManus notes that GenomeQuest's high-throughput extension will be valuable in using DrugBank Pro. He also hopes that the relationship with Wishart's group continues. "Given how creative David is, if there are other ideas he might have, we'd be open to talking to him." •

FURTHER READING: Wishart, D.S. *et al.* 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucl. Acids Res.* 36:D901–6.

Clustering the Power of IP

SparkIP links patent and inventor data to build rich, contextual search.

BY ALLISON PROFFITT

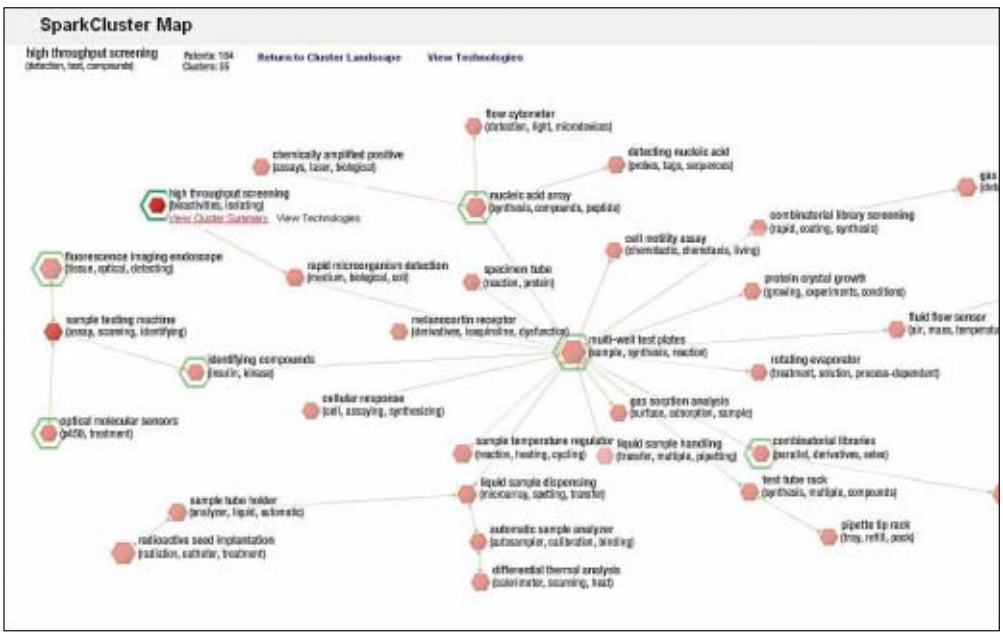
'A living, breathing map," is how SparkIP CEO Ed Trimble describes it. The online company, founded in October, has launched a Beta version of an intellectual property network that presents information on some 3.5 million patents and 3,800 licensable technologies. These are displayed graphically in "SparkClusters," akin to subway maps of connected areas of research with hubs and interchanges.

The company's dataset includes information from the United States Patent Office (USPTO) dating back to the 1960s. "It's the patented ideas, it's in the inventors, it's the innovations that are involved," says Trimble. "We are actively adding new sets of information: patent applications, we're looking at adding scientific journals, and scientific literature and kind of mashing that all together into these maps."

SparkIP adds 3,000 new patents each week. All of the relationships are established via algorithms developed by the company, and the clusters are linked together with semantic analysis. Trimble says it provides a "giant map that covers everything from information technology to pharmaceuticals to sporting goods — all the innovation, the inventions that have led to the formation of these fields."

While the clusters might be visually interesting, Trimble says that the information they contain is far more valuable than simple search results. People can "navigate into" the clusters and "find things that are specific to their search that they would find with a traditional search engine like a Google, but then find a landscape where they can find non-obvious results."

Context is the SparkIP touchstone. Instead of keyword-driven searches that result in a "laundry list of things that aren't ordered or prioritized," SparkIP provides a level of additional "visual contextualization that says if you search on microfluidics, here are all the areas within microfluidics and you can see that visual and here's how they connect." Users can ask: "What do you want to drill down into? It's a much more powerful tool for finding the information that you're seeking."



A SparkCluster shows linked topics. Green-ringed topics have technology available for license.

IP Marketplace

SparkIP's business model doesn't depend on being a better search engine. (When asked if he considers Google competition, Trimble laughs. "You know, I think everybody has to consider Google a competitor!") SparkIP is currently the largest public marketplace for IP in the world. Within each map, licensable technologies are highlighted and for sale. SparkIP has already partnered with eight research institutions — Johns Hopkins, NIH, Duke, Georgia Tech, Stanford, North Carolina State, Tufts, and EPFL of Switzerland — and is talking to "many more." Today, the site boasts 3,800 technologies in the marketplace.

The network is free while the company builds listings, which could last until spring. In the future, Trimble envisions several revenue streams, including, "fees for the marketplace, so listing fees for licensable technologies. Premium subscriber fees for research, so that will be more targeted toward industry, but we think some university and government labs will also pay those, probably site license fees to give those research tools to all their members," explains Trimble. "We see a community forming around this."

Free access to research won't disappear once the subscriptions are in place. Trimble plans on keeping a public face of the website accessible to researchers, supported by advertising. "We think it's important to keep parts of the site open and free." •

IT/Workflow

Symyx Rounds Out Software Offerings

2007 MDL acquisition fuels claims of market leadership.

BY KEVIN DAVIES

The dust is still settling on Symyx Software's acquisition last October of MDL Information Systems, but the company believes the merger has made it the undisputed market leader in life sciences software.

Timothy Campbell, president of Symyx Software, sees the life sciences software market as "an important and critical market," albeit narrower than CRM, ERP, or supply chain. "So in order to be successful in this market, you need to be the leader." The new Symyx Software has more than 200 engineers worldwide "developing our applications for our customers." Symyx Software is one of three business units of Symyx Technologies (the others being Symyx Tools and Research Services), and is destined to become the largest in 2008.

The MDL acquisition from Dutch publishing giant Elsevier (See, Symyx Acquires MDL, *Bio•IT World*, Sept. 2007) tips Symyx' offerings firmly towards life sciences from the chemical industry. Campbell says Symyx recognized that it needed to augment its two major platform areas, automation and electronic lab notebooks, "with a deeper capability in chemistry, broader content offerings, and a complete cheminformatics infrastructure."

MDL Merger

Joining Symyx from MDL is Trevor Heritage, now senior VP of science at Symyx. Heritage had joined MDL in 2005 from Tripos, tasked with transitioning MDL from a platform company to more of a solutions and applications company. He credits Elsevier with doing many things right over the years, including the acquisition of the Beilstein database and the creation of the PharmaPendium database.

Heritage thinks the integration has proceeded "amazingly smoothly" so far. The merger marries Symyx's strengths in ELN and lab execution with MDL's forte in logistics and data access and analysis, "without really causing conflicts between the two," he says.

The new combined strategy offers solutions in four key domains. Two are the traditional strengths of Symyx, namely ELNs and software for experimental design, workflows, execution, and results capture in areas such as developmental chemistry (crystallization, solubility, catalysis).

The third domain is logistic systems for registration, chemical inventory management, and workflow requests, including instrument scheduling and calibration. Heritage says: "A scientist can sit there and looking at only one application environment — their notebook — use that environment to completely design the experiment from A to Z, execute that experiment, as well as integrate with all the other operational systems you need as part of the normal lab workflow."

Fourth is data access, analysis, and the physician support area, which is where users will find former MDL products such as ISIS and Isentris. Says Heritage: "We provide a single environment, whether you're working in development or discovery, where a scientist can accomplish everything they need to do, whether they're setting up an experiment or whether they're actually in the data access and analysis and physician support phase," says Heritage. "If you're looking at biological screening data in some column of your spreadsheet, you'd have access to the systems and the notebook pages that generated that data."

Early Days

Every big pharma is now using Symyx's software, Campbell asserts, although that wasn't the case before the acquisition. He quickly cites Eli Lilly, which two years ago, purchased an enterprise deployment of Symyx' ELN spanning discovery to development. "They are at about 850 seats with an objective to grow to 1,100 seats in 2008 and then with a move to biology and analytical ELN offerings."

Heritage adds that many biopharma companies were already transitioning from ISIS to Isentris for chemical information and data management. Since the merger, he says, "We're already closing business deals where customers have licensed or added to their MDL or Symyx license, with the specific intention of making integration between Isentris and between Symyx lab execution or Symyx notebook." •

The 2008 Symyx Software Symposium ("R&D Integration Success") will be held in conjunction with the 2008 Bio-IT World Conference & Expo, World Trade Center, Boston: April 30-May 2, 2008.

> "We provide a single environment, whether you're working in development or discovery..."
>
> **Trevor Heritage**, Symyx Software

New Products

Solid Workstation

The Zephyr SPE Workstation offers automated Solid Phase Extraction (SPE) at the benchtop. The Caliper product includes a 96-well plate format and SPE-specific software designed to increase efficiency and streamline the sample preparation process for mass spectrometer analysis. Clog detection ensures that operators can reliably process all samples.

Product: Zephyr SPE Workstation
Company: Caliper Life Sciences
Available: Now
For More Information: www.caliperls.com

DNA, Hands-Free

For hands-off DNA isolation, Invitrogen has introduced the iPrep PureLink gDNA Blood Kit. The kit is designed for the iPrep Purification Instrument, and produces high yields of pure genomic DNA from up to 350ul of fresh or frozen blood. The yields are high enough for clinical studies or multiple PCR-based standardized testing assays.

Product: iPrep PureLink gDNA Blood Kit
Company: Invitrogen
Available: Now
For More Information: www.invitrogen.com/iprep

RNA Structure Calculations

CLC Bio has released version 2.0 of its CLC RNA Workbench. The new release offers user-friendly versions of partition function calculations for RNA secondary structure and minimum energy free scanning. The features are ideal for full genome scans of viruses, or for scanning mRNA for structural signals. A graphical interface allows viewing of RNA's secondary structure.

Product: CLC RNA Workbench v. 2.0
Company: CLC Bio
Available: Now
For More Information: www.clcbio.com/rna

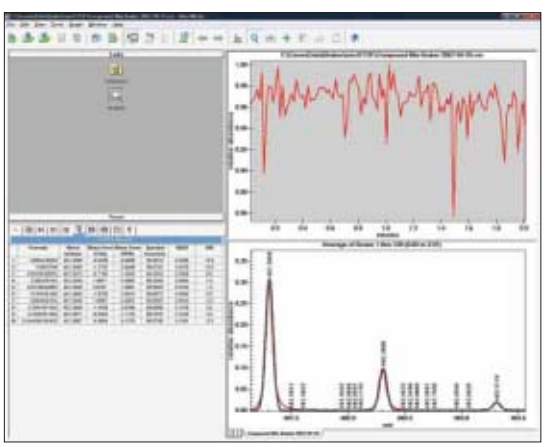
Data Management

SeqMan Genome Assembler, the newest tool released by DNASTAR to aid in processing next generation sequencing data, permits the assembly of fragment data sequenced using 454, Illumina, or Sanger sequencing techniques. SeqMan Genome Assembler integrates data assembly with the Lasergene suite of analysis software, allowing analysis on Windows or Macintosh computers.

Product: SeqMan Genome Assembler
Company: DNASTAR
Available: Now
For More Information: www.dnastar.com

Mass Spec Search

Cerno Bioscience's new MassWorks sCLIPS (self Calibrated Lineshape Isotope Profile Search) will be on display at PITTCON in February. The latest addition to its award-winning MassWorks mass spectrometry software portfolio, sCLIPS significantly improves formula identification results from High Mass Accuracy (HMA) instruments, providing higher-quality, reproducible data through accurate line-shape calibration. For the first time, sCLIPS also enables users to obtain unique formula ID from ion trap mass spectrometers.



Product: sCLIPS
Company: Cerno Bioscience
Available: Now
For More Information: www.cernobioscience.com

The Russell Transcript

Toward a Predictive Model for a Cell

JOHN RUSSELL

In the closing days to 2007, a really nice piece of systems biology work was published in the journal *Cell*, in which researchers developed a predictive model for a free cell, in this case the Archea organism, *Halobacterium salinarum NRC-1*.What's more, the authors suggest that even though their model is for a relatively simply organism (~2400 genes), the approach used to build it can probably be used to tackle complex organisms.

The authors of the paper describe a predictive model called EGRIN (Environmental and Gene Regulatory Influence Network). They used a data-driven discovery approach to determine regulatory and functional interrelationships among roughly 80 percent of NRC-1's genes.

"Using relative changes in 72 transcription factors and 9 environmental factors (EFs) this model accurately predicts dynamic transcriptional responses of all these genes in 147 newly collected experiments representing completely novel genetic backgrounds and environments — suggesting a remarkable degree of network completeness. Using this model we have constructed and tested hypotheses critical to this organism's interaction with its changing hypersaline environment. This study supports the claim that the high degree of connectivity within biological and EF networks will enable the construction of similar models for any organism."

It is perhaps unsurprising that much of the work was done at the Institute for Systems Biology and led by current ISB researcher Nitin Baliga and a former ISB researcher now at the Center for Comparative Functional Genomics, New York University, Richard Bonneau. Indeed, this was a classic systems biology exercise, as espoused by ISB founder, Lee Hood (See, What is Systems Biology, *Bio•IT World*, September 2007), a co-author. The work involved global measurements (genome-wide); quantitative and dynamic measurements; careful system perturbation (genetic and environmental); integrating different data types; and of course adherence to the systems biology cycle of perturbation-measurement-model-hypothesis-perturbation.

Baliga's research depends on data-driven discovery to find relationships between genes.

Construction and Validation

There were several hurdles. For example, roughly 38 percent of NRC-1's genes had little or no functional assignments. The group incorporated functional relationships from comparative genomics as well as predicted structural and domain similarities until achieving "nearly 90 percent... meaningful association with either a characterized protein, a protein family, or a structural fold." Similar techniques were used to boost the number of putative transcription factors.

266 microarray experiments were used to construct the networks and 147 microarray experiments were used to validate model predictions. Network construction was based on the "Inferelator algorithm" (catchy name) developed in large measure by Bonneau. The authors note the number of experiments required was relatively modest given EGRIN's model's high accuracy and suggest the interdependence of many networks and, at least for metabolism, cells may usually function in one or a few dominant states.

"What is powerful about this approach is that it took under six years to move from genome sequence to this level of understanding for a relatively poorly-studied organism. Indeed, it would be significantly quicker to implement the same approach with a newly sequenced organism given that much of the scientific methods including experimental procedures, algorithms, and software have been delineated through our study," write the authors.

Of course, there's still work to be done on EGRIN. Many other regulatory mechanisms — small RNAs, epigenetic modifications, post-translational modifications, metabolite-based feedback — are not included and may account, at least in part, for its failure to predict what 20 percent of the genes are doing.

Further Reading: Bonneau *et al.* 2007. *Cell* 131: 1354-65.

It's Time Your Lab Evolved

- **Flexible architecture** to map diverse processes

  Sequencing & High throughput cores | Proteomics & Genomics labs | Pharmaceutical R&D units | Tissue & Biobanking centers |
- **Fully configurable** to cater diverse functions & end-users
- **PDA and Wireless Support**
- **Full support for regulatory compliance**

  (GxP, 21 CFR Part 11 and ISO 17025)

Now offering
free process consulting
for your LIMS needs.

Biotracker™

Specialty Laboratory Informatics Solutions for Life Sciences

For more details visit us at
www.ocimumbio.com/lims
Talk to us at +1 317-536-5465

India +91 40 665 64400-08 | USA +1 317 228 0600 | The Netherlands +31 (0)30 68 78 788

E-mail: outreach@ocimumbio.com

Ocimum Biosolutions
... enabling R&D™
An ISO 9001:2000 Company

Alternative Thinking About The Size of Blade Systems:

CAREFUL, DON'T TRIP OVER THE INFRASTRUCTURE.

Strength and power, now in a smaller, more convenient size. Introducing the HP BladeSystem c3000. All the technology of our larger BladeSystem in an efficient, compact, affordable package. Careful, watch your step.

Technology for better business outcomes.

Powered by the Quad-Core Intel® Xeon® Processor[1]

Get down to the smallest details. Visit hp.com/go/nocompromise69
1-866-625-1019

1. Intel, the Intel logo, Xeon, and Xeon Inside are trademarks or registered trademarks of Intel Corporation in the U.S. and other countries.