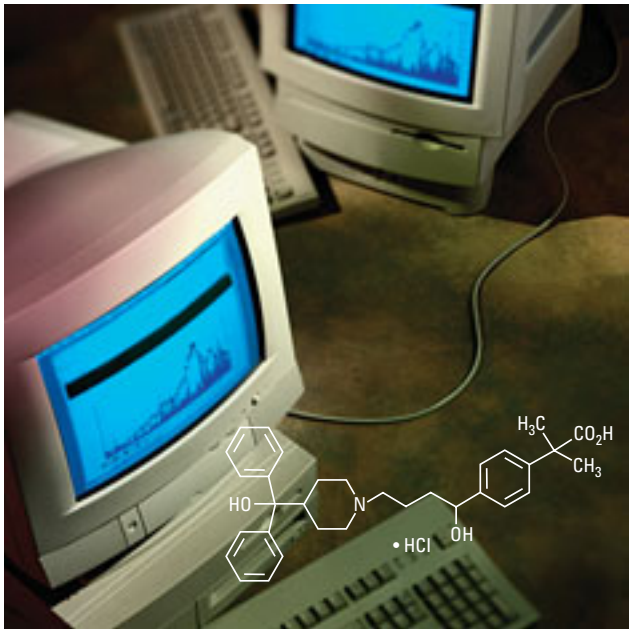
E pluribus unum

A chemical information system for a global R&D organization bridges disparate cultures, synthetic processes, and informatics systems.

BY LUDWIG FRANZISKET

When companies decide to merge, they often cite the advantages of linking many resources and gaining entrance into markets around the world. Making this vision a reality in the new company's R&D division, however, means sharing information among laboratories where different languages are spoken, and different instruments and information technology (IT) systems are already in place. Moreover, research does not come to a halt while the transition is taking place. High-throughput and combinatorial chemistry constantly accelerates the pace of compound synthesis and screening, increasing the speed and efficiency of drug discovery and creating new demands on IT and business systems. These demands are especially acute for large global R&D networks, which need strong informatics infrastructures to make the vast amounts of genomic, biological, and chemical data generated daily accessible at multiple sites while guaranteeing that the data is consistent, high-quality, stored, secure, and organized.



When Hoechst and Rhône-Poulenc merged to form Aventis in December 1999, the fledgling company was faced with the challenge of developing a registration system for chemical information that would link 5000 scientists working in research labs around the globe. Another challenge was the range of synthetic techniques used by Aventis researchers. The system needed to accommodate the full range of contemporary synthesis technologies, including classical preparation of a single batch, parallel synthesis of 1–100 batches, high-throughput medicinal chemistry of 100–10,000 batches, and large-scale combinatorial chemistry production of up to 100,000 batches at a time.

The project team realized early on that it was dealing with multiple sites using many different chemical information systems and numerous processes for registering chemical data. Its members saw the need to define a common process for chemical registration across all sites, creating one process conforming to a common set of business rules and a common grade of data quality.

Tailoring the system to the workflow

To design the system to meet these specifications, the team sought inspiration not from the old legacy registration systems but from those who would be using the system: the laboratory scientists. The guiding principle of the project was that the new system should conform to scientific workflow. Rather than designing the system purely from an IT perspective, they examined the processes employed in the labs, identifying the similar steps and the differences between the workflows at each location.

Arriving at a common process was not an easy task. The design of the system was based on input from scientists from all global R&D sites, which created a new hurdle for the project team. The team had to deal with multiple cultural behaviors and histories, along with different approaches to dealing with chemical data. Many discussions were needed over the course of six months to define a process that was acceptable to scientists at all company sites.

From concept to product

To transform the ideas generated during discussions of system requirements into an actual product, Aventis partnered with MDL Consulting to develop the Aventis Registration Process (ARP), a global registration system and business process implemented with professional guidance and tools from MDL (San Leandro, CA, *www.mdl.com*).

To meet the scientists' challenging requirements, MDL designed a system with a distributed architecture that is a hybrid between a client–server and a true multi-tier system. The Oracle databases that perform the communications between the various components of the ARP system combine the design typical of these databases with the functionality of a chemical database, allowing scientists to search the database by chemical structure.

Although the initial intent was to use a Web-based client to eliminate the hassles of installation and upgrades inherent in a thick client design (one that requires spe-

KEY TERMS: automation, combinatorial chemistry, high throughput, informatics, process

PHOTO: PHOTODISC

cialized software installed on each user's PC), the time frame of the project dictated that the initial production version of ARP consist of a thick client with a number of services running on Web-accessible Windows NT servers. Because the classes are well defined, future iterations of ARP can easily be used in a thin-client environment (a platform-independent design that requires only a Web browser).

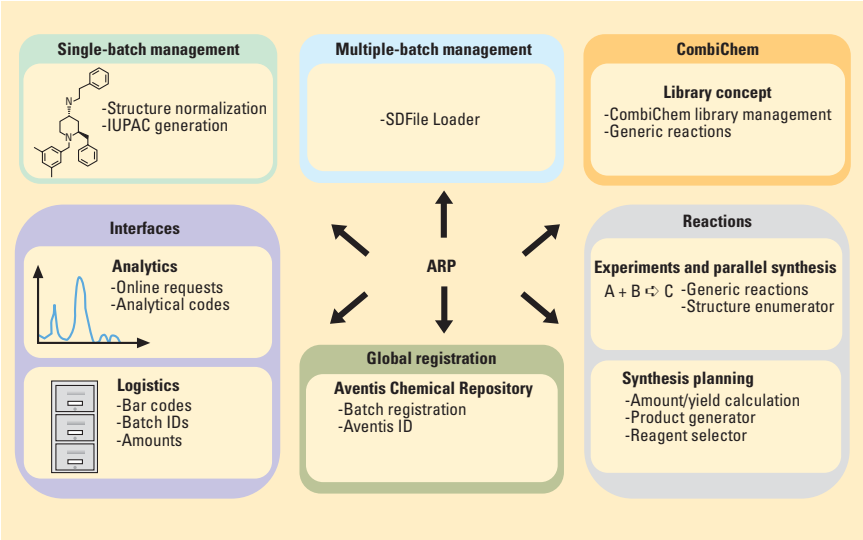


Figure 1. Scope of the Aventis Registration Process (ARP): A modular approach. (The SD file format is commonly used to transfer chemical data between applications.)

ARP achieved its goal of conforming to scientific workflow, as is evident in the user-friendly interface, which allows the scientists to conduct day-to-day research efficiently without complicated IT maneuvers. To enter a new batch, a scientist need only draw the structure or retrieve the structure from a database, name the compound, and provide synthesis information. The compound is scheduled for analytical testing via the system, and all resulting analytical data are then stored in the system. When the necessary information has been gathered for a compound, the scientist merely clicks the "publish" button and receives an Aventis ID number in 30–40 s. The simplified interface masks the underlying complexity of the ARP system. During the half-minute that a scientist is waiting for ARP to generate the ID number, the system sends the chemical information that the scientist provided to multiple databases and servers. After the information is added to the local database, a local server sends the data of newly submitted compounds to a server located at one of Aventis's main research facilities in Frankfurt, Germany. The server in Frankfurt then passes the compound information on to the Aventis Chemical Repository (ACR), the master compound database.

To decrease the demand on the master database, each research site has its own copy of the master database. Every evening, the local ACR compares itself with the global ACR and updates itself to match the global database. In this way, scientists searching for compound information can search a local copy of the ACR for chemical information, decreasing traffic on the global ACR.

The success of ARP can be measured by the response of the system's users.

Measuring success

Since the initial release of ARP in the summer of 2001, the system has expanded to become more than a tool for batch registration (Figure 1). Using ARP, scientists manage experiments, submit online requests for analytics, and generate reports to share with colleagues. The most recent additions to the application include an experiment module and a library concept module.

With the experiment module, chemists enter information about a synthesis, such as the starting materials, catalysts, and reaction conditions. The products of the reaction are all then entered into the database. For multistep syntheses, all intermediates are also entered into the database to generate their unique ID and batch numbers. This module has created an internal searchable reaction database, enabling chemists to learn from the synthetic techniques used by scientists around the globe.

The library concept module allows chemists to register families of compounds prior to their actual synthesis. Researchers can then look for potential active compounds, successful syntheses, and unexpected products in these libraries. They can also access supplemental data, such as internal reports and literature references, associated with specific compounds.

With each subsequent version, ARP moves closer to reaching the long-term goal of becoming an electronic notebook that captures all chemically relevant experimental information. Aside from the fact that the system is well within reach of this goal, the success of ARP can be measured by the response of the system's users.

ARP has gained wide user acceptance among Aventis scientists, largely because of their deep involvement in all phases of developing the system. Our team has also received very positive feedback about the system, which reflects the ability of the ARP system to conform to the scientific workflow and expertly handle the business operations.

ARP succeeded largely because of the sustained collaboration and teamwork between Aventis and MDL, the ability of Aventis scientists to marry chemistry with informatics, and the commitment of Aventis management to integrating chemistry knowledge globally with a standardized IT solution.

Ludwig Franzisket is a global project manager for Drug Innovation & Approval Information Solutions at Aventis (Frankfurt, www.aventis.com). Send your comments or questions about this article to mdd@acs.org or to the Editorial Office address on page 3. ■