

Computers and Automation

New techniques opened new worlds of convenience and possibilities.

DAVID FILMORE

As with much of modern society, computer technology has transformed the chemical sciences and, in the process, launched two major industries—computational chemistry and laboratory automation—that help provide a foundation for the modern research endeavor.

Computable Chemicals

Computational chemistry relies on the fact that molecules can be viewed as mathematical entities. By the 1950s and 1960s, scientists had developed sophisticated mathematical models; they were interested in calculating electronic structures and the relative stabilities of different molecular conformations.

At this time, early computers were becoming available (the first computer, ENIAC, was built for the U.S. Army in the 1940s). The very complex calculations necessary for even simple organic molecules were making evident the applicability of these machines to chemistry. In 1961, for instance, UCLA chemistry professor James Hendrickson used the IBM 709 computer (“capable of 8000 additions or subtractions, 4000 multiplications or divisions, or about 500 or more complex functions per second,” he wrote) to calculate the relative conformational stabilities of cyclohexane based on various geometrical factors.

In 1965, H. L. Morgan of the Chemical Abstracts Service (CAS) published what is called the Morgan algorithm, a system for representing chemical structures in graphical format (“connection tables”) that could be stored and entered into computational operations. This allowed the CAS to develop its modern chemical registry and, generally, provided the basis for chemists to interact with a computer in their own structural language. By this time, the Digital Equipment Corporation (DEC) had introduced the PDP8, the first widely distributed minicomputer.

The concept of chemical structures as logical, computable entities got a major boost in 1969 when Harvard University professor and future

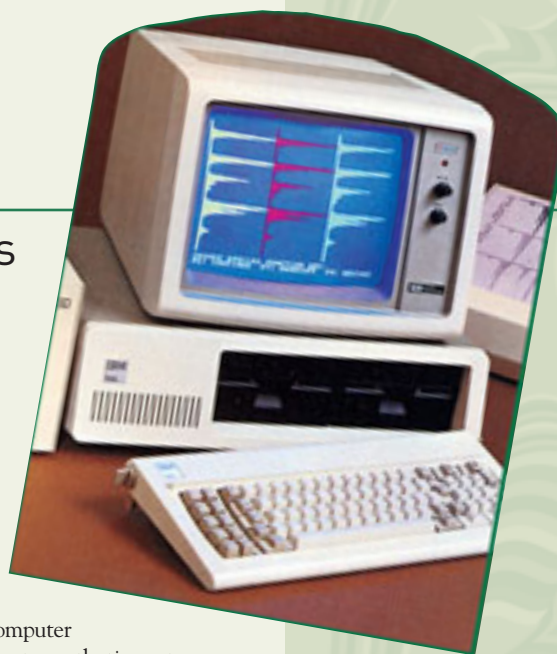
Nobel prize winner (1990) E. J. Corey, with his postdoctoral fellow Todd Wipke, published a paper in *Science* that described computer software that could generate synthetic routes for chemical structures based on Corey’s method of logical retrosynthetic analysis.

The following year, Carnegie Mellon University professor John Pople developed Gaussian 70, a computational chemistry program that quickly became the mainstay of theoretical physical chemists for performing ab initio electronic structure calculations.

Meanwhile, Wipke became an assistant professor at Princeton University, where he improved on the Morgan algorithm by providing a means to computationally represent stereochemically unique structures. He also developed an improved version of the program with Corey, called Simulation and Evaluation of Chemical Synthesis, which could store and retrieve chemical structures and simulate chemical reactions. It was licensed to chemical companies beginning in 1974.

In 1978, three years after Bill Gates and Paul Allen founded Microsoft, and one year after the introduction of the Apple I, Wipke, then at the University of California, Santa Cruz, along with Stuart Marson and Stephen Peacock, started Molecular Designs, the first chemical database company. A year later, the company sold its first product, MACCS.

In 1979, Washington University professor Garland Marshall founded Tripos Associates to manufacture molecular modeling and conformational analysis software for investigating drug molecules. Tripos’s first commercial molecular modeling program, SYBYL, came out in 1982. A string of molecular modeling companies followed, including Chemical Designs and Hypercube in 1983; Biode-



Top: Nelson Analytical ad, *Analytical Chemistry*, 1983

Center: Early integrators, *Today's Chemist at Work*, 2004



COMPUTERS AND AUTOMATION



GOING ROBOTIC

In 1981, Zymark introduced the first laboratory robotic system, Zymate 1. Tecan, PerkinElmer, and Fisher Scientific followed suit with competitive laboratory robots over the next several years.

By 1984, scientists at Purdue University reported using the Zymark system for the “robotic orchestration” of nine simultaneous reactions, workup, and HPLC yield analysis. Companies such as Procter & Gamble, Merck, and Dow applied robotic systems to perform titrations, make reaction mixtures for kinetic studies, and screen compounds for electrochemical activity, respectively.

Automation was also reaching full commercial reality in the early to mid-1980s for the synthesis and sequencing of biological molecules, a major spark for the looming genomic revolution.

sign, Biosym Technologies, and Polygen in 1984; and Oxford Molecular Group in 1989.

In 1984, Molecular Designs introduced MACCS-II, the first system to integrate chemical structures with data, which gave scientists the ability to create data forms paired with each structure and connect them to databases that could be browsed, searched, and updated. Daylight Chemical Information Systems, another important chemical database company, was founded in 1987.

That same year, the popularity of Pople’s Gaussian programs had grown so great that Pople founded Gaussian, Inc., in Pittsburgh to sell and distribute the product—the first commercialization of *ab initio* software. However, Pople left Gaussian in 1991 following disagreements with the company over profit distribution.

During the same period, Benny Johnson, a former Pople graduate student at Carnegie Mellon who was helping develop technology for Gaussian, got into an intellectual property dispute with the company. This led Johnson, along with Peter Gill and Carlos Gonzalez, to found the competing company Q-Chem in 1993. Pople joined Q-Chem’s board of directors in 1999 after winning the 1998 Nobel Prize in Chemistry.

Matching up three-dimensional structures to active protein sites was becoming a challenge of substantial interest in the 1970s and 1980s. The Brookhaven Protein Data Bank (later to become the Research Collaboratory for Structural Bioinformatics, or RCSB, Protein Data Bank) began development in the 1970s. In 1982, Irwin Kuntz at the University of California, San Francisco, developed an algorithm (that would later be developed into the widely used DOCK program) for matching up the shape and electronics of small molecules with protein receptor pockets. At the time, though, the technique was limited because it relied on X-ray crystal structures to be available for the small molecules. But in 1987, another algorithm became widely available to pharmaceutical companies, called CONCORD, which allowed rapid computation of an approximate three-dimensional structure from two-dimensional input. This allowed large three-dimensional databases to be built.

By 1988, Molecular Designs, called MDL Information Systems after being purchased by Maxwell Communications, released the first commercial three-dimensional structural database. Shortly after, Chemical Designs followed up with its Chem-3DBS database.

Experimental Acceleration

Chemists in the mid-1960s were looking to use the newly available minicomputers—which were small enough to fit in a typical laboratory space—not just to precede or supplant experimental efforts, but also to make the experiments themselves faster and more effective.

X-ray crystallographers were among the first to use minicomputers for automating chemical research, specifically to control diffractometers. Chemists soon began to yield control to computers in setting up and collecting data in electrochemistry experiments, and the rest of analytical chemistry soon followed. In 1969, Hewlett-Packard introduced the first robotic sample injector for chromatography, which allowed samples to be analyzed while the system was unattended.

In the 1970s, the microprocessor age began with the introduction of the Intel microchip. Furthermore, standard bus structures began to appear, which allowed computers and instruments to communicate more efficiently via parallel channels. Analytical instrument makers, such as Hitachi, Beckman, PerkinElmer, and Shimadzu, commercialized the first microprocessor-controlled spectrophotometers in this period.

As computers became faster and more available at the turn of the decade—IBM introduced the first PC in 1981—expectations grew for greater efficiency and speed in the laboratory. So when Millipore acquired the chromatography powerhouse Waters Corporation in 1980, Waters president Frank Zenie left and founded Zymark, a company based completely around laboratory automation (see sidebar).

Applied Biosystems, formed in 1981 out of the collaboration between Marvin Caruthers of the University of Colorado and Leroy Hood of the California Institute of Technology, introduced the 470A, the first automated protein sequencer, in 1982. The next year, it launched the first commercially successful DNA synthesizer, followed by a peptide synthesizer the following year and the first automated DNA sequencer in 1986.

In an attempt to handle the massive quantities of data resulting from all of the newfound automation, companies began producing laboratory information management systems (LIMS). PerkinElmer’s LIMS 2000 was released in 1981, and in 1983, Varian and DEC jointly released the VAX-11 LIMS, which included Ethernet capability. Also that year, Nelson Analytical, founded by David Nelson in 1980, offered the first widely used commercial chromatography data system software for desktop computers for performing LC and GC experiments and collecting and storing the data. PerkinElmer acquired Nelson Analytical in 1989.

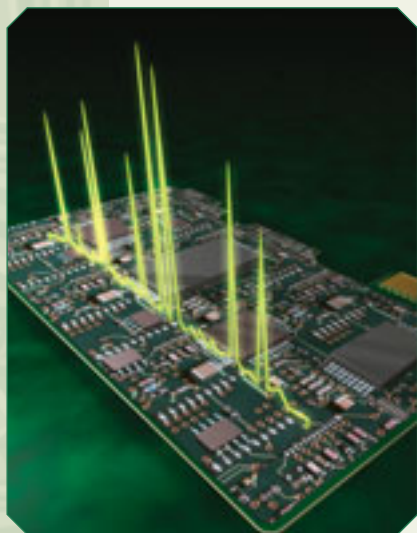
Pervasive Technology

The seeds of today’s computer pervasiveness in the laboratory continued to be sown as the 1980s progressed. MDL offered the first PC-based chem-

Above: Zymark robotics ad, *Analytical Chemistry*

istry database system, CPSS, in 1985, and also that year Tripos introduced *Alchemy*, the first desktop molecular modeling system. In 1986, Cambridge Scientific Computing (later to become Cambridge-Soft) was founded and soon began developing the popular desktop chemical structure drawing program *ChemDraw*.

As the 1990s rolled in, so did the data, particularly in the endeavor of drug discovery and genomics. Increasingly efficient automation and combinatorial chemistry techniques were beginning to produce an enormous number of compounds, and the concept of robot-based parallel high-throughput screening was starting to catch on (in 1990, the term “high-throughput” appeared in seven American Chemical Society journal papers; by 1995, it appeared in a total of 127 papers; and by 2000, in a total of 743 papers).



The need for computational processing of this information as a prerequisite to performing more experiments became a requirement. In the early 1990s, Tripos introduced its first CoMFA quantitative structure-activity relationship tools that create predictive statistical and graphical models of bioactivity. In the mid-1990s, MDL

Information Systems (which became a publicly traded company in 1993) and Tripos offered desktop tools for designing and choosing combinatorial libraries for high-throughput biological screening.

Far-Reaching Project

The Human Genome Project—which began in 1990 and was essentially completed in 2003—and the myriad areas of study that extend from it (e.g., proteomics, pharmacogenomics) epitomize the modern scientific application of automation and computerization.

Applied Biosystems's ABI Prisms 3700 gene sequencers played an essential role in the Celera Genomics “shotgun” approach to sequencing. And Stephen Fodor founded Affymetrix in 1992, based on an automated photolithography method for producing the GeneChip, a transforming technology for the genome project and genomic-based research in general.

Software was and is commonly produced in-house (in industry and academia) and sometimes is provided free over the Internet for genomic analy-

sis, but commercial software development has also been a significant result of the genomic endeavor. MDL, Tripos, Oxford Molecular Group, Molecular Simulations, and other computational chemistry firms were exhibiting new bioinformatics tools for analyzing and tabulating gene and protein sequences by the 1997 ACS National Meeting in San Francisco, according to a *Chemical & Engineering News* report from the conference.

New companies were formed to meet this need as well. LION bioscience was formed in 1997 to exploit intellectual property developed at a consortium of European molecular biology and informatics groups. The company played a significant role in the Human Genome Project, as its informatics tools were used by participating companies such as Celera Genomics and Incyte Genomics (now Incyte Corp.)

In the mid- to late-1990s, computational companies started to more vigorously seek to be comprehensive informatics “solutions” for pharmaceutical companies.

MDL, for instance, which Elsevier acquired in 1997, purchased Interactive Simulations, a molecular modeling company, and SciVision, a QSAR company, to expand its offerings. In 2001, Pharmacoepia combined several of its recent acquisitions—Molecular Simulations, Synopsis Scientific Systems, and Oxford Molecular Group (which, itself, acquired several software companies, including Chemical Design and toxicology prediction software company Health Designs, throughout the 1990s)—to form Accelrys, a company that integrates chemoinformatic, bioinformatic, database, and molecular modeling tools.

On the laboratory automation front, PerkinElmer, Tecan, and Zymark remain big players in the market, along with Beckman Coulter, which was formed in 1998 when Beckman Instruments acquired Coulter Corp. In 2003, Caliper Technologies (now Caliper Life Sciences), producer of lab-on-a-chip technologies, acquired Zymark, combining two core modern drug discovery technologies, microfluidics and liquid-handling robots.

Computational Modeling

The automation and computerization of chemistry have, of course, not only been about drug discovery. Computational modeling tools such as those produced by Accelrys are increasingly applied to the design of catalysts and other advanced materials. Symyx Technologies, Torial Technologies, and several other companies are pursuing combinatorial synthesis and high-throughput analysis of catalyst candidates, which is inevitably leading to large catalyst databases.

As computational speed increases in general, so will the speed of chemistry. The need for greater productivity will continue to drive innovation in automation, computational chemistry, and the overall laboratory research endeavor. ♦

Above: Artist's rendering of LIMS, *Today's Chemist at Work*, 2004