

► **Scaling the heights of productivity**

Knowledge engineering software empowers researchers in processing new molecular entities.

BY JAMES TAM, PETER GIANNOUSIS, AND BILL NELSON

"Understanding the human genome, combined with dynamic research, has resulted in an overload of data that is generating between 2500 and 5000 new drug targets, which has led to major bottlenecks in target validation and lead optimization," said Marc D. Kozin, President of L.E.K. Consulting, at the Analytical & Life Science Systems Association 2002 Fall Meeting.

As R&D approaches continue to evolve, the cost is rising and productivity is falling. "Despite technical advances, current R&D productivity is lagging across the industry, which is evident from the decreasing number of FDA approvals and also the number of 'original submissions' of novel molecular therapeutics," reported Kozin. David Molowa, Managing Director of J. P. Morgan, reported at the same meeting that sales growth rates of U.S. pharmaceutical companies have slowed from 16% in 1999 to an estimated 8% in 2002.

SUGEN, Inc. (www.sugen.com), a life sciences company and a wholly owned subsidiary of Pfizer, discovers and develops novel drugs to fight cancer and other diseases. The company's challenge was to manage the proliferation of data and a voluminous increase in instrument and technical reports that needed to be archived, easily retrieved, and shared. It recognized that these capabilities directly affected its ability to increase productivity and decrease product time to market. That information, typically scattered across different information islands, is a laboratory's most valuable intellectual property. The company sought a solution that would manage the overload of data efficiently by improving the system of creating, searching, retrieving, and archiving

instrument data and reports.

SUGEN specializes in targeting the mechanisms through which messages are sent from one cell to another—the "signal transduction pathway". The company leverages its expertise in identifying these targets through applying its first-rate drug discovery platform to validate multiple targets simultaneously and screen and develop lead compounds across a wide variety of pro-



grams rapidly and efficiently. The company believes this strategy can reduce cost, time, and risk associated with taking potential products to market.

Although SUGEN had integrated the full range of today's state-of-the-art biology and chemistry capabilities, like many companies, it was operating on a paper-based system. This inefficiency slowed production and added to the costs of delivering products to the marketplace, thereby hampering the company's strategic approach to drug discovery.



KEY TERMS: automation, high-throughput, informatics

Challenges of increased data and information

SUGEN's Analytical and Chemical Development Department develops scale-up processes for active pharmaceutical ingredients (APIs) and provides analytical support for both API and product development. The department generates numerous reports, which traditionally were filed as hard copies, and then more recently, scanned and electronically filed in a simple Windows-based folder system. At the time, no automated archival system was available to back up the voluminous data generated from their computer-controlled instruments: a dissolution system, a particle sizer, a differential scanning calorimeter (DSC), a thermogravimetric analyzer (TGA), HPLCs, an FT-IR, and UV spectrophotometers.

Because each instrument uses stand-alone software, data from each instrument needed to be backed up manually on CD-ROMs or manually copied to the network drives. This step was not only extremely time-consuming but also prone to human error. Often, extra copies of reports were generated, there was difficulty in readily locating CD-ROMs, or the backup process simply failed. Backup failure was an especially big problem because of both personnel turnover and the unacceptable possibility of losing valuable company information. These problems were expected to worsen as the number of analysts and instruments grew.

SUGEN also faced the challenge of archiving the processed analytical data and reports. Many of these were printed, copied, and pasted into lab notebooks, but this was not practical for large studies (e.g., >50 HPLC injections). Analysts also kept many of these printout reports in binders, which occupied a lot of shelf space and were not easily retrievable.

Another challenge was managing all of

the department's reports (about 200 per year). Earlier, most reports had been filed as hard copies in filing cabinets. More recently, a system was devised to store scanned copies electronically in a simple Windows shared folder system, with cover pages. This folder system, however, did not allow for easy retrieval or keyword searches.

Managing knowledge

Addressing these critical challenges, SUGEN recognized that a knowledge engineering system would increase its productivity and could decrease its time to market for new molecular entities. SUGEN sought a system that would automatically archive instrument data in a secure way, provide 100% fidelity of data transferred and archived, easily retrieve archived instrument data for reprocessing, store electronic reports printed from instruments, eliminate the voluminous binders of printed HPLC traces, and store and easily search written technical reports.

At the time of purchase, the department was growing; therefore, it needed a system that would provide access to an unlimited number of users (but with a fixed number of concurrent users), providing flexibility and easy management of user accounts. To protect the company's investment in a new application, the preferred system would use a universally adaptable browser as an interface and a standard PDF format that is easy to read now and will likely be in the future.

The company selected CyberLAB Knowledge Engineering System (KES) from Scientific Software, Inc. (www.scisw.com), to leverage SUGEN's well-documented research, securing its records against loss and misuse, protecting its intellectual assets, and improving its organizational productivity. CyberLAB KES is an electronic library that collects, organizes, warehouses, indexes, and safely archives structured and unstructured electronic records, from raw data to laboratory reports to compliance records.

The CyberLAB KES automatically archives data from the instrument PCs.

Reports are generated from the laboratory instruments and automatically filed for secured storage and searchability. The Scheduler Agent archives data at night so that it does not interfere with the day-to-day use of the instruments or burden the network during peak demand hours.

Custom templates

SUGEN staff have designed custom templates for CyberLAB KES to retrieve individual laboratory reports easily, signifi-

metadata that provide a trail to the instrument's binary raw data, which can be retrieved from CyberLAB for reprocessing. Flexible templates can easily accommodate variations of the standard reports and still extract useful metadata. This flexibility is made possible using "key value" searches designated by tokens (values in a document or table) and their separators (characters such as commas that are placed between tokens) that are set up using the Template Key Definition screen.

SUGEN's preconstructed queries allow scientists to easily and directly select criteria such as "chromatogram injection time" and "instrument operator" to search for HPLC chromatogram reports. Scientists are able to retrieve instrument data easily from one location and readily incorporate instrument graphics, such as chromatograms, mass spectra, and DSC and TGA traces into reports using simple copy and paste functions.

If the instrument software can be installed on a remote desktop PC, archived data can be retrieved from the knowledge engineering application and be reprocessed at the remote PC. This allows for more efficient use of the instruments because the instrument can be running experiments and not be tied up for data reprocessing.

In near future, SUGEN will implement the File Association feature, linking written technical reports to the instrument reports. This will

allow easy retrieval and review of raw data that supports a particular conclusion.

With CyberLAB KES's ability to automate raw data archiving, provide easy retrieval of printed instrument reports, and straightforwardly incorporate graphics from instrument reports into written reports, SUGEN staff is saving considerable time and seeing significant gains in productivity.

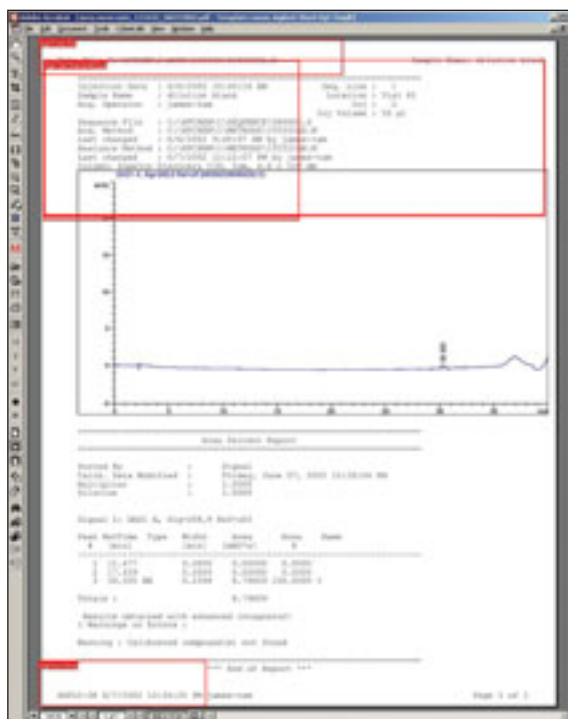


Figure 1. Searchable metadata are extracted from standard instrument output, reducing the time necessary to generate written reports. This CyberLAB KES report template incorporates data imported from Agilent's Chemstation HPLC software.

cantly reducing the time needed to compile final reports and incorporate graphics. The most commonly used standard HPLC instrument reports are templated to extract searchable metadata, such as the instrument name, sample name, time of injection, injection method, and project name from the report header and footer (Figure 1). The estimated time needed to generate a written report is reduced by about two hours per person per week. Additionally, the graphics feature is a huge timesaver and eliminates the need to scan individual graphics manually for inclusion in the reports.

The printed laboratory reports contain

James Tam is an associate scientist, **Peter Giannousis** is the senior director, and **Bill Nelson** is an associate director in the Analytical and Chemical Development Department at SUGEN, Inc. Send your comments or questions about this article to mdd@acs.org or to the Editorial Office address on page 3. ■