A Drug Is Born: Its Information Facets in Pharmaceutical Research and Development

HORACE D. BROWN
7130 Windover Drive, Durham, North Carolina 27712

Received March 11, 1985

A cursory look is taken at pharmaceutical research and development and its related information world in the 1960s, at the present time, and in the future. Several phases in the birth of a new drug are examined in relation to the data/information/knowledge/wisdom ("information") being generated for the new chemical entity along with retrieval of information already existing in proprietary or published stores. Generation, evaluation, transfer ("communication"), and utilization are treated as facets of the new drug's information profile. The increasing utility of computer-based systems for rational drug design and for information resource management is noted.

INTRODUCTION

The pharmaceutical industry is a rainbow-wide spectrum of diversity. Its scientific and technical challenges include both the ancient and newly developing diseases of man and animals. Its customers range from the starving millions to the "jet set". It is under the intense scrutiny of both national and international regulatory bodies as well as a number of consumer-activist groups. Management in such a setting is a real challenge.

Pharmaceutical research and development (R&D), which carries the burden for the scientific and technical aspects of the industry, is an exciting place to work. Upward of 50 major disciplines are required to handle the range of programs normally found in a major company. The skills needed range from robotics to board-certified physicians. Teamwork, project management, and effective communication (information transfer) are critical at both the disciplinary and interdisciplinary levels. Proprietary and published information resource management (IRM) assumes enormous proportions.

The scope of this article will include a cursory look at the period from the 1960s to the present, with a limited glance to the future. Emphasis will center on data/information/knowledge/wisdom (collectively called "information"), generation, evaluation, collection, classification, storage, retrieval, dissemination (*transfer*), publication, and utilization (information has no value if not utilized). The jargon will be largely familiar to information scientists, but for the uninitiated some definitions or a glossary of terms may be helpful.[1-3] In passing, one should note that information science (still in gestation) and scientific information, information resource management (IRM) and information management, information dissemination and information transfer, discussion and communication, and finally data and information are not identical.

Horace D. Brown obtained his B.S. (cum laude) from Berry College in chemistry and mathematics, an M.S. from Emory University in organic chemistry, and a Ph.D. from the Iowa State University in chemistry and entomology. He was an instructor of Navy V-7 trainees and a chemist in the Manhattan Project. He worked 14 years in Medicinal Chemistry at Merck & Co., Inc., on insecticides, fungicides, analgesics, chemotherapeutics, steroids, and antiparasitic agents and 23 years in Scientific Information on structure/substructure searching, proprietary databases, and computer-assisted drug design. He has been a corecipient of the First New Jersey Patent Award for thiabendazole and authored or coauthored 32 U.S. patents and 25 publications. He has retired from Merck & Co., Inc., as Senior Director, Scientific Information. His professional affiliations include the American Chemical Society, Drug Information Association, Association for Computing Machinery, New York Academy of Sciences, Science Information Subsection/Pharmaceutical Manufacturers Association (twice elected chairman), and Industrial Technical Information Managers Group.

PERSPECTIVE

The early 1960s brought significant change to the pharmaceutical industry. It was only a quarter century since the heyday of the vitamins and sulfas and the beginnings of expanded R&D in the industry. Bold steps by a few selected companies attempted to woo the scientific talent needed from academia and Europe. Financial support for R&D required a leap of faith. Georeg Merck's "Medicine is for the people—take care of them and the profits will follow" was not universally accepted. Antibiotics, steroids, mental health drugs, and antitumor agents were major projects in R&D. The bombshell that shook the industry was the new Food, Drug and Cosmetic Act requiring *proof* of efficacy and safety following the thalidomide tragedy. "Medical opinion" was no longer adequate for the marketing of a new chemical entity (NCE).

From an information perspective a major revolution was in the wings. Computers (mostly key punch and sorters) were being used to collect, store, classify, retrieve, and disseminate information. In the face of the "literature explosion" in the published domain and the new requirement for data/information generation in proprietary quarters, this developing computer technology was a godsend. Management of R&D *reluctantly* sacrificed another scientist at the bench for a "staff" person in the library, for data clerks, statisticians, etc. Bench scientists responded to the literature explosion in varied manner. Some trimmed their reading goals; some became members of "invisible colleges" for their discipline; some "threw up their hands" and abdicated their literature responsibilities. *Some* (thank Heaven) came to the scientific information departments seeking help. A few were even willing to be guinea pigs in *S*elective *D*issemination of *I*nformation (SDI) projects

0095-2338/85/1625-0218$01.50/0 © 1985 American Chemical Society

where information scientists (with the help of computer systems) made selections for them.

Publishers, abstracting services, and database builders struggled to keep up. A number turned to the new computer technology (e.g., National Library of Medicine and Chemical Abstracts Service). Early effort was directed to the editing, printing, and photocomposition activities. Fortunately for us, all compound registration (CAS) and machine encoding of chemical structures (or surrogates) got started. Thus, the chemical typewriter,[4] Wiswesser line-formula notation,[5] and substructure searching[6] led to the present-day foundation for the management of chemical substances and their properties—so crucial to drug research and information resource management in R&D.

Proprietary data generation grew to astronomical proportions, both because of the growing number of scientists employed and because of the increasing spectrum of requirements by regulatory bodies. Notebooks, project reports, departmental memoranda, outside correspondence related to research contracts, and most particularly correspondence (phone calls and letters) with regulatory agencies became crucial to IRM and R&D management. Microfilm and microfiche activities blossomed to reduce the size of archival storage requirements. Library automation (e.g., M.I.T. INTREX) got started. The library community began to turn its attention to information transfer in contrast to its traditional "books on the shelf" philosophy.[7]

Another significant development had its beginning. Creation of new (e.g., Drug Information Association) and expansion of existing professional groups concerned with information resource management provided a forum for the exchange of developing methodologies to cope with the burgeoning data/information/knowledge growth. The concept of an information center was reduced to practice.[8] Database vendors (usually lacking scientific staff) slowly responded to "user groups" having a common interest in the improvement of content, format, and automated retrieval.

During the 1960s, a small but significant Literature Subcommittee of the Research and Development Section of the Pharmaceutical Manufacturers Association[9] initiated exploratory discussions on information needs with a number of database generators (e.g., National Library of Medicine, Chemical Abstracts Service, Biological Abstracts, Food and Drug Administration, and Dr. Eugene Garfield—the latter in relation to steroid coding and timely alerting to new chemicals from the published literature). In 1969, this Subcommittee recommended and received approval for Subsection Status for what is today Science Information Subsection/R&D/PMA. This forum has provided useful information transfer in the information resource management arena and has also served to instruct an increasing number of neophytes in such activities.

TODAY

Published Literature Databases. Compounded growth of the literature, user-shared costs for development, production, and maintenance, computer technology, and an improving vendor attitude toward user groups have led to a large number of vendors dealing with the world's published literature and patents. Program, project, and personal-level interests can now be matched (via word profile descriptors) to the computer-based dictionaries for a given profile, and "hits" are printed out according to prescribed formats. With current graphics-based work stations, the "hits" may be displayed on the screen or printed on-line or off-line (cheaper rate). Such profiles may be stored and searches performed at file updates. The technique is applicable to proprietary or to published databases and with improved telecommunications may be done with files across the ocean as well as across the street. Information marketing is now an accepted concept.[10] There appears to be a vertical integration under way with database builders becoming interested in direct sale to users (e.g., Chemical Abstracts Service and Pergamon Press).

Role of the Information Scientist. Over the time span considered, the role of the information scientist has gradually evolved from "staff" to "peer acceptance". In the early 1960s the perennial line vs. staff distinction was much in evidence. Relevance vs. recall studies was a central theme then. Scientists generally preferred "do-it-yourself" philosophy and in addition expressed concern about loss of return from browsing. Ideally, the information scientist should be a full-blown, fully accepted, working member of project teams. From retrospective searching of a proposed topic in both published literature and patents, through the derivation of project goals, to the establishment and operation of program (general), project (selected), and personal (specific) profiles during the research phase, and to monitoring of programs in the market place, the information scientist has a real contribution to make to the project team. Methodical, thorough, on-going searches pose a serious challenge at both the scientific and information levels.

Selection, Training, Evaluation, and Reward of Information Scientists. Information scientists are a valuable part of the drug discovery and development process. Hence, their selection, training, evaluation, and reward (when successful) need serious attention from management. Basic training in a hard science (perferably chemistry), a certain amount of earned peer acceptance, a true sense of psychic reward from helping others, a "nose for news", a keen sense of curiosity, a capacity to appreciate both intricate detail and the "big picture", personal drive, oral and written communication skills, and scientific integrity beyond reproach all contribute to a likely profile for success. Evaluation, motivation, and reward of information scientists require special effort. Since direct involvement with users is highly preferred, some feedback from clients and a measure of user satisfaction (even if subjective) are needed. Involvement with project teams and management through reports and meetings offer a chance for exposure/recognition. Supervisory/management training courses should be freely available to those showing interest in the administrative arena. Job evaluations are a thorny problem for information managers since the average personnel department is still biased by the "books-on-the shelf" image of librarians. Accountability is usually limited to an "assist" at best with scientists and management given major credit for the utilization facet of information.

BIRTH OF A DRUG

For the purpose of discussion, I will consider the following phases in drug discovery and development: (1) project definition; (2) lead discovery; (3) lead development; (4) drug development; (5) market launch; (6) clinical utility. Though space does not permit full discussion, the major scientific activities and information facets of the individual phases will be mentioned.

Project Definition. The proposal by a scientific/marketing team to embark on the search for a new drug for a major human or animal disease is a major undertaking (10 years of time and up to one-hundred million dollars). Therefore, it is imperative that each discipline represented in the project do its homework thoroughly—both in collection and evaluation of the extant data/information/knowledge. Sources must include proprietary stores, published literature and patents, and consultation with "movers and shakers" in the field of interest. In addition, it is not just the initial retrospective search but an ongoing commitment that must be made. True communication (*information transfer*) among members of the

team and with management must be explicit and continuing.

Once the specific objective is defined to the satisfaction of team members and management, some realistic time tables and cost estimates must be derived. Management must remember that scientists are optimistic re time tables and usually underestimate the effort required to reach goals. The objective as agreed to must become part of the project file and, if later amended (by mutual agreement), likewise recorded.

Lead Discovery. In many R&D projects, assignments in the early stage frequently rest heavily with chemists and biologists. Activities center around the collection of compounds and their preliminary biological testing by a defined protocol. The methodology selected by the biologist is a critical part of the project since results of this first-pass test become the eyes and ears for the chemist. Credibility and candor are important in the information transfer that ensues. This applies not only to the new data being generated but to any contributions made by the information scientists from proprietary or published stores. The chemist frequently uses a variety of sources for the test compounds—synthesis, collection from sample/catalog stores, microbial broths, plant extracts, and molecular modification of known drugs. In recent times rational drug design (RDD) has received more attention. This may stem from hypotheses regarding the pharmacological or enzymatic nature of the disease process—frequently involving an agonist or antagonist for a known enzyme and its substrate. Molecular modeling, using three-dimensional models in color, has proven espcially stimulating to medicinal chemists. Whatever the source of the compound, the major goal in this phase is to discover a novel structure with interesting biological effect as measured in the in vitro, enzyme inhibition, perfused organ, or whole animal model chosen for the initial evaluation.

Large volumes of new data are, of course, generated by each of the disciplines involved. It is practical to limit the number of substances that pass this first "screen" to 1–2%. Larger numbers weigh heavily on time and resources for repeat assays, follow-up, and minor structural variations, etc. Criteria for passing on to the next level of evaluation include not only the degree of biological activity but structural type and relation to previously known "actives", as well as chance for patentability and known toxicity patterns. Chemists are wary of early discards based on toxicity bias, but this is a liability not to be forgotten. In all these aspects, judgments (evaluation methodology) are involved. One must balance the desire to pursue all viable leads against the practical aspects of return on investment of time and resources.

In terms of the chemists' activities, it is not just the chemicals synthesized but ancillary data that must be collected, evaluated, and recorded. In both synthetic programs and natural product sources spectral data are extremely valuable. This includes ultraviolet (UV), infrared (IR), nuclear magnetic resonance (NMR), and mass spectrometry (MS). Requests for tests to the various service laboratories or self-service units frequently form the first recorded indication of a given chemical structure in hand. This becomes very valuable in patent application or interference prosecution. Though chemists frequently reject the idea, the writer believes that the "publication" of planned agendas for synthesis in proprietary computer files will prove to be meritorious (not only to retain ownership for the proprietor but to avoid duplicate effort). For spectral data, for sample transmission requests, and, in fact, for laboratory records generally, data/record management protocols need to be followed. These are usually derived by the proprietary information scientists in concert with legal and patent personnel. This is not only wise from a scientific record keeping perspective but may well come under "Good Laboratory Practices" guidelines when compound safety is involved. Biological protocols, instrument calibration, and audit trails through the scientists' notebooks all require attention from information scientist and laboratory personnel.

As the proprietary data is generated/evaluated in the search for a new lead, the published literature and patents must be combed routinely by the information scientists assigned to this task. Bioassay procedures, new reports of biological activity, and especially quick-publish foreign patents are reviewed to keep track of competitive activity in the chemical structure or bioassay areas of interest. Known toxicity (mutagenic, teratogenic, carcinogenic) for closely related chemical structures is collected and disseminated to the team members.

Proprietary reports from the project team members must be collected, indexed, stored for later retrieval, and circulated as required to other team members or those with a "need to know". Progress made and problems encountered are highlighted in profile reports to those concerned.

Success in this phase is marked by the finding of one or more structural types that possess "interesting" biological activity, offer a chance for novelty (patent protection), and represent a class of chemicals not usually perceived as "toxic". Though some residual activity may persist in this phase, the team then takes such a lead to the next step.

Lead Development. In this phase the goal is to exploit the findings from the lead discovery phase. Both offensive and defensive aspects come into play. One wishes to derive the data needed to define the structure–activity relationships (SAR) for the lead compound(s) in question. Efficacy and safety are dominant, but chemical novelty is very important since the high cost of drug development requires some assurance of patent protection.

Systematic structural variation is applied by the medicinal chemists to determine the chemical attributes (electronic, steric, oil/water distribution coefficients, etc.) that have influence (positive or negative) on the specific biological activity of interest. On the biological side it is important to have a whole animal model so that one may begin to consider absorption, distribution, plasma levels, plasma half-life, metabolism, and mode-of-action studies.

If the biological data is sufficiently quantifiable, then regression analyses should be conducted to determine what quantitative structure–activity relationships (QSAR)[11] exist. Such analyses may be helpful in projecting the specific structure to be synthesized to find maximum biological activity.

It is usually during this phase that a sound basis for patent application is found. Thus, careful record keeping re synthetic agendas, dates of request for test, actual test data in relation to standards and literature/patent searches must be carefully documented.

Computer-based systems (e.g., Chemical Abstracts Service Registry System) are very useful in this phase. Structure and substructure searching (vs. biological activity) of proprietary files of chemicals is a priceless tool. Connection table based systems[12,13] that have graphics capabilities are especially useful. In addition, sample transmission records, unique compound numbers, sample inventories, physical properties (especially spectra), and references to original notebooks have all proven valuable. Likewise, it is important to have biological data available in a computer-based system[14] that registers a unique compound number, a chemical name/structure, protocols, test type, test numbers, investigator name, dates (received and tested), raw data from tests, and expert evaluation. Finally, a project document system for notebooks, memos, reports, meeting agendas and minutes, and relevant correspondence is also very important.

Use of computer-based systems for molecular modeling, for drug/receptor docking, for energy minimizations, for estimation of drug–receptor binding energies, for graphic display, and for molecular modeling files for individual or group use

has increased dramatically over the last decade.[15] X-ray crystallography is being regularly applied to enzyme/inhibitor studies. Such patterns when displayed in color as a three-dimensional model are extremely useful to the medicinal chemist in drug design activities. To "jiggle" a bound complex with estimates of the energy consequences is not only beautiful but useful. This methodology will doubtless continue to expand dramatically.

Drug Development. From the lead development stage a prime candidate (plus one or two backups) is selected for development as a product. Proposals to management are made with both research and marketing support. The package includes objectives (potential claims), time tables, manpower, and dollar cost estimates by clinical phase (phase I, safety; phase II, efficacy under carefully controlled conditions; phase III, general clinical/field studies), new drug application, and market support (combinations, new claims, new dosage forms, and publications). With management approval, all required disciplines receive their assigned tasks.

Development chemists assume the task of preparing kilograms of bulk chemical for safety studies, for pharmaceutical dosage forms, and ultimately for clinical studies. Pharmacists initiate studies to determine the salt/derivative form, dosage-form preparations (tablet, capsule, injection), and stability studies. Safety assessment (toxicology) units start the required studies in the appropriate species to file the Investigational New Drug Application (IND) and to assure the physicians of the safety clearance for particular dosage regimens. Radiolabeled compound is prepared at the particular atom in the molecule to ensure that questions of transport, metabolism, and blood half-life are answered. In the case of veterinary drugs, the label will be useful in determining residue in edible carcass.

Physicians initiate planning for the clinical phases. Investigators must be contacted and contracts drawn for the required studies. Claims targets must be defined vs. patient populations to be studied. Case reports need to be carefully designed[16] to collect the patient data.

It is important to note the enhanced role now played by statisticians in laboratory, clinical, and veterinary studies. In the early 1960s most laboratories had a small staff of statisticians who were usually contacted only after the scientists were unable to decipher meaningful results from their data plots ("If the biology isn't convincing on its own, no statistician is going to make it so"). At the present time statisticians abound at every stage in laboratory, clinical, and veterinary groups. Not only are they a key part of the data evaluation, but more importantly, they are dominant in the experimental design. Thus medical opinion (the usual basis for FDA approval in the 1950s) has given way to statistically sound experiments and to statistically sound data derived under controlled (frequently double-blind) conditions.

The data generated by each discipline in the drug development process must be done under "state of the art" conditions with reference to documented protocols. Good Laboratory Practices, Good Clinical Practices, and Good Manufacturing Practices must be rigorously adhered to. Evaluation of the data by the disciplinary expert (study director) is followed by reports to the team, to management, and to the requisite regulatory bodies—nationally and internationally. Documentation is crucial at every step from drug preparation to purity testing, to safety testing, to formulation (safety tested again!), to clinical study, to evaluation and reporting.

Information scientists dealing with proprietary data/information are a dominant factor in this time frame. Reams of paper are flowing! Those dealing with published material, however, are not without responsibilities. Key journals dealing with new compounds, new biological activities, and especially patents must be carefully reviewed. All key commercial databases are routinely scanned at each update to be able to define the degree of competitive activity. It should be remembered that marketing is still years away (even if the studies are successful). Hence, in countries where product patents do not exist or where regulatory controls are less stringent the competition can still cause concern.

The original chemical/biological team is now hard at work extending patent coverage, seeking second generation candidates, and preparing manuscripts for publication (when management feels the patent applications are on firm footing).

A major effort is ongoing in the toxicology (safety assessment) area. Both short-term (90 days) and long-term (2–3 years or lifetime) studies are in progress in the appropriate species (mice, rats, dogs, rabbits, or primates) to define the mutagenic, teratogenic, carcinogenic, or organ toxicity attributes of the drug candidate. These studies, of course, lead the clinical/veterinary studies. For drugs with a residue problem it may be necessary to conduct additional studies with any metabolite found to accumulate in edible carcass. Each new formulation prepared by the pharmacists must be tested before release to physicians. Each new process used by the developmental chemists must be shown to be free of unwanted contaminants.

In preparation for the filing of the Investigational New Drug Applications all major disciplines must search the literature with the help of the information scientists to locate any reports of toxicity for the candidate or its close relatives. Proprietary data must be collected, evaluated, and reported. The chemistry of the compound (synthesis, assay procedures) is of course included. An internal computer system for tracking synthesis of new batches of the chemical and their history is a valuable tool at this stage. Curriculum vitae, Statements of Investigator, study plans, and numerous other documents are needed from the clinical group. An inventory of all correspondence and compound shipment, as well as all incoming patient data, must be maintained. A useful computer-based system for clinical data management has been described.[17]

When process chemists have delivered enough material to initiate safety studies and preliminary clinical tests, they have a long-term task to initiate. This involves an ultimate manufacturing process. First lots are usually delivered from a "blowup" of the laboratory synthesis. There is need for an economical, practical (few steps), safe, low-labor, minimal-capital, patentable process for domestic and/or international manufacture. Though seldom appreciated by academic chemists, some of the most challenging problems in chemistry and engineering in the pharmaceutical industry are solved by this group. High-dosage (1 g/day) or exotic-structure chemicals pose an exceptional challenge. It should be remembered that many drugs contain one to several asymmetric centers—thus requiring expensive resolutions or the design of processes giving rise only to the desired isomer. The information scientists support this group through references to internal notebooks, to spectral data, and to novelty searches for compounds or processes and for safety data (physical or physiological) on the intermediates. Safety in the scaleup is a major concern. Personnel-exposure tolerances for *all* intermediates, solvents, and final products need to be determined. Industrial toxicology as well as the animal toxicology groups thus participate.

Information scientists responsible for proprietary materials must collect, index, store, and retrieve all the agendas, meeting minutes, management decisions, departmental reports, and correspondence associated with the development. This usually means a new file is started when the drug candidate is selected. Project reports are usually circulated to those in a need to know category. In addition, the files must be protected from ca-

tastrophe by microfilming and storing off-site.

Throughout the clinical studies (and later!) adverse-reaction reporting is a major task in drug development. Any toxicity noted in animal or human studies must be reported to the appropriate regulatory bodies. This applies to both internal data as well as the published literature. A Drug Information Association workshop has covered this issue in detail.[18]

During clinical phases I–III, there is a continuing schedule of periodic reports and scheduled meetings with Food and Drug Administration (for domestic) and with the respective foreign agencies where the drug is in study. Enormous volumes of data are generated, and all must be carefully documented by the proprietary information scientists. Compilation of the New Drug Application is a major undertaking, resulting in the creation of hundreds of volumes. The final application is frequently delivered in a truck. Copies of this filing must be kept internally, and copies (usually microfilm or microfiche) are distributed to other divisions or sites that require such. These volumes will be referenced (and updated!) over the period of review by the FDA (1–3 years).

Market Launch. When the regulatory agency gives approval for release of the new drug, the sponsoring organization must complete the steps required for market launch. Risk-basis manufacturing facilities for both bulk drug and dosage form(s) will usually have been completed. Processes for both bulk drug and dosage form must be transferred from research to production (another major information-transfer process). Demonstration of the processes by R&D personnel is followed by monitoring until yield and quality-assurance specifications are met. Ultimately, some of these processes will be released in publications—others may be retained as trade secrets.

Printing of package circulars and labeling for the packaged drug is a major last minute task. Final wording for the claims, indications, and contraindications is usually the last portion to be cleared by FDA. Wholesale/drugstore supplies of the new packages must be manufactured and shipped. Extensive training sessions of the sales representatives are conducted (information-transfer again!). Both background (disease-related) and specific (claims, side effects) aspects must be thoroughly understood by each representative. They in turn will be responsible for bringing this information to the prescribing physician or the hospital pharmacist.

Advertising copy approval is an internal task requiring the cooperative effort of chemist, biologist, toxicologist, physician, marketing representative, and legal personnel. Exact wording is derived for the advertisements to be sent to professional journals so that company policies and FDA guidelines are compiled with. Working committees and upper management review these in detail before release.

Professional seminars are frequently held in connection with the launch of a new drug. These serve to bring together experts who have worked on specific aspects of the drug development. They also serve (through press coverage) to give an announcement to the lay public (usually not familiar with any technical papers that have been published earlier). Happily for information transfer, it is now acceptable practice for organizational physicians who have served as clinical monitors to write summary publications covering a broad range of studies. In the 1960s such practice was not in vogue. Likewise, toxicologists now summarize the findings of their long-range studies. Both of these serve as feedback to medicinal chemists already concerned with the next generation of drug for the disease entity in question.

Clinical Utility. Several years (sometimes earlier!) after the drug launch usually "tells the tale" for the NCE. If the drug finds acceptance in the market place—and especially if it is a *breakthrough* drug—it will not only receive support from its sponsor, but competition will add to the family. Thus for a successful drug one may expect new derivatives, new combinations, new dosage forms, new claims, and a growing body of publications. Occasionally, new uses may be found—projections from laboratory or clinical observations. Thus an antiparasitic agent, ivermectin, was recently shown to have promise in a human disease, "river blindness", which affects many people in Africa.[19] During this fiery furnace testing of the drug in the open marketplace, its strengths and weaknesses come to light. Whatever its fate, the sponsor must continue to monitor both proprietary data and publications and report results that bear on safety to the regulatory bodies. Adverse reaction reporting is now covered internationally. Hence, the incidence of side effects vs. unit doses consumed is available. I would like to note that such activities have been in process for some drugs for the quarter-century time frame covered by this article.

An interesting example of information transfer to busy physicians using microcomputer technology has been developed by Merck Sharp & Dohme.[20] In this program the professional representative carries a microcomputer with video screen to the physician's office. Highlights of information about a new drug may be displayed, or the physician's questions on a broader topic may be answered by dial-up to the National Library of Medicine or other commercial database. This facility thus represents an on the spot source of information for the physician.

Finally from the clinical utility phase information must be collected and fed back to the project team—and competitive teams as well will be following it. There are lessons to be learned—plus and minus—regarding the bioassay procedures used to discover the agent. The rationale or hypothesis followed needs to be evaluated in the light of the utility (or lack of it) found. One cycle has thus been completed. Both in the laboratory and in the clinic, ideas/hypotheses have been generated, evaluated, and confirmed or negated by the data produced. This information now needs to be integrated into the larger store of related knowledge about the disease entity in question. Totally new information may have been generated, existing tenets may have been confirmed, or the whole experience may have been a big failure. Expensive lessons are learned from such failures. However, both successes and failures need to be part of the "grist for the mill" of the medicinal chemist as he/she sits before the computer screen in an effort to design the next miracle drug.

FUTURE

Predictions on almost any topic are frought with the possibility of failure. Surely any topic as complex as pharmaceutical R&D will not be an exception. However, my hazy crystal ball produced some thoughts.

Rational drug design will continue to receive attention. Ongoing studies in the biochemistry of enzymes and receptors will lead to a better basis for design. These coupled with improving computer-assisted facets of drug discovery will find increasing application in the laboratories of medicinal chemists, pharmacologists, and research-minded physicians.

Molecular modeling will grow in sophistication. X-ray coordinates for enzyme/receptor/inhibitor complexes are emerging in growing numbers. Projections of these three-dimensional models on a color screen and the ability to "jiggle" these, to assess the energy consequences of substitute inhibitors, or to measure the impact of an approaching molecular fragment on a selected portion of the enzyme/receptor give powerful information to the design chemist. Improvements in both software and hardware seem imminent.

Compatible systems capable of being integrated into a large communication network (both proprietary and public) will surely come to the drug design laboratory. Security is no small

problem. However, the known data/information relevant to a given problem must be assembled (through *downloading*) into the private working file of the designer; whatever the price for security, the price for lost or missing data/information can be greater. Since security is everybody's business, system designers, system operators, and users must collaborate to provide maximum information availability with minimal security risk.

Creative process and entrepreneurship are oft-studied items. Sometimes the true genius at invention does not have the daring to be an entrepreneur. Difficult as the management of the creative person/entrepreneur is, R&D management must press forward. Though the formula for success is not well-known, some ingredients seem desirable. Among these are competent scientists, clearly defined challenges, appropriate laboratory equipment, and, the writer believes, ready access to *interdisciplinary* information. In fact, a chance to communicate (transfer information!) with scientists in other disciplines on a continuing basis would seem to be a useful stimulant. Hence, break areas, "arranged" lunch table discussions with heterogeneous disciplines, interdisciplinary seminars, and meetings might be helpful. Meetings, especially international ones, are very expensive but so is nonproductive research. Souder[21] has analyzed 100 new product cases and concluded that six conditions were correlated with entrepreneurship and successful developmental efforts.

Artificial intelligence (AI) is a topic drawing much attention. Though some have labeled it as a threat to human thinking, others are working hard to exploit its potential. It seems, like the early days of computer-based information retrieval systems, that a prime requirement is to capture, index, and store the experts' data/information/knowledge/wisdom stores and then to be able through *large* computer systems (software perhaps more crucial than hardware) to "teach" the system to respond to complex topics. If such a "tool" is to be used to support decision making in a corporate setting, then a truly complex task is at hand. Though the process by which a corporate executive arrives at a decision is poorly understood, there are some insights. A sense of value (perhaps peculiar to the individual), a political bent, or "this is how I've always done it" may equal (if not outweight) a massive array of data, the lengthy report (*information*) compiled by the staff assistant, or even the advice of a close aide. Furthermore, there may be hidden agendas, or overriding issues not known to the staff or the computer database. Hence, artificial intelligence will require some insight into these peripheral areas if it is to be helpful in decision making. It does seem clear that the computer and AI will be useful in processing large arrays of data to look for trends or correlations that escape the human capability.

Improved productivity is a goal in most organizations. Nowhere would this be more appreciated than in an R&D organization. Thinking (a major facet of R&D) is tough sledding. Henry Ford Sr. said, "Thinking is the hardest work there is, which is the probable reason why so few engage in it". Not only must R&D scientists think, but then they must put their thoughts (hypotheses) to test. This means generation of data/information/knowledge to fill the gaps in published data/information. Usually, one finds quickly that much that is published is inaccurate or inadequate. The sad state of much of our published data has been described by Branscomb.[22] He has made suggestions for federal support for more reliable databases. One might add that many R&D files of proprietary data could likewise be improved. Though correcting past mistakes is seldom stimulating, we could raise our sights so that future data is beyond reproach. Better data/information and ready access to it should help the R&D process. The management of risk is a major task in pharmaceutical R&D productivity. Though not from the pharmaceutical industry, Baillie[23] has studied federal R&D subcontractors and concluded that these managers are risk averters ("unwilling to commit 10% of firm's resources for a potential 300% increase in sales"). These managers attempted to reduce risk by reducing the scope of the technical challenge, by maintaining good customer rapport, by establishing elaborate planning and control systems, or by transferring financial responsibility to the customer. Sound familiar! More effective information management has the potential for risk minimization. With ready access to data/information one can clearly define where the gaps exist and hopefully delineate the size and nature of the gap. One then needs to estimate the cost of deriving the data/information needed to reduce or eliminate the gap. Information reduces the uncertainty[24] and minimizes the use of our basic types of ignorance based on guessing and belief.

Information. Branscomb[25] in a challenging article has called information the "ultimate frontier". In looking ahead he predicted that long before the year 2078 lasers and a single glass fiber smaller than the human hair could move the contents of 40 000 books from the Library of Congress to Los Angeles in 1 h. This remarkable feat still could not guarantee that the right information could be at the right place, at the right time, in the hands of the right person to solve a given R&D problem. However, it is clear that the real *management of information* is a big challenge for the future. Carlson crusaded originally to promote the concept that information is a manageable resource, later did an about-face.[26] He concluded that accountants do not have a technique for evaluating the acquisition and especially the consumption of information (it is not used up!). Horton[27] disagrees and concludes that corporations that excel in the 1980s will be those that manage information as a major resource.

Bits, *bytes*, and *words* in "computerese" have different meanings than one might find in an old dictionary. Data, information, and knowledge may not be bad equivalents for bits, bytes, and words. We need to work harder to help both the accountants and management find a way to express the value of information. Whether by smoke signals or laser/fiber optics, the *message* must get through.

ACKNOWLEDGMENT

I thank Stanley J. Fidelman and Dr. Myra Williams of Merck & Co., Inc., for reviews of the manuscript, Dr. W. V. Metanomski of Chemical Abstracts Service for his helpful editorial comments, and many colleagues, in the Pharmaceutical Industry, who, over a quarter-century period, shared their views on information resource management.

REFERENCES AND NOTES

(1) Borko, H. "Information Science: What Is It?" *Am. Doc.* **1968**, *19*, 3–5.
(2) Miles, S. A. "An Introduction to the Vocabulary of Information Technology". *Tech. Commun.* **1967** (Fall), 20–24. Cited in "Key Papers in Information Science"; Elias, A. W., Ed.; American Society for Information Science: Washington, DC, 1971; p 5.
(3) Farradane, J. "The Nature of Information". *J. Inf. Sci.* **1979**, *1*, 13–17.
(4) Feldman, A.; Holland, D. B.; Jacobus, D. P. "The Automatic Encoding of Chemical Structures". *J. Chem. Doc.* **1963**, *3*, 187–189.
(5) Smith, E. G.; Baker, P. A., Eds. "The Wiswesser Line-Formula Chemical Notation (WLN)", 3rd ed.; Chemical Information Management, Inc.: Cherry Hill, NJ, 1976.
(6) McNulty, P. J.; Smith, R. B., Eds. "Substructure Searching of Large Chemical Files". "Proceedings of the Manufacturing Chemists Association Meeting", Arlington, VA, August, 1977; Manufacturing Chemists Association: Washington, DC, 1977.
(7) Taylor, R. S. "The Interface Between Librarianship and Information Science and Engineering". *Spec. Libr.* **1967**, *58*, 45–48.
(8) Kertesz, F. "The Information Center Concept"; Oak Ridge National Laboratory, Union Carbide Corp., U.S. Atomic Energy Commission Report ORLN-TM-2281, July 2, 1968. In "Key Papers in Information Science"; Elias, A. W., Ed.; American Society for Information Science: Washington, DC, 1971; pp 67–86.
(9) Pharmaceutical Manufacturers Association, 1100 Fifteenth St., N.W., Washington, DC 20005.

(10) Freeman, J. E.; Katz, R. M. "Information Marketing". *Annu. Rev. Inf. Sci. Technol.* **1978**, *13*, 37-59.
(11) Martin, Y. C. "A Practitioner's Perspective of The Role of Quantitative Structure-Activity Analysis in Medicinal Chemistry". *J. Med. Chem.* **1981**, *24*, 229-237.
(12) Brown, H. D.; Costlow, M.; Cutler, F. A., Jr.; Demott, A. N.; Gall, W. B.; Jacobus, D. P.; Miller, C. J. "The Computer-Based Chemical Structure Information Systems of Merck Sharp and Dohme Research Laboratories". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 5-10.
(13) Howe, W. J.; Hagadone, T. R. "Molecular Substructure Searching: Computer Graphics and Query Entry Methodology". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 8-15.
(14) Eggers, I. M. R.; Gall, W. B.; Cutler, F. A., Jr.; Brown, H. D. "Use of Proprietary Biological and Chemical Data at Merck & Co., Inc.". In "Retrieval of Medicinal Chemical Information". *ACS Symp. Ser.* **1978**, *No. 84*, 85-106.
(15) Gund, P.; Andose, J. D.; Rhodes, J. B.; Smith, G. M. "Three-Dimensional Molecular Modeling and Drug Design". *Science (Washington, D.C.)* **1980**, *208*, 1425-1431.
(16) Bernd, C. L. "Chemical Case Report Forms Design—A Key to Clinical Trial Success". *Drug Inf. J.* **1984**, *18*, 3-8.
(17) Schwartz, S.; George, R.; Owens, J.; Edson, J. "A Clinical Database Management System". *Drug Inf. J.* **1984**, *18*, 33-42.
(18) Herman, R. L., Ed. "Drug-Event Association: Perspectives, Methods, and Uses". *Drug Inf. J.* **1984**, *18*, 195-352.
(19) "Corporate Strategies". *Business Week* **1984**, *November 26*, 114-118.
(20) "Merck World 5"; Merck & Co., Inc.: P.O. Box 2000, Rahway, NJ, 1984; No. 5, December, p 5.
(21) Souder, W. E. "Encouraging Entrepreneurship in the Large Corporations". *Res. Manage.* **1981**, *24* (3), 18-22.
(22) Branscomb, L. M. "Improving R&D Productivity: The Federal Role". *Science (Washington, D.C.)* **1983**, *222*, 133-135.
(23) Baillie, A. S. "Management of Risk and Uncertainty". *Res. Manage.* **1980**, *23* (2), 20-24.
(24) Ozog, S. "On the Value of Information". *J. Am. Soc. Inf. Sci.* **1979**, *30*, 310-315.
(25) Branscomb, L. M. "Information: The Ultimate Frontier". *Science (Washington, D.C.)* **1979**, *203*, 143-147.
(26) Carlson, W. M. "Information Is Not a Manageable Resource". *Inf. Manager* **1980**, *2* (2), 6.
(27) Horton, F. W. "Information is a Manageable Resource". *Inf. Rec. Manage.* **1981**, *15* (1), 9.

The Chemical Information System and Spectral Databases†

STEPHEN R. HELLER‡
Silver Spring, Maryland 20902

Received January 1, 1985

From 1970 to 1984, the U.S. Government cooperated with various organizations in the support of the development, maintenance, and distribution of a computer-based chemical information system of spectral and other numeric databases, known as the NIH/EPA Chemical Information System (CIS). This presentation discusses the history of the project and related activities in the area of numeric database activities and summarizes the current state of the project.

INTRODUCTION

A major activity in modern chemistry is the identification of chemical substances from laboratory measurements made on these substances. It was this activity that lead the National Institutes of Health (NIH), in the early 1970s, to initiate an informal project for the identification of chemicals using a computer-based mass spectral database and associated search software.[1] It rapidly became evident that, in addition to mass spectral data, there were other capabilities that were needed in a modern chemical laboratory. These included other spectral and numeric data, coupled with chemical structure search, manipulation, and retrieval capabilities. Having available an excellent computer facility at NIH, which included a large PDP-10 time-sharing interactive computer, some databases, and the interest of a core group of scientists desiring to explore new areas, a chemical information system began to germinate.

In the early 1970s the main interest of the chemical information community was with bibliographic databases. Thus, the efforts of the group at NIH, later joined by scientists at the National Bureau of Standards Office of Standard Reference Data (NBS, OSRD), the Food and Drug Administration (FDA), and others, began to explore a new and potentially promising area. In late 1973 the Environmental Protection Agency (EPA) initiated an expansion of its support activities to EPA laboratories and state and local government laboratories by joining this budding informal cooperative effort to develop a computer system for support of its environmental legislative mandates.

This paper will describe the development of the spectral databases of the chemical information system, called the NIH/EPA Chemical Information System (CIS), of some related spectral database activities, and of some recent events regarding this project. Further details on the development of

Dr. Stephen Heller received a B.S. in chemistry in 1963 at the State University of New York—Stony Brook and a Ph.D. in organic chemistry in 1967 at Georgetown University. He is presently a research leader at the U.S. Department of Agriculture, Agricultural Research Service, in Beltsville, MD. Previously, he has served as a chemist at the U.S. EPA, as a Senior Staff Fellow at the National Institutes of Health (1970-1973), and as a chemist for the U.S. Army (1967-1969). He also served for 6 months with the U.S. House Subcommittee on Health and the Environment (1979-1980) and as a Lady Davis Visiting Professor of Chemistry at the Hebrew University, Jerusalem (1981). He has published over 100 papers and books during the past 20 years.

† This article does not reflect the official views of the EPA.
‡ Address correspondence to the author at the USDA, ARS, BARC-West, Beltsville, MD 20705.

0095-2338/85/1625-0224$01.50/0 © 1985 American Chemical Society