

PART I

COMPUTERS IN PHARMACEUTICAL RESEARCH AND DEVELOPMENT: A GENERAL OVERVIEW

COPYRIGHTED MATERIAL

1

HISTORY OF COMPUTERS IN PHARMACEUTICAL RESEARCH AND DEVELOPMENT: A NARRATIVE

DONALD B. BOYD AND MAX M. MARSH

Contents

- 1.1 Introduction
- 1.2 Computational Chemistry: the Beginnings at Lilly
- 1.3 Germination: the 1960s
- 1.4 Gaining a Foothold: the 1970s
- 1.5 Growth: the 1980s
- 1.6 Fruition: the 1990s
- 1.7 Epilogue
- Acknowledgments
- References

1.1 INTRODUCTION

Today, computers are so ubiquitous in pharmaceutical research and development that it may be hard to imagine a time when there were no computers to assist the medicinal chemist or biologist. A quarter-century ago, the notion of a computer on the desk of every scientist and company manager was not even contemplated. Now, computers are absolutely essential for generating, managing, and transmitting information. The aim of this chapter is to give a

Computer Applications in Pharmaceutical Research and Development, Edited by Sean Ekins.
ISBN 0-471-73779-8 Copyright © 2006 John Wiley & Sons, Inc.

brief account of the historical development. It is a story of ascendancy and one that continues to unfold.

Owing to the personal interest and experience of the authors, the emphasis in this chapter is on using computers for drug discovery. But the use of computers in laboratory instruments and for analysis of experimental and clinical data is no less important. This chapter is written with young scientists in mind. We feel it is important that the new investigator have an appreciation of how the field evolved to its present circumstance, if for no other reason than to help steer toward a better future for those scientists using or planning to use computers in the pharmaceutical industry.

Computers began to be deployed at pharmaceutical companies as early as the 1940s. These early computers were usually for the payroll and for accounting, not for science. Pharmaceutical scientists did eventually gain access to computers, if not in the company itself, then through contractual agreements with nearby educational institutions or other contractors.

There were several scientific and engineering advances that made possible a computational approach to what had long been exclusively an experimental art and science, namely, discovering a molecule with useful therapeutic potential. One fundamental concept understood by chemists was that chemical structure is related to molecular properties including biological activity. Hence if one could predict properties by calculations, one might be able to predict which structures should be investigated in the laboratory. Another fundamental, well-established concept was that a drug would exert its biological activity by binding to and/or inhibiting some biomolecule in the body. This concept stems from Fischer's famous lock-and-key hypothesis (Schlüssel-Schloss-Prinzip) [1, 2]. Another advance was the development of the theory of quantum mechanics in the 1920s [3]. This theory connected the distribution of electrons in molecules with observable molecular properties. Pioneering research in the 1950s attacked the problem of linking electronic structure and biological activity. A good part of this work was collected in the 1963 book by Bernard and Alberte Pullman of Paris, France, which fired the imagination of what might be possible with calculations on biomolecules [4]. The earliest papers that attempted to mathematically relate chemical structure and biological activity were published in Scotland all the way back in the middle of the nineteenth century [5, 6]. This work and a couple of other papers [7, 8] were forerunners to modern quantitative structure-activity relationships (QSAR) but were not widely known. In 1964, the role of molecular descriptors in describing biological activity was reduced to a simplified mathematical form, and the field of QSAR was propelled toward its modern visage [9, 10]. (A descriptor is any calculated or experimental numerical property related to a compound's chemical structure.) And, of course, there was the engineering development of computers and all that entailed. The early computers were designed for military and accounting applications, but gradually it became apparent that computers would have a vast number of uses.

One of us (MMM) was one of the first people in the pharmaceutical industry to perceive that computer-aided drug design was something that might be practical and worthy of investigation. He pioneered a sustained, industrial research program to use computers in drug design. After retiring from Eli Lilly and Company in 1986, he became a Visiting Research Scientist and later an Adjunct Professor in the Department of Chemistry, Indiana University, Bloomington. Section 1.2 is his personal account of the early steps at Lilly.

1.2 COMPUTATIONAL CHEMISTRY: THE BEGINNINGS AT LILLY

This narrative was first presented at Don Boyd's third annual Central Indiana Computational Chemistry Christmas Luncheon (CICCCCL-3) on December 18, 1997. Although it is specific for Eli Lilly and Company, the progress and problems that transpired there were probably not too different from developments at other large, forward-looking, research-based pharmaceutical companies.

This little story contains mainly my personal recollection about how the computational chemistry project in the Lilly Research Laboratories began. An advantage of living longer than one's contemporaries is that there is no one around among the early participants to contradict my reminiscences. A more comprehensive history of this discipline may be found in the Bolcer and Hermann chapter in *Reviews of Computational Chemistry* [11]. I shall confine this commentary to what I remember about my own involvement.

I began work at Eli Lilly and Company in March 1942 as a laboratory aide in the analytical department. At that time, there was very little sophisticated instrumentation in the laboratory. The most complex calculations were carried out using a slide rule. After military service in World War II and an educational leave of absence to complete my undergraduate studies in chemistry at Indiana University, I returned to the Lilly analytical group in 1947. Slide rules were still much in evidence but were soon augmented with mechanical calculators—usually Monroe or Friden models.

It was not until 1949 that the company actually acquired a stored-program computer; at that time an IBM 704 system was purchased—for about \$1 million. In spite of the fact that it was a vacuum tube machine—with considerable concomitant downtime—several business operations were carried out using it. A number of inventories and the payroll were successfully handled. However, no scientific calculations were performed with it. The system was replaced in a few years with an IBM 709—again only for business and financial operations.

In the late 1950s or early 1960s, the first computers to have stored programs of scientific interest were acquired. One of these was an IBM 650; it had a rotating magnetic drum memory consisting of 2000 accessible registers. The programs, the data input, and the output were all in the form of IBM punched cards. A major concern was keeping those card decks intact and in order as they were

moved about from user to machine and back. My recollection is that some statistical calculations by Lilly's research statistics group under Dr. Edgar King were carried out on this machine.

At about the same time, one of the business groups obtained an IBM 610 computer. This device was simpler to use than the 650, and it utilized punched paper tape input, program, and output. The tape was generated on a typewriter. Programs were developed using an essentially algebraic language peculiar to the machine. After the program tape was read in, a tape containing sequentially the data to be processed was fed in. The output tape was carried back to a tape reader linked to a typewriter where the results were ultimately typed out. I used this machine.

My interest at that time revolved around evaluating optical rotary dispersion data [12]. The paired values of optical rotation vs. wavelength were used to fit a function called the Drude equation (later modified to the Moffitt equation for William Moffitt [Harvard University] who developed the theory) [13]. The coefficients of the evaluated equation were shown to be related to a significant ultraviolet absorption band of a protein and to the amount of alpha-helix conformation existing in the solution of it.

Interest in possible applications of computers at the Lilly Research Laboratories began to broaden in the early 1960s. Dr. King (then director of the statistical research group) and I appeared before the Lilly board of directors, submitting a proposal to acquire a computer and ancillary equipment to be devoted primarily to research needs. The estimated cost was a little more than \$250,000. In those days, an expenditure of that large an amount required board approval. Today, I suppose a division director or even a department head could sign off on a personal computer with vastly more power than any computer of the 1960s!

The board of directors approved our proposal. The system that was purchased was an IBM 1620 with the necessary card punches and reader plus tape drives. In addition to statistical and some analytical chemistry applications, Dr. Charles Rice (then head of the radiochemistry group) and I initiated Lilly's first computer-based information retrieval system. Through an agreement with the Institute for Scientific Information (ISI, Philadelphia), Lilly was able to receive magnetic tapes containing computer-searchable title information on current scientific journals from ISI every 10 days. Interest profiles of individual Lilly scientists were then used to generate the famous (or infamous!) "hit" cards that were distributed to members of the research staff. The cards contained journal citations to articles matching the recipient scientist's profile. This service continued until the advent of electronic literature alerts in the 1980s.

Stemming from my growing interest in and enthusiasm for the potential use of computed values of atomic and molecular properties in pharmaceutical research, I was able to gain approval for a requisition for a scientist who knew how to use computers to determine molecular properties. The person I hired was Dr. Robert B. Hermann, our first theoretical chemist. It was 1964. He obtained his Ph.D. with Prof. Norman L. Allinger at Wayne State and then did postdoctoral research with Prof. Joseph O. Hirschfelder at Wisconsin and with Prof. Peter Lykos at the Illinois Institute of Technology. When Bob joined us, he brought along a semiempirical molecular orbital program that he had personally written. He

planned to use this to estimate molecular properties of drug-type molecules, but Lilly computers were incapable of handling the necessary matrix multiplication steps. This obstacle was overcome by going outside the company. We were able to develop a working agreement with the engineering component of Allison Transmission Division of General Motors to use their IBM 7094 after regular working hours. Since the system was used only by Allison and Lilly, data security was not an issue. However, considerable time was spent transporting punched card decks and printouts between the Lilly Research Laboratories near downtown Indianapolis and the Allison facility in nearby Speedway, Indiana.

Looking back, it is difficult for me to pinpoint the factors leading to my initiation of the molecular modeling and drug design effort at Lilly. Certainly, the developments of Prof. Lou Allinger and his associates (at Wayne State and the University of Georgia) in the 1960s to use calculations to study conformation played an important part [14]. Similarly, the publishing of an EHT program by Prof. Roald Hoffmann (Harvard University) in 1963 was a significant impetus. The introduction of the pi-sigma correlation equation by Prof. Corwin Hansch (Pomona College) in 1964 added another facet of interest. Also that year, Dr. Margaret Dayhoff (a theoretical chemist who became the first prominent woman in what would become the field of bioinformatics and who was at the National Biomedical Research Foundation in Maryland) published a method for arriving at the geometry of a polypeptide or protein via internal coordinates [15]. This methodology also encouraged me to begin thinking about enzyme-inhibitor interactions and the three-dimensional requirements for molecular design.

It was not until 1968, when Don Boyd joined us as the second theoretical chemist in our group, that the computers at Lilly started to reach a level of size, speed, and sophistication to be able to handle some of the computational requirements of our various evaluation and design efforts. Don brought with him Hoffmann's EHT program from Harvard and Cornell. Due to the length of our calculations and due to the other demands on the computer, the best we could obtain was a one-day turnaround.

The preceding years involved not only the Allison agreement (for which we paid a modest fee) but also later ones with Purdue University (West Lafayette, Indiana) and Indiana University, Bloomington computing centers. These latter arrangements involved Control Data Corporation (CDC) systems that were much faster than the IBM 7094. Use of the Purdue computer, which continued after Don joined our group, involved driving to the near north side of Indianapolis where the Purdue extension campus was located. In the basement of their science building was a computer center connected to the CDC 7600 in West Lafayette. Computer card decks of data and the associated program for approximate molecular orbital calculations could be left with the machine operators. With luck, the card decks and computer printouts could be retrieved the next day. Security was more of a problem with the academic facilities because they had a large number of users. The concern was enhanced when—on one occasion—I received, in addition to my own output, the weather forecast data and analysis for the city of Kokomo, Indiana! Even though it was unlikely that anyone could make use of our information except Bob, Don, or myself, it was a relief to research management when we were able to carry out all our computations in-house.

These reminiscences cover about the first 15 years of the Lilly computational chemistry effort. Considering the strong tradition of lead generation emanating from the organic chemistry group, the idea that molecular modeling could make a significant contribution to drug design was slow to be accepted. Nevertheless, enough research management support was found to spark the small pioneering project and to keep it going in the face of strong skepticism. Regrettably, a considerable amount of my time in this critical period was spent attempting to convince management and the scientific research staff of the logic and significance of these studies. Because we entered the field at a very early stage, a great deal of effort went into the testing, evaluation, and establishment of the limits of application of the various computational methods. This kind of groundwork was not always well understood by the critics of our approach.

In what follows, we review events, trends, hurdles, successes, people, hardware, and software. We attempt to paint a picture of happenings as historically correct as possible but, inevitably, colored by our own experiences and memories. The time line is broken down by decade from the 1960s through the turn of the century. We conclude with some commentary on where the field is headed and lessons learned. For some of the topics mentioned, we could cite hundreds of books [16] and thousands of articles. We hope that the reader will tolerate us citing only a few examples. We apologize to our European and Japanese colleagues for being less familiar with events at their companies than with events in the United States. Before we start, we also apologize sincerely to all the many brilliant scientists who made landmark contributions that we cannot cover in a single chapter.

1.3 GERMINATION: THE 1960s

We can state confidently that in 1960 essentially 100% of the computational chemists were in academia, not industry. Of course, back then they were not called computational chemists, a term not yet invented. They were called theoretical chemists or quantum chemists. The students coming from those academic laboratories constituted the main pool of candidates that industry could hire for their initial ventures into using computers for drug discovery. Another pool of chemists educated using computers were X-ray crystallographers. Some of these young theoreticians and crystallographers were interested in helping solve human health challenges and steered their careers toward pharmaceutical work.

Although a marvel at the time, the workplace of the 1960s looks archaic in hindsight. Computers generally resided in computer centers, where a small army of administrators, engineers, programming consultants, and support people would tend the mainframe computers then in use. The computers were kept in locked, air-conditioned rooms inaccessible to ordinary users. One of the largest computers then in use by theoretical chemists and crystallographers was the IBM 7094. Support staff operated the tape readers, card readers,

and printers. The users' room at the computer centers echoed with the clunk-clunk-clunk of card punches that encoded data as little rectangular holes in the so-called IBM cards [see reference 11]. The cards were manufactured in different colors so that users could conveniently differentiate their many card decks. As a by-product, the card punches produced piles of colorful rectangular confetti. There were no Delete or Backspace keys; if any mistake was made in keying in data, the user would need to begin again with a fresh blank card. The abundance of cards and card boxes in the users' room scented the air with a characteristic paper smell. Programs were written in FORTRAN II. Programs used by the chemists usually ranged from half a box to several boxes long. Carrying several boxes of cards to the computer center was good for physical fitness. If a box was dropped or if a card reader mangled some of the cards, the tedious task of restoring the deck and replacing the torn cards ensued. Input decks were generally smaller—consisting of tens of cards—and were sandwiched between JCL (job control language for IBM machines) cards and bound by rubber bands. Computer output usually came in the form of ubiquitous pale green and white striped paper (measuring 11 by 14-7/8 inches per page). Special cardboard covers and long nylon needles were used to hold and organize stacks of printouts.

Mathematical algorithms for common operations such as matrix diagonalization had been written and could be inserted as a subroutine in a larger molecular orbital program, for instance. Programs for chemistry were generally developed by academic groups, with the graduate students doing most or all of the programming. Partly, this was standard practice because the professors at different universities were in competition with each other and wanted a better program than their competitors had access to. (Better means running faster, handling larger matrices, and doing more.) Partly, this situation was standard practice so that the graduate students would learn by doing. Obviously, this situation led to much duplication of effort: the proverbial reinventing the wheel. To improve this situation, Prof. Harrison Shull and colleagues at Indiana University, Bloomington, conceived and sold the concept of having an international repository of software that could be shared. Thus was born in 1962 the Quantum Chemistry Program Exchange (QCPE). Competitive scientists were initially slow to give away programs they worked so hard to write, but gradually the depositions to QCPE increased. We do not have room here to give a full recounting of the history of QCPE [17], but suffice it to say that QCPE proved instrumental in advancing the field of computational chemistry including that at pharmaceutical companies. Back in the 1960s and 1970s, there were no software companies catering to the computational chemistry market, so QCPE was the main resource for the entire community. As the name implies, QCPE was initially used for exchanging subroutines and programs for ab initio and approximate electronic structure calculations. But QCPE evolved to encompass programs for molecular mechanics and a wide range of calculations on molecules. The quarterly *QCPE Newsletter* (later renamed the *QCPE Bulletin*), which was edited by Mr. Richard W. Counts,

was for a long time the main vehicle for computational chemists to announce programs and other news of interest. QCPE membership included industrial computational chemists.

Finally in regard to software, we note one program that came from the realm of crystallography. That program was ORTEP (Oak Ridge Thermal Ellipsoid Program), which was the first widely used program for (noninteractive) molecular graphics [18]. Output from the program was inked onto long scrolls of paper run through expensive, flat-bed printers. The ball-and-stick ORTEP drawings were fine for publication, but routine laboratory work was easier with graph paper, ruler, protractor, and pencil to plot the Cartesian coordinates of a molecule the chemist wanted to study. Such handmade drawings quantitated molecular geometry. Experimental bond lengths and bond angles were found in a British compilation [19].

Also to help the chemist think about molecular shape, hand-held molecular models were widely used by experimentalists and theoreticians alike. There were two main types. One was analogous to stick representations in which metal or plastic rods represented bonds between atoms, which were balls or joints that held the rods at specific angles. Metal wire Drieding models were among the most accurate and expensive. The other type was space filling. The expensive CPK (Corey–Pauling–Koltun) models [20, 21] consisted of three-dimensional spherical segments made of plastic that were color-coded by element (white for hydrogen, blue for nitrogen, red for oxygen, etc.). From this convention, came the color molecular graphics we are familiar with today.

In the 1960s, drug discovery was by trial and error. Interesting compounds flowed from two main sources in that period. The smaller pipeline was natural products, such as soil microbes that produced biologically active components or plants with medicinal properties. The dominant pipeline, however, was classic medicinal chemistry. A lead compound would be discovered by biological screening or by reading the patent and scientific literature published by competitors at other pharmaceutical companies. From the lead, the medicinal chemists would use their ingenuity, creativity, and synthetic expertise to construct new compounds. These compounds would be tested by the appropriate in-house pharmacologists, microbiologists, and so forth. Besides the intended biological target, the compounds would often be submitted to a battery of other bioactivity screens being run at the company so that leads for other drug targets could be discovered. The most potent compounds found would become the basis for another round of analog design and synthesis. Thus would evolve in countless iterations a structure-activity relationship (SAR), which in summary would consist of a table of compounds and their activities. In fortuitous circumstances, one of the medicinal chemists would make a compound with sufficient potency that a project team consisting of scientists from drug discovery and drug development would be assembled to oversee further experiments on the compound to learn whether it had the appropriate characteristics to become a pharmaceutical product. The formula

for career success was simple: The medicinal chemist who invented or could claim authorship of a project team compound would receive kudos from management.

What happens when a theoretical chemist is thrown into this milieu? Well, initially not much because the only theoretical methods of the 1960s that could treat drug-sized (200–500 Da) molecules were inaccurate and limited. These molecular orbital methods were extended Hückel theory [22, 23] and soon thereafter CNDO/2 (complete-neglect-of-differential-overlap/second parameterization) [24, 25]. Although approximate by today's standards and incapable of giving accurate, energy-minimized ("optimized"), three-dimensional molecular geometries (bond lengths, bond angles, and torsional angles), they were far more appropriate for use than other methods available at the time. One of these other methods was Hartree–Fock [26–29] (also called self-consistent field or nonempirical in the early literature, or *ab initio* in recent decades). Although Hartree–Fock did fairly well at predicting molecular geometries, the computers of the era limited treatment to molecules not much larger than ethane. Another class of methods such as simple Hückel theory [30–32] and Pariser–Parr–Pople (PPP) theory [33] could treat large molecules but only pi electrons. Hence, they were formally limited to planar molecules, but not many pharmaceuticals are planar.

In addition to the quantum chemistry in use in the 1960s, an independent approach was QSAR, as already alluded to. Here the activity of a compound is assumed to be a linear (or quadratic or higher) function of certain molecular descriptors. One of the commonly used descriptors was the contribution of an atom or a functional group to the lipophilicity of a molecule; this descriptor was designated pi (π). Other famous descriptors included the Hammett sigma (σ) values for aromatic systems and the Taft sigma (σ^*) values for aliphatic systems; both came from physical organic chemistry [34–36]. The sigma values measured the tendency of a substituent to withdraw or donate electron density in relation to the rest of the molecule.

Abbott, Schering-Plough, and Upjohn were among the first companies, besides Lilly, to venture into the area of using computers for attempts at drug discovery. Dow Chemical, which had pharmaceutical interests, also initiated an early effort. Generally, the first steps consisted of either hiring a person with theoretical and computer expertise or allowing one of the company's existing research scientists to turn attention to learning about this new methodology. Much effort was expended by these early pioneers in learning the scope of applicability of the available methods. Attempts to actually design a drug were neither numerous nor particularly successful. This generalization does not imply that there were no successes. There were a few successes in finding correlations and in better understanding what was responsible for biological activity at the molecular level. For example, early work at Lilly revealed the glimmer of a relationship between the calculated electronic structure of the beta-lactam ring of cephalosporins and antibacterial activity. The work was performed in the 1960s but was not published until 1973 [37] because of delays by cautious research management and patent

attorneys at the company. (The relationship was elaborated in subsequent years [38, 39], but no new pharmaceutical product resulted [40].)

1.4 GAINING A FOOTHOLD: THE 1970s

Some of the companies that first got into this game dropped out after a few years (but returned later), either for lack of management support or because the technology was not intellectually satisfying to the scientist involved. Other companies, like Lilly, persisted. Lilly's pioneering effort paid off in establishing a base of expertise. Also, quite a few papers were published, almost like in an academic setting. In hindsight, however, Lilly may have gotten in the field too early because the initial efforts were so limited by the science, hardware, and software. First impressions can be lasting. Lilly management of the 1970s thwarted further permanent growth but at least sustained the effort. (It was not until near the end of the 1980s that Lilly resumed growing its computational chemistry group to catch up to the other large pharmaceutical companies.) It was generally recognized that Lilly was a family-oriented company committed to doing what was right in all phases of its business. There was great mutual loyalty between the company and the employees. Other companies such as Merck and Smith Kline and French (using the old name) entered the field a few years later. Unlike Lilly, they hired chemists trained in organic chemistry and computers and with a pedigree traceable back to Prof. E. J. Corey at Harvard and his attempts at computer-aided synthesis planning [41–43].

Regarding hardware of the 1970s, pharmaceutical companies invested money from the sale of their products to buy better and better mainframes. Widely used models included members of the IBM 360 and 370 series. Placing these more powerful machines in-house made it easier and more secure to submit jobs and retrieve output. But output was still in the form of long print-outs. Input had advanced to the point where punch cards were no longer needed. So-called dumb terminals, that is, terminals with no local processing capability, could be used to set up input jobs for batch running. For instance, at Lilly an IBM 3278 and a Decwriter II (connected to a DEC-10 computer) were used by the computational chemistry group. The statistics program MINITAB was one of the programs that ran on the interactive Digital Equipment Corporation machine. Card punches were not yet totally obsolete, but received less and less use. The appearance of a typical office for computational chemistry is shown in Figure 1.1.

The spread of technology at pharmaceutical companies also meant that secretaries were given word processors (such as the Wang machines) to use in addition to typewriters, which were still needed for filling out forms. Keyboarding was the domain of the secretaries, the data entry technicians, and the computational chemists. Only a few managers and scientists would type their own memos and articles.

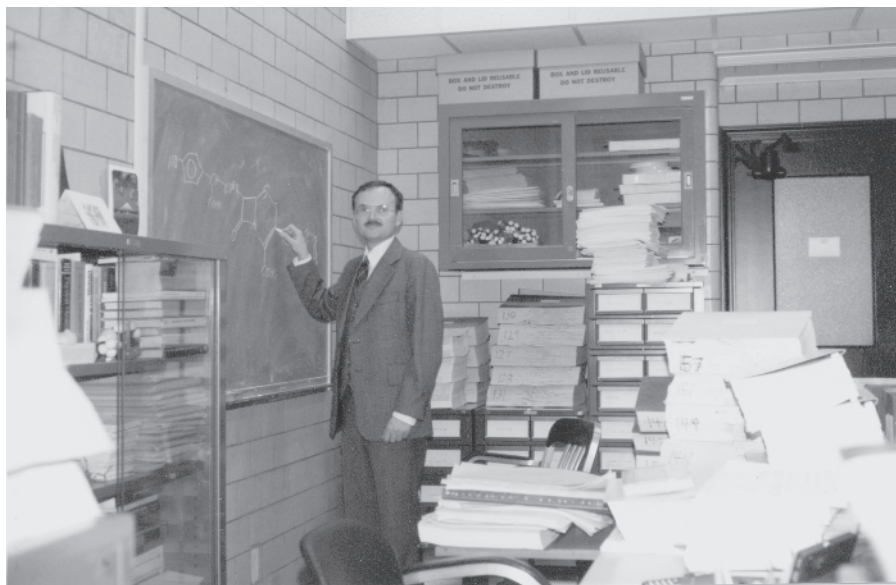


Figure 1.1 Offices used by computational chemists were filled with stacks of print-outs and banks of file cabinets with legacy card decks. This photograph was taken in 1982, but the appearance of the environs had not changed much since the mid-1970s.

Software was still written primarily in FORTRAN, now mainly FORTRAN IV. The holdings of QCPE expanded. Among the important acquisitions was Gaussian 70, an *ab initio* program written by Prof. John A. Pople's group at Carnegie-Mellon University. Pople made the program available in 1973. (He later submitted Gaussian 76 and Gaussian 80 to QCPE, but they were withdrawn when the Gaussian program was commercialized by Pople in 1987.) Nevertheless, *ab initio* calculations, despite all the élan associated with them, were still not very practical or helpful for pharmaceutically interesting molecules. Semiempirical molecular orbital methods (EHT, CNDO/2, MINDO/3) were the mainstays of quantum chemical applications (MINDO/3 [44] was Prof. Michael J. S. Dewar's third refinement of a modified intermediate neglect-of-differential-overlap method).

The prominent position of quantum mechanics led a coterie of academic theoreticians to think that their approach could solve research problems facing the pharmaceutical industry. These theoreticians, who met annually in Europe and on Sanibel Island in Florida, invented the new subfields of quantum biology [45] and quantum pharmacology [46]. These names may seem curious to the uninitiated. They were not meant to imply that some observable aspect of biology or pharmacology stems from the wave-particle

duality seen in the physics of electrons. Rather, the names conveyed to cognoscenti that they were applying their trusty old quantum mechanical methods to compounds discussed by biologists and pharmacologists [47]. However, doing a calculation on a system of pharmacological interest is not the same as designing a drug. For instance, calculating the molecular orbitals of serotonin is a far cry from designing a new serotonin reuptake inhibitor that could become a pharmaceutical product.

Nonetheless, something even more useful came on the software scene in the 1970s. This was Prof. N. L. Allinger's MMI/MMPI program [48, 49] for molecular mechanics. Classic methods for calculating conformational energies date to the 1940s and early 1960s [50, 51]. Copies of Allinger's program could be purchased at a nominal fee from QCPE. Molecular mechanics has the advantage of being much faster than quantum mechanics and capable of generating common organic chemical structures approaching "chemical accuracy" (bond lengths correctly predicted to within about 0.01 Å). Because of the empirical manner in which force fields were derived, molecular mechanics was an anathema to the quantum purists, never mind that Allinger himself used quantum chemistry, too. Molecular mechanics became an important technique in the armamentarium of industrial researchers. Meanwhile, a surprising number of academic theoreticians were slow to notice that the science was transitioning from quantum chemistry to multifaceted computational chemistry [52, 53].

Computational chemists in the pharmaceutical industry also expanded from their academic upbringing by acquiring an interest in force field methods, QSAR, and statistics. Computational chemists with responsibility to work on pharmaceuticals came to appreciate the fact that it was too limiting to confine one's work to just one approach to a problem. To solve research problems in industry, one had to use the best available technique, and this did not mean going to a larger basis set or a higher level of quantum mechanical theory. It meant using molecular mechanics or QSAR or whatever.

Unfortunately, the tension between the computational chemists and the medicinal chemists at pharmaceutical companies did not ease in the 1970s. Medicinal chemists were at the top of the pecking order in corporate research laboratories. This was an industry-wide problem revealed in conversations at scientific meetings where computational chemists from industry (there were not many) could informally exchange their experiences and challenges. (Readers should not get the impression that the tension between theoreticians and experimentalists existed solely in the business world. It also existed in academic chemistry departments.)

The situation was that as medicinal chemists pursued an SAR, calculations by the computational chemists might suggest a structure worthy of synthesis. Maybe the design had the potential of being more active. But the computational chemist was totally dependent on the medicinal chemist to test the hypothesis. Suddenly, the medicinal chemist saw himself going from being the

wellspring of design ideas to being a technician who was implementing someone else's idea. Although never intended as a threat to the prestige and hegemony of the organic chemistry hierarchy, proposals from outside that hierarchy were often perceived as such.

Another problem was that on a computer it was easy to change a carbon to a nitrogen or any other element. It was easy to attach a substituent at any position in whatever stereochemistry seemed best for enhancing activity. It was easy to change a six-member ring to a five-member ring or vice versa. Such computer designs were frequently beyond the possibilities of synthetic organic chemistry, or at least beyond the fast-paced chemistry practiced in industry. This situation contributed to the disconnect between the computational chemists and medicinal chemists. What good is a computer design if the molecule is impossible to make?

If the computational chemist needed a less active compound synthesized to establish a computational hypothesis, such as for a pharmacophore, that was totally out of the question. No self-respecting medicinal chemist would want to admit to his management that he purposely spent valuable time making a less active compound. Thus the 1970s remained a period when the relationship between computational chemists and medicinal chemists was still being worked out. Management people, who generally rose from the ranks of medicinal chemists, were often unable to perceive a system for effective use of the input computational approaches might provide. In addition, many managers were not yet convinced that the computational input was worth anything.

The computational chemists at Lilly tackled this problem of a collaboration gap in several ways. One was to keep the communication channels open and constantly explain what was being done, what might be doable, and what was beyond the capabilities of the then-current state of the art. For organic chemists who had never used a computer, it was necessary to gently dispel the notion that one could push a button on a large box with blinking lights and the chemical structure of the next \$200 million drug would tumble into the output tray of the machine. (Back in those days, \$200 million in annual sales was equivalent to a blockbuster drug.) The limited capability to predict molecular properties accurately was stressed by the computational chemists. Moreover, it was up to the human, not the machine, to use chemical intuition to capitalize on relationships found between calculated physical properties and sought-after biological activities. Also, it was important for the computational chemist to avoid theory and technical jargon when talking with medicinal chemists. The computational chemists, to the best of their ability, had to speak the language of the organic chemists, not vice versa.

In an outreach to the medicinal chemists at Lilly, a one-week workshop was created and taught in the research building where the organic chemists were located. (The computational chemists were initially assigned office space with the analytical chemists and later with the biologists.) The workshop covered the basic and practical aspects of performing calculations on

molecules. The input requirements (which included the format of the data fields on the punch cards) were taught for several programs. One program was used to generate Cartesian atomic coordinates. Output from that program was then used as input for the molecular orbital and molecular mechanics programs. Several of the adventurous young Ph.D. organic chemists took the course. The outreach was successful in that it empowered a few medicinal chemists to do their own calculations for testing molecular design ideas. It was a foot in the door. These young medicinal chemists could set an example for the older ones. An analogous strategy was used at some other pharmaceutical companies. For instance, Merck conducted a workshop on synthesis planning for their chemists [54].

Despite these efforts, medicinal chemists were slow to accept what computers were able to provide. Medicinal chemists would bring a research problem to the computational chemists, sometimes out of curiosity about what computing could provide, sometimes as a last resort after a question was irresolvable by any other approach. The question might range from explaining why adding a certain substituent unexpectedly decreased activity in a series of compounds to finding a QSAR for a small set of compounds. If the subsequent calculations were unable to provide a satisfactory answer, there was a tendency among some medicinal chemists to dismiss the whole field. This facet of human nature of scientifically educated people was difficult to fathom. A perspective that was promoted by one of us (DBB) to his colleagues was that computers should be viewed as just another piece of research apparatus. Experiments could be done on a computer just like experiments could be run on a spectrometer or in an autoclave. Sometimes the instrument would give the results the scientist was looking for; other times, the computational experiment would fail. Not every experiment—at the bench or in the computer—works every time. If a reaction failed, a medicinal chemist would not dismiss all of synthetic chemistry. Instead, another synthetic route would be attempted. However, the same patience did not seem to extend to computational experiments.

Finally, in regard to the collaboration gap, the importance of a knowledgeable (and wise) mentor—an advocate—cannot be overstated. For a nascent effort to take root in a business setting, the younger scientist(s) had to be shielded from excessive critiquing by higher management and powerful medicinal chemists.

The computational chemists were able to form collaborations with their fellow physical chemists. Some of the research questions dealt with molecular conformation and spectroscopy. The 1970s were full of small successes such as finding correlations between calculated and experimental properties. Some of these correlations were published. Even something so grand as the *de novo* design of a pharmaceutical was attempted but was somewhat beyond reach.

Two new computer-based resources were launched in the 1970s. One was the Cambridge Structural Database (CSD) [55], and the other was the Protein

Data Bank (PDB) [56]. Computational chemists recognized that these compilations of 3D molecular structures would prove very useful, especially as more pharmaceutically relevant compounds were deposited. The CSD was supported by subscribers, including pharmaceutical companies. On the other hand, the PDB was supported by American taxpayers.

We have not discussed QSAR very much, but one influential book of the 1970s can be mentioned [57]. Dr. Yvonne Martin began her scientific career as an experimentalist in a pharmaceutical laboratory, but after becoming interested in the potential of QSAR she spent time learning the techniques at the side of Prof. Corwin Hansch and also Prof. Al Leo of Pomona College in California. As mentioned in her book, she encountered initial resistance to a QSAR approach at Abbott Laboratories. Another significant book that was published in the late 1970s was a compilation of substituent constants [58]. These parameters were heavily relied upon in QSAR investigations.

The decade of the 1970s saw the administration in Washington, DC, set the laudatory goal of conquering cancer. Large sums of taxpayer dollars were poured into the National Cancer Institute for redistribution to worthy academic research projects. Naturally, many professors, including those whose work was related to cancer in the most tenuous and remote way, lined up to obtain a grant. The result was that many academic theoretical chemistry papers published in the 1970s included in their introduction rather farfetched claims as to how the quantum chemical calculations being reported were going to be applicable (someday) to the design of anticancer agents. Computational chemists in industry were not touched by this phenomenon because they were supported by the sales efforts of the manufacturers of the pharmaceuticals and were more focused on the real task of aiding drug discovery.

1.5 GROWTH: THE 1980s

If the 1960s were the Dark Ages and the 1970s were the Middle Ages, the 1980s were the Renaissance, the Baroque Period, and the Enlightenment all rolled into one. The decade of the 1980s was when the various approaches of quantum chemistry, molecular mechanics, molecular simulations, QSAR, and molecular graphics coalesced into modern computational chemistry.

In the world of scientific publishing, a seminal event occurred in 1980. Professor Allinger launched his *Journal of Computational Chemistry*. This helped stamp a name on the field. Before the journal began publishing, the field was variously called theoretical chemistry, calculational chemistry, modeling, etc. Interestingly, Allinger first took his proposal to the business managers for publications of the American Chemical Society (ACS). Unfortunately, they rejected the concept. Allinger turned to publisher John Wiley & Sons, which went on to become the premier producer of journals and books in the field. Nearly 25 years passed before the ACS moved to rectify its mistake, and in 2005 it remodeled its *Journal of Chemical Information and Computer*

Sciences (JCICS) in an attempt to meet the needs of today's computational chemists. JCICS was becoming the most popular venue for computational chemists to publish work on combinatorial library designs (see Fig. 1.2 and Section 1.6 on the 1990s).

Several exciting technical advances fostered the improved environment for computer use at pharmaceutical companies in the 1980s. The first was a development of the VAX 11/780 computer by Digital Equipment Corporation (DEC) in 1979. The machine was departmental size, that is, the price, dimen-

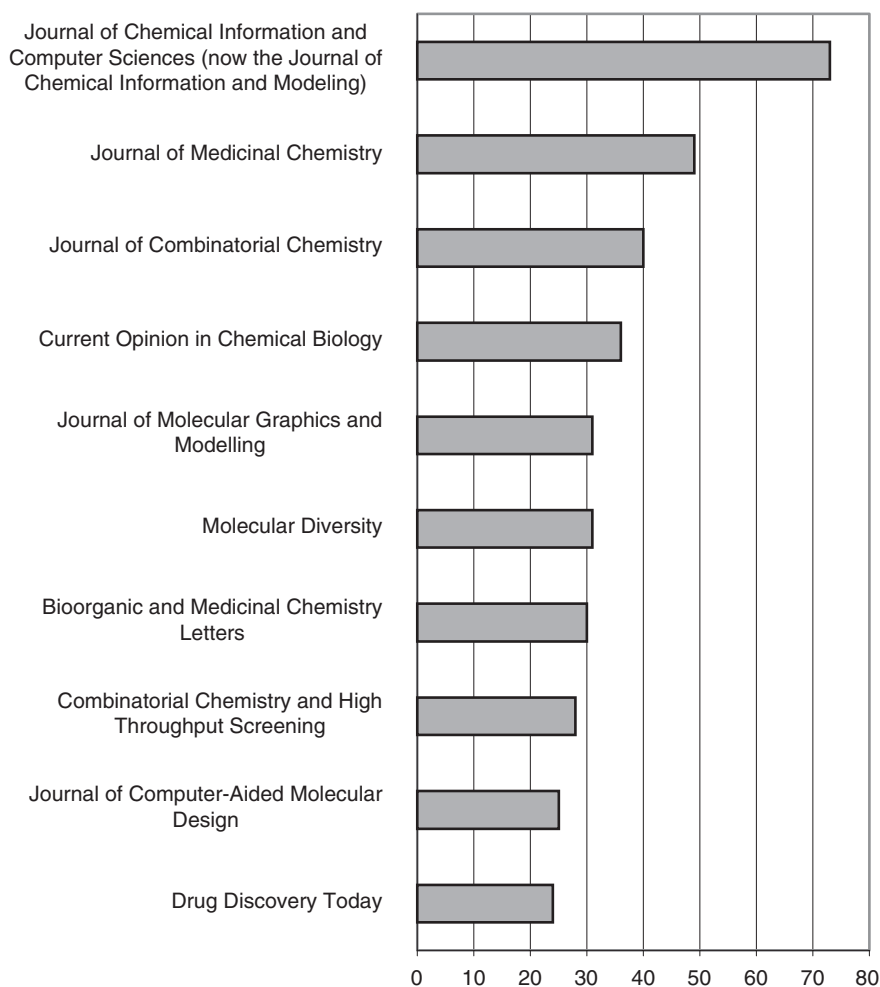


Figure 1.2 Journals that have published the most papers on combinatorial library design. Total number of papers published on this subject according to the Chemical Abstract Service's CAPLUS and MEDLINE databases for all years through 2004 plus three-quarters of 2005.

sions, and easy care of the machine allowed each department or group to have its own superminicomputer. This was a start toward noncentralized control over computing resources. At Lilly, the small-molecule X-ray crystallographers were the first to gain approval for the purchase of a VAX, around 1980. Fortunately, the computational chemists and a few other scientists were allowed to use it, too. The machine was a delight to use and far better than any of the batch job-oriented IBM mainframes of the past. The VAX could be run interactively. Users communicated with the VAX through interactive graphical terminals. The first terminals were monochrome. The first VAX at Lilly was fine for one or two users but would get bogged down, and response times would slow to a crawl if more than five users were logged on simultaneously. Lilly soon started building an ever more powerful cluster of VAXes (also called VAXen in deference to the plural of "ox"). Several other hardware companies that manufactured superminicomputers in the same class as the VAX sprung up. But DEC proved to be a good, relatively long-lasting vendor to deal with, and many pharmaceutical companies acquired VAXes for research. (However, DEC and those other hardware companies no longer exist.)

The pharmaceutical companies certainly noticed the development of the IBM personal computer (PC), but its DOS (disk operating system) made learning to use it difficult. Some scientists bought these machines. The Apple Macintosh appeared on the scene in 1984. With its cute little, lightweight, all-in-one box including monochrome screen, the Mac brought interactive computing to a new standard of user friendliness. Soon after becoming aware of these machines, nearly every medicinal chemist wanted one at work. The machines were great at word processing, graphing, and managing small (laboratory) databases. The early floppy disks formatted for the Macs held only 400 KB, but by 1988 double-sided, double-density disks had a capacity of 1400 KB, which seemed plenty in those days. In contrast to today's huge applications requiring a compact disk for storage, a typical program of the 1980s could be stuffed on one or maybe two floppy disks.

On the software front, three advances changed the minds of the medicinal chemists from being diehard skeptics to almost enthusiastic users. One advance was the development of electronic mail. As the Macs and terminals to the VAX spread to all the chemists in drug discovery and development, the desirability of being connected became obvious. The chemists could communicate with each other and with management and could tap into databases and other computer resources. As electronic traffic increased, research buildings had to be periodically retrofitted with each new generation of cabling to the computers. A side effect of the spread of computer terminals to the desktop of every scientist was that management could cut back on secretarial help for scientists, so they had to do their own keyboarding to write reports and papers.

The second important software advance was ChemDraw, which was released first for the Mac in 1986 [59–62]. This program gave chemists the ability to quickly create two-dimensional chemical diagrams. Every medicinal chemist

could appreciate the aesthetics of a neat ChemDraw diagram. The diagrams could be cut and pasted into reports, articles, and patents. The old plastic ring templates for hand drawing chemical diagrams were suddenly unnecessary.

The third software advance also had an aesthetic element. This was the technology of computer graphics, or when 3D structures were displayed on the computer screens, molecular graphics. Whereas a medicinal chemist might have trouble understanding the significance of the highest occupied molecular orbital or the octanol-water partition coefficient of a structure, he or she could readily appreciate the stick, ball-and-stick, tube, and space-filling representations of 3D molecular structures [63–65]. The graphics could be shown in color and, on more sophisticated terminals, in stereo. These images were so stunning that one director of drug discovery at Lilly decreed that terms like “theoretical chemistry,” “molecular modeling,” and “computational chemistry” were out. The whole field was henceforth to be called “molecular graphics” as far as he was concerned. A picture was something he could understand!

Naturally, with the flood of new computer technology came the need to train the research scientists in its use. Whereas ChemDraw running on a Mac was so easy that medicinal chemists could learn to use it after an hour or less of training, the VAX was a little more formidable. One of the authors (DBB) was involved in preparing and teaching VAX classes offered to the medicinal chemists and process chemists at Lilly.

Computer programs that the computational chemists had been running on the arcane IBM mainframes were ported to the VAXes. This step made the programs more accessible because all the chemists were given VAX accounts. So, although the other programs (e.g., e-mail and ChemDraw) enticed the medicinal chemist to sit down in front of the computer screen, he or she was now more likely to experiment with molecular modeling calculations. (As discussed elsewhere [66], the terms computational chemistry and molecular modeling were used more or less interchangeably at pharmaceutical companies.) Besides the classes and workshops, one-on-one training was offered to help the medicinal chemists run the computational chemistry programs. This was generally fruitful but occasionally led to amusing results such as when one medicinal chemist burst out of his lab to happily announce his discovery that he could obtain a correct-looking 3D structure from MM2 optimization even if he did not bother to attach hydrogens to the carbons. However, he had not bothered to check the bond lengths and bond angles for his molecule.

On a broader front, large and small pharmaceutical companies became aware of the potential for computer-aided drug design. Although pharmaceutical companies were understandably reticent to discuss what compounds they were pursuing, they were quite free in disclosing their computational chemistry infrastructure. For instance, Merck, which had grown its modeling group to be one of the largest in the world, published its system in 1980 [67]. Lilly’s infrastructure was described at a national meeting of the American Chemical Society in 1982 [68].

A few years later, a survey was conducted of 48 pharmaceutical and chemical companies that were using computer-aided molecular design methods and were operating in the United States [69]. Between 1975 and 1985, the number of computational chemists employed at these companies increased from less than 30 to about 150, more than doubling every five years. Thus more companies were jumping on the bandwagon, and companies that were already in this area were expanding their efforts. Hiring of computational chemists accelerated through the decade [70]. Aware of the polarization that could exist between theoretical and medicinal chemists, some companies tried to circumvent this problem by hiring organic chemistry Ph.D.s who had spent a year or two doing postdoctoral research in molecular modeling. This trend was so pervasive that by 1985 only about a fifth of the computational chemists working at pharmaceutical companies came from a quantum mechanical background. Students, too, became aware of the fact that if their Ph.D. experience was in quantum chemistry, it would enhance their job prospects if they spent a year or two in some other area such as molecular dynamics simulations of proteins.

The computational chemistry techniques used most frequently were molecular graphics and molecular mechanics. *Ab initio* programs were in use at 21 of the 48 companies. Over 80% of the companies were using commercially produced software. Two-thirds of the companies were using software sold by Molecular Design Ltd. (MDL). A quarter were using SYBYL from Tripos Associates, and 15% were using the molecular modeling program CHEM-GRAF by Chemical Design Ltd.

The following companies had five or more scientists working full-time as computational chemists in 1985: Abbott, DuPont, Lederle (American Cyanamid), Merck, Rohm and Haas, Searle, SmithKline Beecham, and Upjohn. Some of these companies had as many as 12 people working on computer-aided molecular design applications and software development. For the 48 companies, the mean ratio of the number of synthetic chemists to computational chemists was 29:1. This ratio reflects not only what percentage of a company's research effort was computer based, but also the number of synthetic chemists that each computational chemist might serve. Hence, a small ratio indicates more emphasis on computing or a small staff of synthetic chemists. Pharmaceutical companies with low ratios (less than 15:1) included Abbott, Alcon, Allergan, Norwich Eaton (Proctor & Gamble), and Searle. The most common organizational arrangement (at 40% of the 48 companies) was for the computational chemists to be integrated in the same department or division as the synthetic chemists. The other companies tried placing their computational chemists in a physical/analytical group, in a computer science group, or in their own unit.

About three-quarters of the 48 companies were using a VAX 11/780, 785, or 730 as their primary computing platform. The IBM 3033, 3083, 4341, etc. were being used for molecular modeling at about a third of the companies. (The percentages add up to more than 100% because larger companies had several types of machines.) The most commonly used graphics terminal was

the Evans and Sutherland PS300 (E&S PS300) (40%), followed by Tektronix, Envision, and Retrographics VT640 at about one-third of the companies each and IMLAC (25%). The most-used brands of plotter in 1985 were the Hewlett-Packard and Versatec.

As cited above, the most widely used graphics terminal in 1985 was the E&S PS300. This machine was popular because of its very high resolution, color, speed, and stereo capabilities. (It is stunning to think that a company so dominant during one decade could totally disappear from the market a decade later. Such are the foibles of computer technology.) At Lilly, the E&S PS300 was set up in a large lighted room with black curtains enshrouding the cubicle with the machine. Lilly scientists were free to use the software running on the machine. In addition, the terminal also served as a showcase of Lilly's research prowess that was displayed to visiting Lilly sales representatives and dignitaries. No doubt a similar situation occurred at other companies.

The 1980s saw an important change in the way software was handled. In the 1970s, most of the programs used by computational chemists were exchanged essentially freely through QCPE, exchanged person to person, or developed in-house. But in the 1980s, many of the most popular programs—and some less popular ones—were commercialized. The number of software vendors mushroomed. For example, Pople's programs for *ab initio* calculations were withdrawn from QCPE; marketing rights were turned over to a company he helped found, Gaussian Inc. (Pittsburgh, Pennsylvania). This company also took responsibility for continued development of the software. In the molecular modeling arena, Tripos Associates (St. Louis, Missouri) was dominant by the mid-1980s. Their program SYBYL originally came from academic laboratories at Washington University (St. Louis) [71].

In the arena of chemical structure management, MDL (then in Hayward, California) was dominant. This company, which was founded in 1978 by Prof. Todd Wipke and others, marketed a program called MACCS for management of databases of compounds synthesized at or acquired by pharmaceutical companies. The software allowed substructure searching and later similarity searching [72, 73]. The software was vastly better than the manual systems that pharmaceutical companies had been using for recording compounds on file cards that were stored in filing cabinets. Except for some companies such as Upjohn that had their own home-grown software for management of their corporate compounds, many companies bought MACCS and became dependent on it. As happens in a free market where there is little competition, MACCS was very expensive. Few if any academic groups could afford it. A serious competing software product for compound management did not reach the market until 1987, when Daylight Chemical Information Systems was founded. By then, pharmaceutical companies were so wedded to MACCS that there was great inertia against switching their databases to another platform, even if it was cheaper and better suited for some tasks. In 1982, MDL started selling REACCS, a database management system for chemical reactions. Medicinal chemists liked both MACCS and REACCS. The former could be

used to check whether a compound had previously been synthesized at a company and how much material was left in inventory. The latter program could be used to retrieve information about synthetic transformations and reaction conditions that had been published in the literature.

Some other momentous advances occurred on the software front. One was the writing of MOPAC, a semiempirical molecular orbital program, by Dr. James J. P. Stewart, a postdoctoral associate in Prof. Michael Dewar's group at the University of Texas at Austin [74–76]. The program was the first widely used program capable of automatically optimizing the geometry of molecules. This was a huge improvement over prior programs that could only perform calculations on fixed geometries. Formerly, a user would have to vary a bond length or a bond angle in increments, doing a separate calculation for each, then fit a parabola to the data points and try to guess where the minimum was. Hence MOPAC made the determination of 3D structures much simpler and more efficient. The program could handle molecules large enough to be of pharmaceutical interest. In the days of the VAX, a geometry optimization could run in two or three weeks. An interruption of a run due to a machine shutdown meant rerunning the calculation from the start. For the most part, however, the VAXes were fairly stable.

MOPAC was initially applicable to any molecule parameterized for Dewar's MINDO/3 or MNDO molecular orbital methods (i.e., common elements of the first and second rows of the periodic table). The optimized geometries were not in perfect agreement with experimental numbers but were better than what could have been obtained by prior molecular orbital programs for large molecules (those beyond the scope of *ab initio* calculations). Stewart made his program available through QCPE in 1984, and it quickly became (and long remained) the most requested program from QCPE's library of several hundred [77]. Unlike commercialized software, programs from QCPE were attractive because they were distributed as source code and cost very little.

In the arena of molecular mechanics, Prof. Allinger's continued, meticulous refinement of a experimentally based force field for organic compounds was welcomed by chemists interested in molecular modeling at pharmaceutical companies. The MM2 force field [78, 79] gave better results than MMI. To fund his research, Allinger sold distribution rights for the program initially to Molecular Design Ltd. (At the time, MDL also marketed some other molecular modeling programs.)

A program of special interest to the pharmaceutical industry was CLOGP. This program was developed by Prof. Al Leo (Pomona College) in the 1980s [80–82]. It was initially marketed through Daylight Chemical Information Systems (then of New Orleans and California). CLOGP could predict the lipophilicity of organic molecules. The algorithm was based on summing the contribution from each fragment (set of atoms) within a structure. The fragment contributions were parameterized to reproduce experimental octanol-water partition coefficients, $\log P_{o/w}$. There was some discussion among

scientists about whether octanol was the best organic solvent to mimic biological tissues, but this solvent proved to be the most used. To varying degrees, lipophilicity is related to many molecular properties including molecular volume, molecular surface area, transport through membranes, and binding to receptor surfaces, and hence to many different bioactivities. The calculated $\log P_{o/w}$ values were widely used as a descriptor in QSAR studies in both industry and academia.

Yet another program was Dr. Kurt Enslein's TOPKAT [83, 84]. It was sold through his company, Health Designs (Rochester, New York). The software was based on statistics and was trained to predict the toxicity of a molecule from its structural fragments. Hence compounds with fragments such as nitro or nitroso would score poorly, basically confirming what an experienced medicinal chemist already knew. The toxicological end points included carcinogenicity, mutagenicity, teratogenicity, skin and eye irritation, and so forth. Today, pharmaceutical companies routinely try to predict toxicity, metabolism, bioavailability, and other factors that determine whether a highly potent ligand has what it takes to become a medicine. But back in the 1980s, the science was just beginning to be tackled. The main market for the program was probably government laboratories and regulators. Pharmaceutical laboratories were aware of the existence of the program but were leery of using it much. Companies trying to develop drugs were afraid that if the program, which was of unknown reliability for any specific compound, erroneously predicted danger for a structure, it could kill a project even though a multitude of laboratory experiments might give the compound a clean bill of health. There was also the worry about litigious lawyers. A compound could pass all the difficult hurdles of becoming a pharmaceutical, yet some undesirable, unexpected side effect might show up in some small percentage of patients taking it. If lawyers and lay juries, who frequently have trouble understanding science, the relative merits of different experiments, and the benefit-risk ratio associated with any pharmaceutical product, learned that a computer program had once put up a red flag for the compound, the pharmaceutical company could be alleged to be at fault.

We briefly mention one other commercially produced program. That program was SAS, a comprehensive data management and statistics program. The program was used mainly for handling clinical data that was analyzed by the statisticians at each company. Computational chemists also used SAS and other programs when statistical analyses were needed. SAS also had unique capabilities for graphical presentation of multidimensional numerical data [85] (this was in the days before Spotfire).

With the widespread commercialization of molecular modeling software in the 1980s, came both a boon and a bane to the computational chemist and pharmaceutical companies. The boon was that the software vendors sent marketing people to individual companies as well as to scientific meetings. The marketers would extol the virtues of the programs they were pushing.

Great advances in drug discovery were promised if only the vendor's software systems were put in the hands of the scientists. Impressive demonstrations of molecular graphics, overlaying molecules, and so forth convinced company managers and medicinal chemists that here was the key to increasing research productivity. As a result of this marketing, most pharmaceutical companies purchased the software packages. The bane was that computer-aided drug design (CADD) was oversold, thereby setting up unrealistic expectations of what could be achieved by the software. Also, unrealistic expectations were set for what bench chemists could accomplish with the software. Unless the experimentalists devoted a good deal of time to learning the methods and limitations, the software was best left in the hands of computational chemistry experts.

Also in the 1980s, structure-based drug design (SBDD) underwent a similar cycle. Early proponents oversold what could be achieved through SBDD, thereby causing pharmaceutical companies to reconsider their investments when they discovered that SBDD too was no panacea for filling the drug discovery cornucopia with choice molecules for development. Nevertheless, SBDD was an important advance.

All through the 1970s, computational chemists were often rhetorically quizzed by critics about what if any pharmaceutical product had ever been designed by computer. Industrial computational chemists had a solid number of scientific accomplishments but were basically on the defensive when challenged with this question. Only a few computer-designed structures had ever been synthesized. Only a very tiny percentage of molecules—from any source—ever makes it as far as being a clinical candidate. The stringent criteria set for pharmaceutical products to be used in humans winnowed out almost all molecules. The odds were not good for any computational chemist achieving the ultimate success, a drug derived with the aid of the computer. In fact, many medicinal chemists would toil diligently their whole career and never have one of their compounds selected as a candidate for clinical development.

Another factor was that there were only a few drug targets that had had their 3D structures solved prior to the advancing methods for protein crystallography of the 1980s. One such early protein was dihydrofolate reductase (DHFR), the structures of which became known in the late 1970s [86, 87]. This protein became a favorite target of molecular modeling/drug design efforts in industry and elsewhere in the 1980s. Many resources were expended trying to find better inhibitors than the marketed pharmaceuticals of the antineoplastic methotrexate or the antibacterial trimethoprim. Innumerable papers and lectures sprung from those efforts. Scientists do not like to report negative results, but finally a frank admission came in 1988. A review concluded that none of the computer-based efforts at his company or disclosed by others in the literature had yielded better drugs [88].

Although this first major, widespread effort at SBDD was a disappointment, the situation looked better on the QSAR front. In Japan, Koga employed

classic (Hansch-type) QSAR while discovering the antibacterial agent norfloxacin around 1982 [89–91]. Norfloxacin was the first of the third-generation analogs of nalidixic acid to reach the marketplace. This early success may not have received the notice it deserved, perhaps because the field of computer-aided drug design continued to focus heavily on computer graphics, molecular dynamics, X-ray crystallography, and nuclear magnetic resonance spectroscopy [92]. Another factor may have been that medicinal chemists and microbiologists at other pharmaceutical companies capitalized on the discovery of norfloxacin to elaborate even better quinoline antibacterials that eventually dominated the market.

As computers and software improved, SBDD became a more popular approach to drug discovery. One company, Agouron in San Diego, California, set a new paradigm for discovery based on iterations of crystallography and medicinal chemistry. As new compounds were made, some could be cocrystallized with the target protein. The 3D structure of the complex was solved by rapid computer techniques. Observations of how the compounds fit into the receptor suggested ways to improve affinity, leading to another round of synthesis and crystallography. Although considered by its practitioners and most others as an experimental science, protein crystallography (now popularly called structural biology, see also Chapter 12) often employed a step whereby the refraction data were refined in conjunction with constrained molecular dynamics (MD) simulations. Dr. Axel Brünger's program X-PLOR [93] met this important need. The force field in the program had its origin in CHARMM developed by Prof. Martin Karplus's group at Harvard [94]. Pharmaceutical companies that set up protein crystallography groups acquired X-PLOR to run on their computers.

The SBDD approach affected computational chemists positively. The increased number of 3D structures of therapeutically relevant targets opened new opportunities for molecular modeling of the receptor sites. Computational chemists assisted the medicinal chemists in interpreting the fruits of crystallography for design of new ligands.

Molecular dynamics simulations can consume prodigious amounts of computer time. Not only are proteins very large structures, but also the MD results are regarded as better the longer they are run because more of conformational space is assumed to be sampled by the jiggling molecules. Even more demand for computer power appeared necessary when free energy perturbation (FEP) theory appeared on the scene. Some of the brightest luminaries in academic computational chemistry proclaimed that here was a powerful new method for designing drugs [95, 96]. Pharmaceutical companies were influenced by these claims [97]. On the other hand, computational chemists closer to the frontline of working with medicinal chemists generally recognized that whereas FEP was a powerful method for accurately calculating the binding energy between ligands and macromolecular targets, it was too slow for extensive use in actual drug discovery. The molecular modifications that could be simulated with FEP treatment, such as changing one substituent

to another, were relatively minor. Because the FEP simulations had to be run so long to obtain good results, it was often possible for a medicinal chemist to synthesize the new modification in less time than it took to do the calculations. Also, for cases in which a synthesis would take longer than the calculations, not many industrial medicinal chemists would rate the modification worth the effort. Researchers in industry are under a great deal of pressure to tackle problems quickly and not spend too much time on them.

The insatiable need for more computing resources in the 1980s sensitized the pharmaceutical companies to the technological advances leading to the manufacture of supercomputers [98]. Some pharmaceutical companies opted for specialized machines such as array processors. By the mid-1980s, for example, several pharmaceutical companies had acquired the Floating Point System (FPS) 164. Other pharmaceutical companies sought to meet their needs by buying time and/or partnerships with one of the state or national supercomputing centers formed in the United States, Europe, and Japan. For instance, in 1988 Lilly partnered with the National Center for Supercomputing Applications (NCSA) in Urbana-Champaign, Illinois. Meanwhile, supercomputer manufacturers such as Cray Research and ETA Systems, both in Minnesota, courted scientists and managers at the pharmaceutical companies.

A phrase occasionally heard in this period was that computations were the “third way” of science. The other two traditional ways to advance science were experiment and theory. The concept behind the new phrase was that computing could be used to develop and test theories and to stimulate ideas for new experiments.

1.6 FRUITION: THE 1990s

The 1990s was a decade of fruition because the computer-based drug discovery work of the 1980s yielded an impressive number of new chemical entities reaching the pharmaceutical marketplace. We elaborate on this statement later in this section, but first we complete the story about supercomputers in the pharmaceutical industry.

Pharmaceutical companies were accustomed to supporting their own research and making large investments in it. In fact, the pharmaceutical industry has long maintained the largest self-supporting research enterprise in the world. However, the price tag on a supercomputer was daunting. To help open the pharmaceutical industry as a market for supercomputers, the chief executive officer (CEO) of Cray Research took the bold step of paying a visit to the CEO of Lilly in Indianapolis. Apparently, Cray’s strategy was to entice a major pharmaceutical company to purchase a supercomputer, and then additional pharmaceutical companies might follow suit in an attempt to keep their research competitive. Lilly was offered a Cray-2 at an irresistible price. Not only did Lilly buy a machine, but other pharmaceutical companies

either bought or leased a Cray. Merck, Bristol-Myers Squibb, Marion Merrell Dow (then a large company in Cincinnati, Ohio), Johnson & Johnson, and Bayer were among the companies that chose a Cray. Some of these machines were the older X-MP or the smaller J90 machine, the latter being less expensive to maintain.

After Lilly's purchase of the Cray 2S-2/128, line managers were given the responsibility to make sure the purchase decision had a favorable outcome. This was a welcome opportunity because line management was fully confident that supercomputing would revolutionize research and development [99]. The managers believed that a supercomputer would enable scientists to test more ideas than would be practical with older computers. Management was optimistic that a supercomputer would foster collaborations and information sharing among employees in different disciplines at the company. The managers hoped that both scientific and business uses of the machine would materialize. Ultimately, then, supercomputing would speed the identification of promising new drug candidates. Scientists closer to the task of using the supercomputer saw the machine primarily as a tool for performing longer molecular dynamics simulations and quantum mechanical calculations on large molecules. However, if some other computational technique such as QSAR or data mining was more effective at discovering and optimizing new lead compounds, then the supercomputer might not fulfill the dreams envisioned for it. A VAX cluster remained an essential part of the technology infrastructure best suited for management of the corporate library of compounds (see more about this below).

Lilly organized special workshops to train potential users of the Cray. This pool of potential users included as many willing medicinal chemists and other personnel as possible. In-house computational chemists and other experts were assigned the responsibility of conducting the off-site, week-long workshops. The workshops covered not only how to submit and retrieve jobs but also the general methods of molecular modeling, molecular dynamics, quantum chemistry, and QSAR. The latter, as mentioned, did not require supercomputing resources, except perhaps occasionally to generate quantum mechanical descriptors. Mainly, however, the training had the concomitant benefit of exposing more medicinal chemists, including younger ones, to what could be achieved with the current state of the art of computational chemistry applied to molecular design.

As the role of the computational chemist became more important, attitudes toward them became more accepting. At some large, old pharmaceutical houses, and at many smaller, newer companies, it was normal practice to allow computational chemists to be co-inventors on patents if the computational chemists contributed to a discovery. Other companies, including Lilly, had long had a company-wide policy that computational chemists could not be on patents. The policy was changed at Lilly as the 1990s dawned. Computational chemists were becoming nearly equal partners in the quest to discover drugs.

Lilly's Cray also served as an impressive public relations showcase. The machine was housed in a special, climate-controlled room. One side of the darkened room had a wall of large glass windows treated with polymer-dispersed liquid crystals. The thousands of visitors who came to Lilly each year were escorted into a uniquely designed observation room where an excellent video was shown that explained the supercomputer and how it could be used for drug discovery. The observation room was automatically darkened at the start of the video. At the dramatic finish of the video, the translucent glass wall was turned clear and bright lights were turned on inside the computer room, revealing the Cray-2 and its cooling tower for the heat transfer liquid. The visitors enjoyed the spectacle.

To the disappointment of Lilly's guest relations department, Lilly's Cray-2 was later replaced with a Cray J90, a mundane-looking machine. But the J90 was more economical, especially because it was leased. The supercomputers were almost always busy with molecular dynamics and quantum mechanical calculations [100]. Of the personnel at the company, the computational chemists were the main beneficiaries of supercomputing.

At the same time supercomputers that were creating excitement at a small number of pharmaceutical companies, another hardware development was attracting attention at just about every company interested in designing drugs. Workstations from Silicon Graphics Inc. (SGI) were becoming increasingly popular for molecular research. These high-performance, UNIX-based machines were attractive because of their ability to handle large calculations quickly and because of their high-resolution, interactive computer graphics. Although a supercomputer was fine for CPU-intensive jobs, the workstations were better suited for interactive molecular modeling software being used for drug research. The workstations became so popular that some medicinal chemists wanted them for their offices, not so much for extensive use but rather as a status symbol.

Another pivotal event affecting the hardware situation of the early 1990s merits mention. As already stated, the Apple Macintoshes were well liked by scientists. However, in 1994 Apple lost its lawsuit against Microsoft regarding the similarities of the Windows graphical user interface (GUI) to Apple's desktop design. Also, the price of Windows-based PCs dropped significantly below that of Macs. The tables tilted in favor of PCs. More scientists began to use PCs. At Lilly, and maybe other companies, the chief information officer (a position that did not even exist until computer technology became so critical to corporate success) decreed that the company scientists would have to switch to PCs whether they wanted to or not. The reasons for this were severalfold. The PCs were more economical. With PCs being so cheap, it was likely more people would use them, and hence there was a worry that software for Macs would become less plentiful. Also, the problem of incompatible files would be eliminated if all employees used the same type of computer and software.

On the software front, the early 1990s witnessed a continued trend toward commercially produced programs being used in pharmaceutical companies.

Programs such as SYBYL (Tripos), Insight/Discover (BIOSYM), and Quanta/CHARMm (Polygen, and later Molecular Simulations Inc., and now Accelrys) were popular around the world for molecular modeling and simulations. Some pharmaceutical companies bought licenses to all three of these well-known packages. Use of commercial software freed the in-house computational chemists from the laborious task of code development, documentation, and maintenance, so that they would have more time to work on actual drug design. Another advantage of using commercial software was that the larger vendors would have a help desk that users could telephone for assistance when software problems arose, as they often did. The availability of the help desk meant that the in-house computational chemists would have fewer interruptions from medicinal chemists who were having difficulty getting the software to work. On the other hand, some companies, particularly Merck and Upjohn, preferred to develop software in-house because it was thought to be better than what the vendors could provide.

Increasing use of commercial software for computational chemistry meant a declining role for software from QCPE. QCPE had passed its zenith by about 1992, when it had almost 1900 members and over 600 programs in its catalog. This catalog included about 15 molecular modeling programs written at pharmaceutical companies and contributed for the good of the community of computational chemists. Among the companies contributing software were Merck, DuPont, Lilly, Abbott, and Novartis. When distribution rights for MOPAC were acquired by Fujitsu in 1992, it was a severe blow to QCPE. After a period of decline, the operations of QCPE changed in 1998. Today only a web-based operation continues at Indiana University, Bloomington.

The 1990s was a decade of change for the software vendors also. The California company that started out as BioDesign became Molecular Simulations Inc. (MSI). MSI went on a buying spree starting in 1991. It grew large by acquiring other small software companies competing in the same drug design market, including Polygen, BIOSYM, BioCAD, Oxford Molecular (which had already acquired several other start-ups), and others [101]. Pharmaceutical companies worried about this accretion because it could mean less competition and it could mean that their favorite molecular dynamics (MD) program might no longer be supported in the future. This latter possibility has not come to pass because there has been sufficient loyalty and demand for each MD package to remain on the market.

Researchers from pharmaceutical companies participated in user groups set up by the software vendors. Pharmaceutical companies also bought into consortia created by the software vendors. These consortia, some of which dated back to the 1980s, aimed at developing new software tools or improving existing software. The pharmaceutical companies hoped to get something for their investments. Sometimes the net effect of these investments was that it enabled the software vendors to hire several postdoctoral research associates who worked on things that were of common interest to the investors. Although the pharmaceutical companies received some benefit from the consortia, other

needs such as more and better force field parameters remained underserved. Inspired by the slow progress in one force field development consortium, Merck single-handedly undertook the de novo development of a force field they call the Merck Molecular Force Field (MMFF94). This force field, which targeted the modeling of pharmaceutically interesting molecules well, was published [102–108], and several software vendors subsequently incorporated it in their molecular modeling programs. The accolades of fellow computational chemists led to the developer being elected in 1992 to become chairman of one of the Gordon Research Conferences on Computational Chemistry [109].

On the subject of molecular modeling and force fields, a general molecular modeling package was developed in an organic chemistry laboratory at Columbia University in New York City [110]. Perhaps because MacroModel was written with organic chemists in mind, it proved popular with industrial medicinal chemists, among others. The program was designed so that versions of several good force fields could easily be invoked for any energy minimization or molecular simulation.

The 1990s witnessed other exciting technological developments. In 1991, Dr. Jan K. Labanowski, then an employee of the Ohio Supercomputer Center (Columbus, Ohio), launched an electronic bulletin board called the Computational Chemistry List (CCL). Computational chemists rapidly joined because it was an effective forum for informal exchange of information. Computational chemists at pharmaceutical companies were among the 2000 or so members who joined in the 1990s. Often these employees would take the time to answer questions from beginners, helping them learn about the field of computer-aided drug design. The CCL was a place where the relative merits of different methodologies and computers and the pros and cons of various programming languages could be debated, sometimes passionately.

In 1991, MDL came out with a new embodiment of its compound management software called ISIS (Integrated Scientific Information System). Pharmaceutical companies upgraded to the new system, having become so dependent on MDL. In general, managers of information technology at pharmaceutical companies preferred one-stop solutions. On the other hand, computational chemists found Daylight Chemical Information Systems software more useful for developing new research applications.

MACCS and then ISIS gave researchers exceptional new tools for drug discovery when similarity searching came along. Chemical structures were stored in the database as connectivity tables (showing which atoms were connected by bonds). In addition, chemical structures could be stored as a series of on-off flags (“keys”) indicating the presence or absence of specific atoms or combinations of atoms and/or bonds. The similarity of compounds could be quantitated by the computer in terms of the percentage of keys that the compounds shared in common. Thus, if a researcher was aware of a lead structure from in-house work or the literature, it was possible to find compounds in the corporate database that were similar and then get these compounds assayed for biological activities. Therefore the technique of data

mining became important. Depending on how large the database was, it was fairly easy to find compounds with low levels of activity by this method. Some of these active compounds might have a skeleton different from the lead structure. The new skeleton could form the basis for subsequent lead optimization. As Dr. Yvonne C. Martin (Abbott) has wryly commented in her lectures at scientific meetings, one approach to drug discovery is to find a compound that the target receptor sees as the same as an established ligand but which a patent examiner sees as a different compound (and therefore satisfying the novelty requirement for patentability).

Many or most of the results from data mining in industry went unpublished. More recently, when a few academic researchers gained access to data mining software, the weakly active compounds they found were excitedly published. This difference between industry and academia in handling similar kinds of results is a matter of priorities. In industry, the first priority is to find marketable products and get them out the door. In academia, the priority is to publish (especially in high-impact journals). Contrary to a common misconception, scientists in industry do publish, a point we return to below.

Software use for drug discovery and development can be classified in various ways. One way is technique based. Examples would be programs based on force fields or on statistical fitting (the latter including log *P* prediction or toxicity prediction). Another way to classify software is according to whether the algorithm can be applied to cases in which the 3D structure of the target receptor is known or not. An example of software useful when the receptor structure is not known is Catalyst [111]. This program, which became available in the early 1990s, tried to produce a 3D model of a pharmacophore based on a small set of compounds with a range of activities against a given target. The pharmacophore model, if determinable, could be used as a query to search databases of 3D structures in an effort to find new potential ligands.

In situations in which the computational chemist had the benefit of the 3D structure of the target receptor, three methodologies came into increased usage. One was docking, that is, letting an algorithm try to fit a ligand structure into a receptor. Docking methodology dates back to the 1980s, but the 1990s saw more crystal structures of pharmaceutically relevant proteins being solved and used for ligand design [112]. A second technique of the 1990s involved designing a computer algorithm to construct a ligand *de novo* inside a receptor structure. The program would assemble fragments or “grow” a chemical structure such that the electrostatic and steric attributes of the receptor would be complemented by the ligand [113–115]. The third technique of the 1990s was virtual screening [116, 117]. The computer would screen hypothetical ligand structures, not necessarily compounds in bottles, against the 3D structure of a receptor in order to find those most likely to fit and therefore worthy of synthesis and experimentation.

A new approach to drug discovery came to prominence around 1993. The arrival of this approach was heralded with optimism reminiscent of earlier

waves of new technologies. The proponents of this innovation—combinatorial chemistry—were organic chemists. The thinking behind combinatorial chemistry seemed to be as follows. The chance of finding a molecule with therapeutic value was extremely low (one in 5000 or one in 10,000 were rough estimates that were often bandied about). Attempts at rational drug design had not significantly improved the odds of finding those rare molecules that could become a pharmaceutical product. Because the low odds could not be beat, make tens of thousands, . . . no, hundreds of thousands, . . . no, millions of compounds! Then, figuratively fire a massive number of these molecular bullets at biological targets and hope that some might stick. New computer-controlled robotic machinery would permit syntheses of all these compounds much more economically than the traditional one-compound-at-a-time process of medicinal chemistry. Likewise, computer-controlled robotic machinery would automate the biological screening and reduce the cost per assay.

Proponents promised that combinatorial chemistry was the way to keep the drug discovery pipeline full. Pharmaceutical companies made massive investments in people and machinery to set up the necessary equipment in the 1990s. Some companies built large refrigerated storage rooms where the libraries of compounds could be stored and retrieved by robots. The computers to run the equipment had to be programmed. This work was done by instrument engineers, although chemists helped set up the systems that controlled the syntheses.

Combinatorial chemistry increased the rate of output of new compounds by three orders of magnitude. Before combi-chem, a typical SAR at a pharmaceutical company might have consisted of fewer than a couple hundred compounds, and a massive effort involving 10–20 medicinal chemistry laboratories might have produced two or three thousand compounds over a number of years. In 1993, with traditional one-compound-at-a-time chemistry it took one organic chemist on average one week to make one compound for biological testing. Some years later, with combi-chem a chemist could easily produce 2000 compounds per week.

With the arrival of combi-chem, computational chemists had a new task in addition to what they had been doing. Computational chemistry was needed so that the combinatorial chemistry was not mindlessly driven by whatever reagents were available in chemical catalogs or from other sources. There were several strategies to library design [118]. The first was to cover as much of “compound space” as possible, that is, to produce a variety of structures to increase the likelihood that one of the compounds would stick to the target. Then after the drug discovery researchers had gained a general idea of what structure(s) would bind to the target receptor, a second strategy would come into play: to design compounds similar to the lead(s). Another need was to assess the value of libraries being offered for sale by various outside intermediaries. Computational chemists could help determine whether these libraries complemented or duplicated a company’s existing libraries of compounds and

determine the degree of variety in the compounds being offered. How does one describe chemical space and molecular similarity? Computational chemists had already developed the technologies of molecular descriptors and substructure keys, which we mentioned above. With these tools, the computational chemist could discern where structures were situated in multidimensional compound or property space and provide advice to the medicinal chemists.

Along with all the data generated by combi-chem and high-throughput screening (HTS) came the need to manage and analyze the data. Hence, computers and the science of informatics became increasingly vital.

The computational chemist was now more important to drug discovery research than ever before. Hence by 1993–1994, these technological changes possibly helped save the jobs of many computational chemists at a time when pharmaceutical companies in the United States were downsizing, as we now explain. In 1992–1993 an acute political force impinged on the pharmaceutical industry in the United States. That force was the healthcare reform plan proposed by Hillary and Bill Clinton. Readers who are well versed in history of the 1930s will be aware of the economic system handed down from pre-World War II Europe. Under that system, the means of production, that is, industry, remains in private ownership but the prices that the companies can ask for their products are regulated by government. That was the scheme underlying the healthcare reform proposal. Pharmaceutical companies in the United States generally favored any proposal that would increase access to their products but feared this specific proposal because of the great uncertainty it cast over the status quo and future growth prospects. As a result, thousands of pharmaceutical workers—including research scientists—were laid off or encouraged to retire. Rumors swirled around inside each pharmaceutical company about who would be let go and who would retain their jobs. When word came down about the corporate decisions, the computational chemists were generally retained, but the ranks of the older medicinal chemists were thinned. A new generation of managers at pharmaceutical companies now realized that computer-assisted molecular design and library design were critical components of their company's success. One is reminded of the observation of the Nobel laureate physicist Max Planck, "An important scientific innovation rarely makes its way by gradually winning over and converting its opponents. . . . What does happen is that its opponents gradually die out and the growing generation is familiarized with the idea from the beginning."

Nevertheless, the Clintons' healthcare reform scheme had a deleterious effect on the hiring of new computational chemists. The job market for computational chemists in the United States fell from a then record high in 1990 to a depression in 1992–1994 [119]. This happened because pharmaceutical companies were afraid to expand until they were sure that the business climate was once again hospitable for growth. The healthcare reform proposal was defeated in the United States Congress, but it took a year or two before pharmaceutical companies started rebuilding their workforces.

Toward the mid-1990s, a new mode of delivering content came to the fore: the web browser. Information technology (IT) engineers and computational chemists help set up intranets at pharmaceutical companies. This allowed easy distribution of management memos and other information to the employees. In addition, biological screening data could be posted on the intranet so that medicinal chemists could quickly access it electronically. Computational chemists made their applications (programs) web-enabled so that medicinal chemists and others could perform calculations from their desktops.

The hardware situation continued to evolve. Personal computers became ever more powerful in terms of speed and the amount of random access memory (RAM) and hard drive capacity. The price of PCs continued to fall. Clusters of PCs were built. Use of the open-source Linux operating system spread in the 1990s. Distributed processing was developed so a long calculation could be farmed out to separate machines. Massively parallel processing was tried. All these changes meant that the days of the supercomputers were numbered.

Whereas the trend in the 1980s was toward dispersal of computing power to the departments and the individual user, the IT administrators started bringing the PCs under their centralized control in the 1990s. Software to monitor each machine was installed so that what each user did could be tracked. Following the example of other industries, some pharmaceutical companies turned over the technical work of managing their networks of PCs to outside contractors. Gradually, computational chemists and other workers lost control over what could and could not be installed on their office machines. One type of hardware, however, persisted through the 1990s and even to today: the SGI workstations. These UNIX machines became more powerful and remained popular for molecular modeling. Silicon Graphics Inc. acquired the expiring Cray technology, but it did not seem to have much effect on their workstation business.

Traditionally, in pursuit of their structure-activity relationships, medicinal chemists had focused almost exclusively on finding compounds with greater and greater potency. However, these SARs often ended up with compounds that were unsuitable for development as pharmaceutical products. These compounds would be too insoluble in water, or were not orally bioavailable, or were eliminated too quickly or too slowly from mammalian bodies. Pharmacologists and pharmaceutical development scientists for years had tried to preach the need for medicinal chemists to also think about other factors that determined whether a compound could be a medicine. Table 1.1 lists a number of factors that determine whether a potent compound has what it takes to become a drug. Experimentally, it was difficult to quantitate these other factors. Often, the necessary manpower resources would not be allocated to a compound until it had already been selected for project team status.

At the beginning of the 1990s, the factors in Table 1.1 were generally beyond the capability of computational chemistry methods to predict reliably. However, as the decade unfolded, computational chemists and other scientists

TABLE 1.1 What It Takes for a Compound (Ligand) to Become a Pharmaceutical Product

Absorption into the body, i.e., bioavailability
Behavior in humans as anticipated from preliminary testing in animal models, i.e., no untoward species differences
Distribution among the appropriate tissues of the body
Metabolism by the body or organisms living in the body
Ease of production, including, for instance, the existence of environmentally safe routes of isolation or synthesis
Efficacy, i.e., whatever the compound does at its site(s) of action; the net effect is to elicit a desirable therapeutic outcome
Elimination from the body, i.e., excretion
Medical need, which affects marketability
Novelty, which determines patentability
Pharmaceutical “elegance,” which encompasses factors related to route of administration (taste, color, mixability with excipients, etc.)
Side effects of the compound and its degradation products are minimal or at least tolerable
Solubility, preferably in water
Stability, so the compound will not degrade before being consumed and can reach its site of action in a bioactive form
Therapeutic ratio, so that the concentration of the compound to elicit its therapeutic effect is much lower than the concentration that would cause untoward effects
Toxic effects of the compound and its degradation products are minimal

created new and better methodologies for selecting compounds with the characteristics necessary to become a drug. In 1997, Lipinski’s now famous “Rule of Five” was published [120]. These simple rules were easily encoded in database mining operations at every company, so that compounds with low prospects of becoming an orally active, small-molecule drug (less than 500 MW) could be weeded out by computer.

The computational methods used in the 1980s focused, like medicinal chemistry, on finding compounds with ever-higher affinity between the ligand and its target receptor. That is why in the past we have advocated use of the term computer-aided ligand design (CALD) rather than CADD. However, with increased attention to the factors listed in Table 1.1, the field was finally becoming more literally correct in calling itself CADD.

Another important change started in the mid-1990s. Traditionally, a QSAR determined at a pharmaceutical company might have involved only 5–30 compounds. The size depended on how many compounds the medicinal chemist had synthesized and submitted to testing by the biologists. Sometimes this size data set sufficed to reveal useful trends. Other times, though, the

QSARs were not very robust in terms of predictability. As large libraries of compounds were produced, data sets available for QSAR analysis became larger. With all that consistently produced (although not necessarily perfectly accurate) biological data and a plethora of molecular descriptors, it was possible to find correlations with good predictability. In fact, QSAR proved to be one of the best approaches to providing assistance to the medicinal chemist in the 1990s. Computational chemists were inventive in creating new molecular descriptors. Hundreds have been described in the literature [121–123].

As stated in the opening of this section, the 1990s witnessed the fruition of a number of drug design efforts. Making a new pharmaceutical product available to patients is a long, difficult, and costly enterprise. It takes 10–15 years from the time a compound is discovered in the laboratory until it is approved for physicians to prescribe. Hence, a molecule that reached the pharmacies in the 1990s was probably first synthesized at a pharmaceutical company well back in the 1980s. (Most of today's medicines come from the pharmaceutical industry rather than from government or academic laboratories.) The improved methodologies of computational chemistry that became available in the 1980s would therefore start to show their full impact in the 1990s. (Likewise, the improved experimental and computational methodologies of the 1990s should be bearing fruit now.)

Table 1.2 lists medicines whose discovery was aided in some way by computer-based methods. Those compounds marked “CADD” were pub-

TABLE 1.2 Marketed Pharmaceuticals Whose Discovery Was Aided by Computers

Generic Name	Brand Name	Year Approved in United States	Discovery Assisted by	Activity
Norfloxacin	Noroxin	1983	QSAR	Antibacterial
Losartan	Cozaar	1994	CADD	Antihypertensive
Dorzolamide	Trusopt	1995	CADD/SBDD	Antiglaucoma
Ritonavir	Norvir	1996	CADD	Antiviral
Indinavir	Crixivan	1996	CADD	Antiviral
Donepezil	Aricept	1997	QSAR	Anti-Alzheimer's
Zolmitriptan	Zomig	1997	CADD	Antimigraine
Nelfinavir	Viracept	1997	SBDD	Antiviral
Amprenavir	Agenerase	1999	SBDD	Antiviral
Zanamivir	Relenza	1999	SBDD	Antiviral
Oseltamivir	Tamiflu	1999	SBDD	Antiviral
Lopinavir	Aluviran	2000	SBDD	Antiviral
Imatinib	Gleevec	2001	SBDD	Antineoplastic
Erlotinib	Tarceva	2004	SBDD	Antineoplastic
Ximelagatran	Exanta	2004	SBDD	Anticoagulant

licized in a series of earlier publications [references 124–128; see also references 129 and 130, and Chapter 16 for other discussion]. These examples of CADD successes were gathered in 1997 when one of us (DBB) undertook a survey of the corresponding authors of papers published after 1993 in the prestigious *Journal of Medicinal Chemistry*. Authors were asked whether calculations were crucial to the discovery of any compounds from their laboratory. Of the hundreds of replies, we culled out all cases in which calculations had not led to drug discovery or had been done post hoc on a clinical candidate or pharmaceutical product. We have always felt strongly that the term “computer-aided drug design” should be more than just doing a calculation; it should be providing information or ideas that directly help with the conception of a useful new structure. We retained only those cases where the senior author of a paper (usually a medicinal chemist) vouched that computational chemistry had actually been critically important in the research process that led to the discovery of a compound that had reached the market. As seen in Table 1.2, there were seven compounds meeting this criterion in the period 1994–1997. The computational techniques used to find these seven compounds included QSAR, ab initio molecular orbital calculations, molecular modeling, molecular shape analysis [131], docking, active analog approach [132], molecular mechanics, and SBDD.

More recently, a group in England led by a structural biologist compiled a list of marketed medicines that came from SBDD [133]. These are labeled “SBDD” in Table 1.2. It can be seen that there is only a little overlap between the two compilations (CADD and SBDD). It can also be seen that the number of pharmaceuticals from SBDD is very impressive. Computer-based technologies are clearly making a difference in helping bring new medicines to patients.

Looking at the success stories, we see that it has often been a team of researchers working closely together that led to the success. It took quite a while for other members of the drug discovery research community to appreciate what computational chemistry could provide. There remains room for further improvement in this regard. Computational chemistry is probably most effective when researchers work in an environment where credit is shared [134]. If management adopts a system whereby company scientists are competing with each other, then collaborations are tempered. On the other hand, if all members of an interdisciplinary team of scientists will benefit when the team succeeds, then collaboration increases, synergies can occur, and the team is more likely to succeed. Sometimes it helps to put the computational chemistry techniques in the hands of the medicinal chemists, but it seems that only some of these chemists have the time and inclination to use the techniques to best advantage. Therefore, computational chemistry experts play an important role in maximizing the potential benefits of computer-based technologies.

1.7 EPILOGUE

To close, we distill in Figure 1.3 the essence of what we have described about the history of computing at pharmaceutical companies over the last four decades. We plot the number of papers published (and abstracted by Chemical Abstracts Service) for each year from 1964 through 2004, the most recent

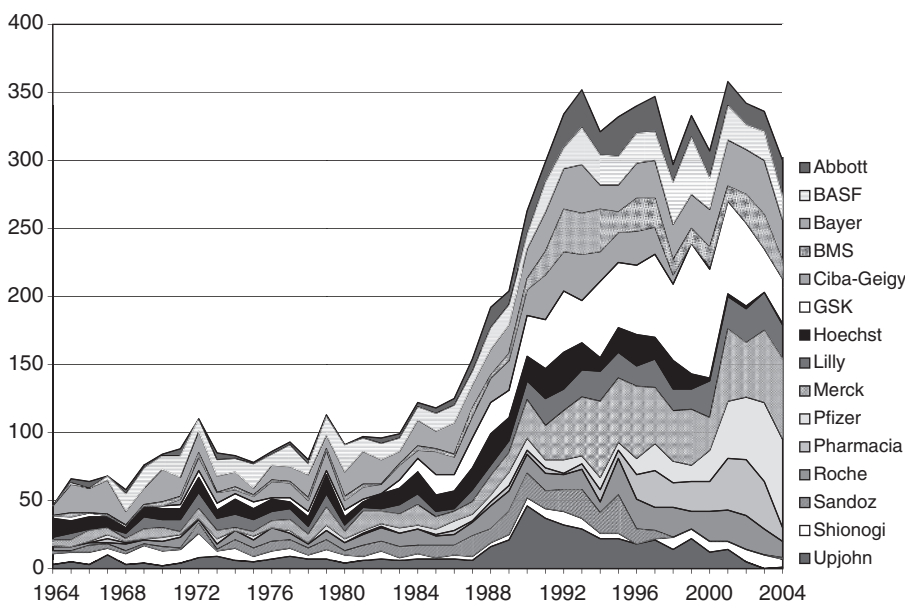


Figure 1.3 Annual number of papers published by researchers at pharmaceutical companies during a 41-year period. The data were obtained by searching the CAPLUS and MEDLINE databases for papers related to “computer or calculation.” Then these hits were refined with SciFinder Scholar by searching separately for 48 different company names. Well-known companies from around the world were included. Companies with more than 250 total hits in the period 1964–2004 are included in the plot. The indexing by CAS is such that a search on SmithKline Beecham gave the same number of hits as for GlaxoSmithKline (GSK) but much more than Smith Kline and French. The downward trend of the sum after 2001 can be traced to fewer papers coming from GSK. Initially, we had wanted to structure our SciFinder Scholar search for all papers using terms pertaining to computational chemistry, molecular modeling, computer-aided drug design, quantitative structure-activity relationships, and so forth. However, CAS classifies these terms as overlapping concepts, and so SciFinder Scholar was unable to do the searches as desired. Searching on “computer or calculation” yields many relevant hits but also a small number of papers that are of questionable relevance. This contamination stems from the subjective way abstractors at CAS have indexed articles over the years. The irrelevant papers introduce noise in the data, but hopefully the systematic error is relatively constant over the period covered. CAPLUS covers 1500 journals. See color plate.

complete year. These are papers that were indexed as pertaining to “computer or calculation” and that came from pharmaceutical companies. We can learn several things from Figure 1.3. First, industrial scientists do publish. Second, the figure includes a list of pharmaceutical companies that have done the most publishing of papers pertaining to computers or calculations. There are 15 companies with more than 250 papers each in the 41-year period. Some of these companies ceased publishing simply because they were acquired by other pharmaceutical companies, and hence the affiliation of the authors changed. The companies are headquartered in the United States, Switzerland, Germany, and Japan. Third, the number of publications increased slowly but fairly steadily from 1964 through the mid-1980s. Then, from 1986 through 1992, the annual number of papers grew rapidly. This period is when the super-minicomputers, supercomputers, and workstations appeared on the scene. We also learn from Figure 1.3 that since 1994 the sum of the annual number of papers published by the 15 companies has zigzagged around 325 papers per year. Curiously, the recent years show no evidence of an upward trend.

As the twentieth century came to a close, the job market for computational chemists had recovered from the 1992–1994 debacle. In fact, demand for computational chemists leaped to new highs each year in the second half of the 1990s [135]. Most of the new jobs were in industry, and most of these industrial jobs were at pharmaceutical or biopharmaceutical companies. As we noted at the beginning of this chapter, in 1960 there were essentially no computational chemists in industry. But 40 years later, perhaps well over half of all computational chemists were working in pharmaceutical laboratories. The outlook for computational chemistry is therefore very much linked to the health of the pharmaceutical industry itself. Forces that adversely affect pharmaceutical companies will have a negative effect on the scientists who work there as well as at auxiliary companies such as software vendors that develop programs and databases for use in drug discovery and development.

Over the last four decades, we have witnessed waves of new technologies sweep over the pharmaceutical industry. Sometimes these technologies tended to be oversold at the beginning and turned out to not be a panacea to meet the quota of the number of new chemical entities that each company would like to launch each year. Experience has shown that computer technology so pervasive at one point in time can almost disappear 10 years later.

Discovering new medicines is a serious, extremely difficult, and expensive undertaking. Tens of thousands of scientists are employed in this activity. Back in 1980, pharmaceutical and biotechnology companies operating in the United States invested in aggregate about \$2000 million in R&D. The sum has steadily increased (although there was the slight pause in 1994 that we mentioned above). By 2003, R&D investments had grown to \$34,500 million. In 2004, the total jumped to \$38,800 million. The United States pharmaceutical industry invests far more in discovering new and better therapies than the pharmaceutical industry in any other country or any government in the world. Despite the ever-increasing investment in R&D each year, the annual number

of new chemical entities (NCEs) approved for marketing in the United States (or elsewhere) has not shown any overall increase in the last 25 years. The number has fluctuated between 60 and 20 NCEs per year and has been around 30 per year recently. This very uncomfortable fact was not widely discussed before the late 1990s [124] but is now well known. A recent analysis of NCE data was able to find some reason for optimism that innovation is bringing to market drugs with substantial advantage over existing treatments [136]. However, deciding whether R&D is becoming more productive depends on how the NCE data are handled. Generally, most people in the field realize that discovery research is not as easy or as productive as they would like.

In an attempt to boost NCE output, executives at pharmaceutical companies have put their researchers under extreme pressure to focus and produce. Since the early 1990s, this pressure has moved in only one direction: up.

During two million years of human evolution, better intelligence at creating and using tools has meant the difference between survival and extinction. In a similar way, those pharmaceutical companies with scientists who are best at creating and using tools will be able to innovate their way to the future. In contrast to the days of the hunter-gatherer cracking stones, today the tools are computers and software, as well as chemistry and biology. With combinatorial chemistry, high-throughput screening, genomics, and structural biology firmly embedded in modern drug discovery efforts, computers are indispensable.

All musical composers work with the same set of notes, but the geniuses put the notes together in an extraordinarily beautiful way. Synthetic chemists all have available to them the same elements. The successful medicinal chemist will combine atoms such that amazing therapeutic effect is achieved with the resulting molecule. The computational chemist's goal should be to help the medicinal chemist by providing information about structural and electronic requirements to enhance activity, namely, information about which regions of compound space are most propitious for exploration.

Fortunately, all the effort that goes into pharmaceutical R&D does benefit society. In nations where modern medicines are available, life expectancy has increased and disability rates among the elderly have declined. Considering all of the things that can go wrong with the human body, many challenges remain for the pharmaceutical researcher. Hopefully, this chapter will inspire some young readers to take up the challenge and join the noble quest to apply science to help find cures to improve people's lives.

ACKNOWLEDGMENTS

We are grateful to Dr. Sean Ekins for the opportunity to participate in this book and for his painstaking editing. We thank our many colleagues over the years for what they taught us or tried to teach us. Some of the historical events reviewed here occurred during the period of our respective tenures as members of the staff at the Lilly Research Laboratories of Eli Lilly and Company. We

thank Prof. Norman L. Allinger, Mr. Douglas M. Boyd, Dr. David K. Clawson, Dr. Richard D. Cramer III, Dr. David A. Demeter, Mr. Gregory L. Durst, Dr. Richard W. Harper, Dr. Robert B. Hermann, Dr. Stephen W. Kaldor, Dr. Yvonne C. Martin, Dr. Samuel A. F. Milosevich, and Dr. Terry R. Stouch for aiding us as we wrote this review. Creation of this review was also assisted by the computer resources of SciFinder Scholar, Google, and Wikipedia. No slight is intended for any scientist not mentioned or any publication not cited in this brief history. Our small recitation here should not add to or subtract from the already established achievements of our fellow scientists living or dead.

REFERENCES

1. Fischer E. Einfluß der Konfiguration auf die Wirkung der Enzymen. *Ber Dtsch Chem Ges* 1894;27:2985–93.
2. Silverman RB. *The organic chemistry of drug design and drug action*. San Diego: Academic Press, 1992.
3. Messiah A. *Quantum mechanics*, Vol. I. (Translated from the French by Temmer GM). New York: Wiley, 1966.
4. Pullman B, Pullman A. *Quantum biochemistry*. New York: Interscience Publishers, Wiley, 1963.
5. Crum Brown A, Frazer TR. On the connection between chemical constitution and physiological action. Part I. On the physiological action of the salts of the ammonium bases, derived from strychnia, brucia, thebaia, codeia, morphia, and nicoti. *Trans R Soc Edinburgh* 1869;25:151–203.
6. Crum Brown A, Frazer TR. On the connection between chemical constitution and physiological action. Part II. On the physiological action of the ammonium bases derived from atropia and conia. *Trans R Soc Edinburgh* 1869;25:693–739.
7. Bruice TC, Kharasch N, Winzler RJ. A correlation of thyroxine-like activity and chemical structure. *Arch Biochem Biophys* 1956;62:305–17.
8. Zahradnik R. Correlation of the biological activity of organic compounds by means of the linear free energy relations. *Experimentia* 1962;18:534–6.
9. Hansch C, Fujita T. ρ - σ - π Analysis; method for the correlation of biological activity and chemical structure. *J Am Chem Soc* 1964;86:1616–26.
10. Free SM Jr, Wilson JW. A mathematical contribution to structure-activity studies. *J Med Chem* 1964;7:395–9.
11. Bolcer JD, Hermann RB. The development of computational chemistry in the United States. In: Lipkowitz KB, Boyd DB, editors. *Reviews in computational chemistry*, Vol. 5. New York: VCH, 1994. p. 1–63.
12. Marsh MM. Spectropolarimetric studies on proteins. Bovine plasma albumin and insulin. *J Am Chem Soc* 1962;84:1896–900.
13. Moffitt W. Optical rotatory dispersion of helical polymers. *J Chem Phys* 1956;25:467–78.

14. Allinger N, Allinger J. *Structures of organic molecules*. Englewood Cliffs, NJ: Prentice-Hall, 1965.
15. Dayhoff MO. Computer search for active-site configurations. *J Am Chem Soc* 1964;86:2295–7.
16. Lipkowitz KB, Boyd DB, editors. Books Published on the Topics of Computational Chemistry. In: *Reviews in Computational Chemistry*, Vol. 17. New York: Wiley-VCH, 2001. p. 255–357.
17. Lipkowitz KB, Boyd DB, editors. A tribute to the halcyon days of QCPE. In: *Reviews in Computational Chemistry*, Vol. 15. New York: Wiley-VCH, 2000. p. v–xi.
18. Johnson CK. Drawing crystal structures by computer. *Crystallographic Computing*, Proceedings of the International Summer School held 1969, 1970. p. 227–30.
19. Sutton LE, Jenkin DG, Mitchell AD, Cross LC, editors. *Tables of interatomic distances and configuration in molecules and ions*, Special Publ No. 11. London: The Chemical Society, 1958.
20. Koltun WL. Precision space-filling atomic models. *Biopolymers* 1965;3:665–79.
21. Boyd DB. Space-filling molecular models of four-membered rings. Three-dimensional aspects in the design of penicillin and cephalosporin antibiotics. *J Chem Educ* 1976;53:483–8.
22. Hoffmann R, Lipscomb WN. Theory of polyhedral molecules. I. Physical factorizations of the secular equation. *J Chem Phys* 1962;36:2179–89.
23. Hoffmann, R. An extended Hückel theory. I. Hydrocarbons. *J Chem Phys* 1963;39:1397–412.
24. Pople JA, Segal GA. Approximate self-consistent molecular orbital theory. II. Calculations with complete neglect of differential overlap. *J Chem Phys* 1965;43: S136–49.
25. Pople JA, Beveridge DL. *Approximate molecular orbital theory*. New York: McGraw-Hill, 1970.
26. Hartree, DR. The wave mechanics of an atom with a non-coulomb central field. I. Theory and methods. *Proc Cambridge Phil Soc* 1928;24:89–110.
27. Hartree, DR. The wave mechanics of an atom with a non-coulomb central field. II. Some results and discussion. *Proc Cambridge Phil Soc* 1928;24:111–32.
28. Hartree, DR. Wave mechanics of an atom with a non-coulomb central field. III. Term values and intensities in series in optical spectra. *Proc Cambridge Phil Soc* 1928;24 (Pt. 3):426–37.
29. Fock, V. “Self-consistent field” with interchange for sodium. *Zeitschrift für Physik* 1930;62:795–805.
30. Roberts JD. *Notes on molecular orbital calculations*. New York: Benjamin, 1962.
31. Neely WB. The use of molecular orbital calculations as an aid to correlate the structure and activity of cholinesterase inhibitors. *Mol Pharmacol* 1965;1:137–44.
32. Schnaare RS, Martin AN. Quantum chemistry in drug design. *J Pharmaceut Sci* 1965;54:1707–13.

33. Parr RG. *Quantum theory of molecular electronic structure*. New York: Benjamin, 1963.
34. Hammett LP. *Physical organic chemistry; reaction rates, equilibria, and mechanisms*, 2nd edition. New York: McGraw-Hill, 1970.
35. Taft RW Jr. Linear free-energy relationships from rates of esterification and hydrolysis of aliphatic and ortho-substituted benzoate esters. *J Am Chem Soc* 1952;74:2729–32.
36. Gould ES. *Mechanism and structure in organic chemistry*. New York: Holt Reinhart Winston, 1959.
37. Hermann RB. Structure-activity correlations in the cephalosporin C series using extended Hückel theory and CNDO/2. *J Antibiot* 1973;26:223–7.
38. Boyd DB, Hermann RB, Presti DE, Marsh MM. Electronic structures of cephalosporins and penicillins. 4. Modeling acylation by the beta-lactam ring. *J Med Chem* 1975;18:408–17.
39. Boyd DB, Herron DK, Lunn WHW, Spitzer WA. Parabolic relationships between antibacterial activity of cephalosporins and beta-lactam reactivity predicted from molecular orbital calculations. *J Am Chem Soc* 1980;102:1812–14.
40. Boyd DB. Beta-lactam antibacterial agents: Computational chemistry investigations. In: Greenberg A, Breneman CM, Liebman JF, editors. *The amide linkage: Structural significance in chemistry, biochemistry, and materials science*. New York: Wiley, 2000. p. 337–75.
41. Corey EJ, Wipke WT, Cramer RD III, Howe WJ. Computer-assisted synthetic analysis. Facile man-machine communication of chemical structure by interactive computer graphics. *J Am Chem Soc* 1972;94:421–30.
42. Corey EJ, Wipke WT, Cramer RD III, Howe WJ. Techniques for perception by a computer of synthetically significant structural features in complex molecules. *J Am Chem Soc* 1972;94:431–9.
43. Wipke, WT, Gund P. Congestion. Conformation-dependent measure of steric environment. Derivation and application in stereoselective addition to unsaturated carbon. *J Am Chem Soc* 1974;96:299–301.
44. Bingham RC, Dewar MJS, Lo DH. Ground states of molecules. XXV. MINDO/3. Improved version of the MINDO semiempirical SCF-MO method. *J Am Chem Soc* 1975;97:1285–93.
45. Löwdin PO, editor. Proceedings of the international symposium on quantum biology and quantum pharmacology, Held at Sanibel Island, Florida, January 17–19, 1974. *Int J Quantum Chem, Quantum Biol Symp* No. 1. New York: Wiley, 1974.
46. Richards WG. *Quantum pharmacology*. London, UK: Butterworths, 1977.
47. Olson EC, Christoffersen RE, editors. *Computer-assisted drug design*, Based on a symposium sponsored by the Divisions of Computers in Chemistry and Medicinal Chemistry at the ACS/CSJ Chemical Congress, Honolulu, Hawaii, April 2–6, 1979. ACS Symposium Series 112. Washington, DC: American Chemical Society, 1979.
48. Allinger NL, Miller MA, Van Catledge FA, Hirsch JA. Conformational analysis. LVII. The calculation of the conformational structures of hydrocarbons

- by the Westheimer-Hendrickson-Wiberg method. *J Am Chem Soc* 1967;89: 4345–57.
49. Gygax R, Wirz J, Sprague JT, Allinger NL. Electronic structure and photophysical properties of planar conjugated hydrocarbons with a 4n-membered ring. Part III. Conjugative stabilization in an “antiaromatic” system: The conformational mobility of 1,5-bisdehydro[12]annulene. *Helv Chim Acta* 1977;60:2522–9.
50. Westheimer FH, Mayer JE. The theory of the racemization of optically active derivatives of biphenyl. *J Chem Phys* 1946;14:733–8.
51. Hendrickson JB. Molecular geometry. I. Machine computation of the common rings. *J Am Chem Soc* 1961;83:4537–47.
52. Boyd DB, Lipkowitz KB. Molecular mechanics. The method and its underlying philosophy. *J Chem Educ* 1982;59:269–74.
53. Boyd DB, Lipkowitz KB, editors. Preface on the meaning and scope of computational chemistry. In: *Reviews in Computational Chemistry*, Vol. 1. New York: VCH, 1990. p. vii–xii.
54. Gund P, Grabowski EJJ, Smith, GM, Andose JD, Rhodes JB, Wipke WT. In: Olson EC, Christoffersen RE, editors. *Computer-assisted drug design*. ACS Symposium Series 112. Washington, DC: American Chemical Society, 1979. p. 526–51.
55. Allen FH, Bellard S, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummelink-Peter BG, Kennard O, Motherwell WDS, Rodgers JR, Watson DG. The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information. *Acta Crystallogr, Sect B* 1979; B35:2331–9.
56. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr., Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 1977;112:535–42.
57. Martin YC. *Quantitative drug design. A critical introduction*. New York: Dekker, 1978.
58. Hansch C, Leo A. *Substituent constants for correlation analysis in chemistry and biology*. New York: Wiley, 1979.
59. Corey EJ, Long AK, Rubenstein SD. Computer-assisted analysis in organic synthesis. *Science* 1985;228:408–18.
60. Rubenstein SD. Electronic documents in chemistry, from ChemDraw 1.0 to present. Abstracts of Papers, 228th ACS National Meeting, Philadelphia, PA, August 22–26, 2004. CINF-054.
61. Helson HE. Structure diagram generation. In: Boyd DB, Lipkowitz KB, editors. *Reviews in Computational Chemistry*, Vol. 13. New York: Wiley-VCH, 1999. p. 313–98.
62. Monmaney T. Robert Langridge: His quest to peer into the essence of life no longer seems so strange. *Smithsonian* 2005;36 (8):48.
63. Langridge R. Computer graphics in studies of molecular interactions. *Chem Industry (London)* 1980 (12):475–7.
64. Langridge R, Ferrin TE, Kuntz ID, Connolly ML. Real-time color graphics in studies of molecular interactions. *Science* 1981;211 (4483):661–6.

65. Vinter JG. Molecular graphics for the medicinal chemist. *Chem Britain* 1985;21(1), 32, 33–5, 37–8.
66. Boyd DB. Molecular modeling—Industrial relevance and applications. In: *Ullmann's Encyclopedia of Industrial Chemistry*, 6th edition. Weinheim, Germany: Wiley-VCH, 1998.
67. Gund P, Andose JD, Rhodes JB, Smith GM. Three-dimensional molecular modeling and drug design. *Science* 1980;208 (4451):1425–31.
68. Boyd DB, Marsh MM. Computational chemistry in the design of biologically active molecules at Lilly. Abstracts of 183rd National Meeting of the American Chemical Society, Las Vegas, Nevada, March 28–April 2, 1982.
69. Boyd DB. Profile of computer-assisted molecular design in industry. *Quantum Chemistry Program Exchange (QCPE) Bulletin* 1985;5:85–91.
70. Lipkowitz KB, Boyd DB. Improved job market for computational chemists. In: *Reviews in Computational Chemistry*, Vol. 12. New York: Wiley-VCH, 1998. p. v–xiii.
71. Marshall GR, Barry CD, Bosshard HE, Dammkoehler RA, Dunn DA. The conformational parameter in drug design: The active analog approach. *Computer-assisted drug design*. ACS Symposium Series 112. Washington DC: American Chemical Society, 1979. p. 205–26.
72. Martin YC, Bures MG, Willett P. Searching databases of three-dimensional structures. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 1. New York: VCH, 1990. p. 213–63.
73. Grethe G, Moock TE. Similarity searching in REACCS. A new tool for the synthetic chemist. *J Chem Inf Comput Sci* 1990;30:511–20.
74. Dewar MJS, Healy EF, Stewart JJP. Location of transition states in reaction mechanisms. *J Chem Soc, Faraday Transactions 2: Mol Chem Phys* 1984;80:227–33.
75. Stewart JJP. MOPAC: A semiempirical molecular orbital program. *J Computer-Aided Mol Des* 1990;4:1–105.
76. Stewart JJP. Semiempirical molecular orbital methods. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 1. New York: VCH, 1990. p. 45–81.
77. Stewart JJP. MOPAC: A semiempirical molecular orbital program. *Quantum Chemistry Program Exchange*, 1983. Prog. 455.
78. Allinger NL. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J Am Chem Soc* 1977;99:8127–34.
79. Burkert U, Allinger NL. *Molecular mechanics*. ACS Monograph 177. Washington DC: American Chemical Society, 1982.
80. Leo AJ. Some advantages of calculating octanol-water partition coefficients. *J Pharmaceut Sci* 1987;76:166–8.
81. Leo AJ. Hydrophobic parameter: measurement and calculation. *Methods Enzymol* 1991;202:544–91.
82. Leo, AJ. Calculating log P_{oct} from structures. *Chem Rev* 1993;93:1281–306.
83. Enslein, Kurt. Estimation of toxicological endpoints by structure-activity relationships. *Pharmacol Rev* 1984;36 (2, Suppl.):131–5.

84. Enslein, K. An overview of structure-activity relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. *Toxicol Industrial Health* 1988;4:479–98.
85. Boyd DB. Application of the hypersurface iterative projection method to bicyclic pyrazolidinone antibacterial agents. *J Med Chem* 1993;36:1443–9.
86. Matthews DA, Alden RA, Bolin JT, Filman DJ, Freer ST, Hamlin R, Hol WG, Kisliuk RL, Pastore EJ, Plante LT, Xuong N, Kraut J. Dihydrofolate reductase from *Lactobacillus casei*. X-ray structure of the enzyme methotrexate-NADPH complex. *J Biol Chem* 1978;253:6946–54.
87. Matthews DA, Alden RA, Freer ST, Nguyen HX, Kraut J. Dihydrofolate reductase from *Lactobacillus casei*. Stereochemistry of NADPH binding. *J Biol Chem* 1979;254:4144–51.
88. Everett AJ. Computing and trial and error in chemotherapeutic research. In: Leeming PR, editor. *Topics in Medicinal Chemistry*, Proceedings of the 4th SCI-RSC Medicinal Chemistry Symposium, Cambridge, UK, Sept. 6–9, 1987. Special Publication 65. London: Royal Society of Chemistry, 1988. p. 314–31.
89. Ito A, Hirai K, Inoue M, Koga H, Suzue S, Irikura T, Mitsuhashi S. In vitro antibacterial activity of AM-715, a new nalidixic acid analog. *Antimicrob Agents Chemother* 1980;17:103–8.
90. Koga H, Itoh A, Murayama S, Suzue S, Irikura T. Structure-activity relationships of antibacterial 6,7- and 7,8-disubstituted 1-alkyl-1,4-dihydro-4-oxoquinoline-3-carboxylic acids. *J Med Chem* 1980;23:1358–63.
91. Koga H. Structure-activity relationships and drug design of pyridonecarboxylic acid type (nalidixic acid type) synthetic antibacterial agents. *Kagaku no Ryoiki, Zokan* 1982;136:177–202.
92. Perun TJ, Propst CL, editors. *Computer-aided drug design: Methods and applications*. New York: Dekker, 1989.
93. Brünger A, Karplus M, Petsko GA. Crystallographic refinement by simulated annealing: application to crambin. *Acta Crystallogr, Sect A* 1989; A45:50–61.
94. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4:187–217.
95. Kollman, P. Molecular modeling. *Annu Rev Phys Chem* 1987;38:303–16.
96. McCammon, JA. Computer-aided molecular design. *Science* 1987;238(4826): 486–91.
97. Reddy MR, Erion MD, Agarwal A. Free energy calculations: Use and limitations in predicting ligand binding affinities. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 16. New York: Wiley-VCH, 2000. p. 217–304.
98. Karin S, Smith NP. *The supercomputer era*. Boston: Harcourt Brace Jovanovich, 1987.
99. Wold JS. Supercomputing network: A key to U.S. competitiveness in industries based on life-sciences excellence. Testimony before the U.S. Senate, Commerce, Science and Transportation Committee Science, Technology and Space Subcommittee. <http://www.funet.fi/pub/sci/molbio/historical/biodocs/wold.txt>

100. Milosevich SAF, Boyd, DB. Supercomputing and drug discovery research. *Perspect Drug Discovery Design* 1993;1:345–58.
101. Richon AB. A history of computational chemistry. <http://www.netsci.org/Science/Compchem/feature17a.html>.
102. Halgren TA. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *J Am Chem Soc* 1992;114:7827–43.
103. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94. *J Comput Chem* 1996;17:490–519.
104. Halgren TA. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J Comput Chem* 1996;17:520–52.
105. Halgren, T. A. Merck Molecular Force Field. III. Molecular geometrics and vibrational frequencies for MMFF94. *J Comput Chem* 1996, 17:553–86.
106. Halgren TA, Nachbar RB. Merck molecular force field. IV. Conformational energies and geometries. *J Comput Chem* 1996;17:587–615.
107. Halgren TA. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data and empirical rules. *J Comput Chem* 1996;17:616–41.
108. Halgren TA. MMFF VI. MMFF94s option for energy minimization studies. *J Comput Chem* 1999;20:720–29.
109. Boyd DB, Lipkowitz KB, History of the Gordon Research Conferences on Computational Chemistry. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 14. New York: Wiley-VCH, 2000. p. 399–439.
110. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC. MacroModel—An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J Comput Chem* 1990;11:440–67.
111. Sprague PW. Catalyst: A computer aided drug design system specifically designed for medicinal chemists. *Recent Advances in Chemical Information*, Special Publication 100, Royal Society of Chemistry, 1992. p. 107–11.
112. Blaney JM, Dixon JS. A good ligand is hard to find: Automated docking methods. *Perspect Drug Discovery Design* 1993;1:301–19.
113. Boehm H-J. Fragment-based de novo ligand design. *Proceedings of the Alfred Benzon Symposium No. 39*, 1996. p. 402–13.
114. Murcko MA. Recent advances in ligand design methods. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 11. New York: Wiley-VCH, 1997. p. 1–66.
115. Clark DE, Murray CW, Li J. Current issues in de novo molecular design. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 11. New York: Wiley-VCH, 1997. p. 67–125.
116. Lauri G, Bartlett PA. CAVEAT: A program to facilitate the design of organic molecules. *J Comput-Aided Mol Des* 1994;8:51–66.
117. Walters WP, Stahl MT, Murcko MA. Virtual screening—An overview. *Drug Discovery Today* 1998;3:160–78.

118. Lewis RA, Pickett SD, Clark DE. Computer-aided molecular diversity analysis and combinatorial library design. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 16. New York: Wiley, 2000. p. 1–51.
119. Boyd DB, Lipkowitz KB, editors. Trends in the job market for computational chemists. *Reviews in Computational Chemistry*, Vol. 7. New York: VCH, 1996. p. v–xi.
120. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25.
121. Hansch C, Leo A, Hoekman D. *Exploring QSAR: Hydrophobic, electronic, and steric constants*. Washington DC: American Chemical Society, 1995.
122. Todeschini R, Consonni V. *Handbook of molecular descriptors*. Berlin: Wiley-VCH, 2000.
123. Karelson M. *Molecular descriptors in QSAR/QSPR*. New York: Wiley, 2000.
124. Boyd DB. Progress in rational design of therapeutically interesting compounds. In: Liljefors T, Jørgensen FS, Krosgaard-Larsen P, editors. *Rational molecular design in drug research*. Proceedings of the Alfred Benzon Symposium No. 42. Copenhagen: Munksgaard, 1998. p. 15–23.
125. Boyd DB. Innovation and the rational design of drugs. *CHEMTECH* 1998;28 (5):19–23.
126. Boyd DB. Rational drug design: Controlling the size of the haystack. *Modern Drug Discovery*, November/December 1998. p. 41–8.
127. Boyd DB. Drug design. In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman P, Schaefer HF III, editors. *Encyclopedia of computational chemistry*, Vol. 1. Chichester: Wiley, 1998. p. 795–804.
128. Boyd DB. Is rational design good for anything? In: Parrill AL, Reddy MR, editors, *Rational drug design: Novel methodology and practical applications*. ACS Symp Series 719. Washington, DC: American Chemical Society, 1999. p. 346–56.
129. Zurer P. Crixivan. *Chem Eng News*, June 20, 2005. p. 54.
130. Dyason JC, Wilson JC, Von Itzstein M. Sialidases: Targets for rational drug design. In: Bultinck P, De Winter H, Langenaeker W, Tollenaere JP, editors, *Computational medicinal chemistry for drug discovery*. New York: Dekker, 2004.
131. Walters DE, Hopfinger AJ. Case studies of the application of molecular shape analysis to elucidate drug action. *THEOCHEM* 1986 27:317–23.
132. DePriest SA, Mayer D, Naylor CB, Marshall GR. 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors: a comparison of CoMFA models based on deduced and experimentally determined active site geometries. *J Am Chem Soc* 1993;115:5372–84.
133. Congreve M, Murray CW, Blundell TL. Structural biology and drug discovery. *Drug Discovery Today* 2005;10:895–907.
134. Boyd DB, Palkowitz AD, Thrasher KJ, Hauser KL, Whitesitt CA, Reel JK, Simon RL, Pfeifer W, Lifer SL, Takeuchi K, Vasudevan V, Kossoy AD, Deeter JB, Steinberg MI, Zimmerman KM, Wiest SA, Marshall WS. Molecular modeling and quantitative structure-activity relationship studies in pursuit of highly

- potent substituted octanoamide angiotensin II receptor antagonists. In: Reynolds CH, Holloway MK, Cox HK, editors. *Computer-aided molecular design: Applications in agrochemicals, materials, and pharmaceuticals*, ACS Symp. Series 589, Washington, DC: American Chemical Society, 1995. p. 14–35.
135. Boyd DB, Lipkowitz KB. Examination of the employment environment for computational chemistry. In: Lipkowitz KB, Boyd DB, editors. *Reviews in Computational Chemistry*, Vol. 18. New York: Wiley-VCH, 2002. p. 293–319.
136. Schmid EF, Smith DA. Is declining innovation in the pharmaceutical industry a myth? *Drug Discovery Today* 2005;10:1031–9.