Information Requirements for Chemists in the Pharmaceutical Industry†

HORACE D. BROWN
Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065

Received December 5, 1983

The principal information needs of chemists in the pharmaceutical industry center around a substance, methods used for its characterization, its reactions, its physicochemical and biomedical properties, and its metabolic fate in biological media. Novelty searches for a specific chemical are critical because of the need for patent protection for product candidates. Chemical relatives are important, necessitating substructure searching. Therapeutic neighbors may give insight into its biological mode of action. The phases of drug discovery and development are used to indicate the diversity of the information needs of chemists and to emphasize the dual process of generation and use of information.

INTRODUCTION

The pharmaceutical industry has as its mission the discovery, development, manufacture, and sale of useful, safe therapeutic agents for improving the quality of life in human and animal health and for protection of our environment (air, water, etc.). Therapeutic agents include drugs, vaccines for the prevention of diseases, and diagnostic agents (chemical, photo, electro, and mechanical). Since I was asked to talk about information requirements of chemists, and since that is my background, I am going to largely limit my remarks to drugs (chemicals) and, hence, vaccines, diagnostics, etc., will be slighted—though many of the same principles regarding information needs would apply there equally well.

One of the major functions of a chemist in the pharmaceutical industry is data/information/knowledge management. The data base will be derived from internal/proprietary/published literature and patent information. The chemist is concerned with chemical substances, their synthesis, and physicochemical and biomedical properties, including absorption, transport, and metabolic fate. Toxicological properties, both acute and chronic, are crucial.

A cursory examination of the major literature search requests made to one of our Merck information groups in 1982 indicated the following: (1) of the 368 questions examined, 269 of them expressed an interest in a substance (or family); (2) 117 requested physicochemical or biomedical properties; (3) 239 were interested in a therapeutic category or related biochemical, biological, or medical aspect.

Chemistry has the best data/information base of the many disciplines required in the pharmaceutical industry. This may stem in part from the nature of the science, which tends to be exact, reproducible, a blend of hypothesis, theory, and practice, and based on a long history. However, I believe no small part of its availability stems from an early recognition by chemists of the value of reviews, compilations, and compendia. *Chemical Abstracts*, prepared in the early days largely by volunteers, was carefully edited according to rigorous schemes. These, coupled with the CAS Registry Number system for compounds and their recent structure/substructure search schemes, provide powerful links between a substance, its characterization, its synthesis, its properties, and its owner.

As we shall see later, these become useful facets of the data/information/knowledge requirements for a chemist engaged in these activities. Thus, thanks are due to those individuals who have contributed time and talent to the organization and dissemination of chemical information from Columbus, OH, for the better part of this century.

Definitions. "Information" is generally used in a very generic sense. I like to dissect the term into three facets. These are data (quantitative measurement), information (evaluated data), and knowledge (integrated information). Some of the milestones in information science[1] and the major indexing classification schemes[2] have been reviewed.

Sources. The sources of data/information/knowledge for a chemist in the pharmaceutical industry are not that different from those in other activities. The major inputs are from published or proprietary sources. The relative importance of these sources varies with the stage of a project, but both types are essential.

It is important to say a word about the commercial bibliographic data bases. Though I do not know the relative volumes of sales by the vendors to particular industries, my impression is that the pharmaceutical industry is near the top. Thus, a word of commendation is due the early members of the PMA's Literature Subcommittee and the several data base builders, who together (along with the chemical industry) launched what is today a sizable industry in its own right. Without the availability of such services, our information problems would be considerably more complex. Through the valuable help of information scientists, our chemists are able to receive program/project/personal profiles of information tailored to specified needs.

PHASES OF DRUG DISCOVERY/DEVELOPMENT

I am going to group my remarks around some arbitrary (classification) steps to give an indication of the progression that occurs in the birth of a new drug for human or animal health as seen through the eyes of a chemist. Fifty or more distinct disciplines are frequently required to bring a safe, effective drug to market—hence, the entire story will not be told. Chemists are a focal point in the discovery, development, and manufacture of new drugs.

The phases that I will use to outline the information needs of a chemist in the birth of a new drug are discovery, lead development, product development, and market support. The average research and development expenditure (including capital) for introduction of a new drug in the U.S. is approximately $70 million. From initiation of a research project to marketing of a new product may take 15 years.[3]

By a cursory description of these phases and the varied activities of the chemists, we can infer the major proprietary and published information needs relevant in that phase. Both generation and usage of information are essential at each step.

Discovery. Research programs and projects within the pharmaceutical industry are usually approved after one or more of the following events: (1) perceived need for new or better therapy for a particular disease, (2) increased incidence of a particular disease in the population of concern, (3) new idea for a different method of detecting useful agents for a new or established field or therapy, (4) new pharmacologic or biochemical rationale related to the cause of a given disease,

† Presented in part at the Herman Skolnik Award Symposium on Indexing Chemical Information before the Division of Chemical Information, 186th National Meeting of the American Chemical Society, Washington, DC, Aug 30, 1983.

0095-2338/84/1624-0155$01.50/0

and (5) glowing success by a competitor with a new wonder drug, etc. With the approval of the project comes a well-defined objective (e.g. "a novel, safe, effective, orally active agent for the treatment of schizophrenia").

As the project is launched, specific assignments of personnel are made. It is important that the "minimal critical mass" of individuals representing the required disciplines be committed. In the discovery phase, a minimum requirement would usually be chemists and biologists (since our focus is on chemists/chemistry, we will minimize the description of the other activities).

The chemist begins by a broad state-of-the-art review of the disease in question—incidence, suspected causes, current therapy, and drugs used (advantages and disadvantages). Of particular concern are the structural/substructural features of any compounds known to be active against the condition, causative organism, etc. Suspected modes of action of useful drugs may suggest new ideas.

The biologist, meanwhile, is responsible for designing/selecting new assays that will be used for sorting out the active vs. inactive candidates. Such tests frequently begin with simplified in vitro cell culture or related types and later move to laboratory animal assays.

If no structures active in the bioassay are known, chemists frequently resort to selecting a wide spectrum of compounds; antimetabolites, antagonists, natural products, or even random structural types ("screening") are submitted for test. The major goal is to find some structures that "ring the bell". From this effort (or from the literature in case of established chemotherapeutic classes), the chemist must let his imagination loose, turn on his creativity spigot, or begin to draw analogies based on the structural features present in the active molecules.

Discovery strategy may dictate testing of microbial broths as a source for new leads. Here, the microbiologists collect cultures from soil or other sources, grow them in tubes or shake flasks, and submit the fermentation broths (treated in standard fashion) for test. Should actives be found, the chemist now has an isolation/characterization/structure determination problem. A whole set of parameters (profiles) needs to be compared to rule out previously known agents: UV spectra, column running rates, antibiotic activity (including known resistant strains of organisms), or any combination of physicochemical or biological properties. Books/compilations or other literature sources are crucial at this stage. If the conviction persists that the agent is new, then larger quantities are produced in the fermentation. Isolation (using physicochemical properties or antimicrobial activity to guide), purification (chromatography), and then structure identification follow. Mass spectrometry is a priceless tool. Again, compilations of data for comparison of the mass fragments are useful.

Piecing together of the "jigsaw" pieces leads to proposed structures and then to the literature for this family of structures (if known) or near relatives if not known. Interest permitting, the structure needs to be synthesized: hence, a need to analyze potential routes to the structure, assemble required starting materials, etc. We should note at this point that *practical* synthesis is not required; the shortest, most certain route is chosen to give the desired compound. This comment would apply as well to any compounds synthesized as *new* agents for preliminary biological testing. The goal is to get the required quantity of material into the biologist's hands for evaluation—if it is inactive, that is the end of that dream.

Recently, medicinal chemists have begun to use computer technology as a tool for drug design.[4] Molecular modeling—especially 3D in color—brings an enhanced perspective over the classical stick and ball models for viewing chemical structures. If one couples the ability to compute preferred confirmations, then a new dimension is added. Where enzyme inhibition is envisioned as a principal target, and if perchance the enzyme (or a near relative) has been crystallized and defined by crystallographers, one may dock known or proposed "inhibitors" into the active site (cleft) of the enzyme image. This exercise in the hands of the expert brings opportunity for exciting chemist–machine interactions. As the file of available enzyme/protein/receptor structures grows and as theoretical chemists are able to improve their tables of parameters, this will become a routine part of drug design activities in the future. Thus, index descriptors for software packages (e.g. MM_2), parameter sets (amides and cyclic lactams), energy minimization values, and inhibitor/enzyme interactions will need to be in CA for use by medicinal chemists.

A crucial aspect of chemists' information needs during the discovery phase is interdiscplinary knowledge. Understanding of the biomedical/biochemical aspects of the target disease and cross-fertilization of ideas and concepts with the interdisciplinary team members are useful.

Lead Development. Once the chemist/biologist team has found a lead structure with the desired qualitative biological response, the next assignment is to explore the structure–activity relationships (SAR).[5,6] To minimize the effort required, a systematic approach is needed. Electronic, steric, lipophilic and isosteric matters must be studied to determine the breadth and depth of the SAR. Here, careful work by both disciplines (chemistry and biology) is essential. Most crucial is the reliability/reproducibility of the bioassay and routine use of standard substances for comparisons. Both *negative* and positive assay results are important.

Proprietary chemical files are searched on a structure/substructure basis. Usually, the reports are accompanied by the biological activities (including toxicity) known for the compounds. The recent availability of substructure search from published data bases is of considerable interest to our chemists.

The chemist has three major concerns in the lead development phase: efficacy, safety (therapeutic index), and *novelty*. Hence, in addition to planning the synthetic agenda, he/she must be looking into chemical, pharmacological, toxicological, and patent literature for chemical relatives of the lead. The biologist is busy exploring dose–response curves ("firepower" of the candidates), exploring type and degree of secondary (side) effects, and defining the novel attributes of the compounds over existing standards. Routes of administration (oral, im, iv, topical) and interactions with other drugs likely to be encountered in therapy must be considered. If the agent is chemotherapeutic, genus, species, and strain of causative agents must be examined. Activity, or lack of it, against resistant strains is always of concern.

Thus, at this stage the team is ranging across disciplines (chemistry, biology, toxicology, medicine) to track a substance or family and its physicochemical and biomedical properties. In addition, they may utilize computer science, crystallography, and statistics.

The major focus of this exercise is to select a product candidate for recommendation for development. To survive the scrutiny to which it will be submitted, it must be novel, effective for the therapy proposed (usually with a significant enhancement over existing therapy), and safe under conditions of proposed usage—with a definition of the expected side effects. In addition (because of the $50–100 million price tag for development), it must either have a potential market or be of sufficient interest clinically to meet the criteria for "orphan drug" class. Finally, if it is to be introduced into human or animal medicine, it must pass the scrutiny of the appropriate regulatory bodies (Investigational New Drug or

New Drug Application, IND/NDA). Patent applications for all of the countries where marketing is considered must be filed and prosecuted.

When the two or three product candidates have been selected, a preparation laboratory chemist enters the picture to prepare additional quantities for preliminary toxicity studies and to permit the pharmacists to start stability and formulation (dosage-form preparation) studies. Though the initial steps may simply amount to scaling up the research chemists' recipe, serious thought must now be given to the economics and to the practicality of producing the substance in quantity.

With the availability of more material, the biologists/pharmacologists will begin to measure bioavailability (percentage of material absorbed by various routes), transport (is it bound to protein or plated out in tissue?), and excretion patterns (urine, feces). In addition, metabolic fate—hydrolysis, esterification, methylation, acylation, oxidation, reduction—needs to be studied. This frequently requires radiolabeling—which in turn means that the labeled molecule must be synthesized by radiochemists under Atomic Energy Commission guidelines for safety. The identity of the metabolites must also be determined. In the instance of drugs intended for veterinary medicine, the extent and duration of residues retained in the edible carcass must be measured down to the limit of detection (parts per million, billion, or trillion?).

A continuing concern, as the selection process for product candidate unfolds, is the patent status for the candidate, its functional derivatives, and close analogues. This means a diligent watch for the literature on such compounds and particular care for foreign patent applications.

Once the candidate and its backups are selected, the pharmacists begin to examine it for suitability for dosage-form preparation. Oral administration (tablet, capsule, syrup) is preferred, but almost always an *intravenous* formulation will be required—especially if the agent is to be used for chemotherapeutic or cardiovascular therapy (where crisis-type situations may be anticipated). For the pharmacist, existing literature on such compounds, known methods of assay, stability to heat and light, crystalline forms, degree of hydration, and likely contaminants (precursors in synthesis) all must be studied. Solubility, particle size, and compatibility with granulating materials must be resolved. Test formulations are studied for absorption and for toxicity tests by safety assessment toxicologists, as well as for physicochemical process controls. The above steps (or equivalent) are a necessary "evil" to permit the selection of the single candidate for recommendation.

One of the major exercises that must go on in this transition from research to development is *information transfer*. A new team is about to take charge of the new fledgling. Some of the early stages in development will repeat the above-described exercises but will be done in much more intense fashion and with an aim toward economic practicality, rather than data/information generation and evaluation.

Product Development. Once an exciting (i.e. effective therapy, good therapeutic index, and patentable) product candidate has been found, management and regulatory requirements must be met before a full-scale developmental program can be launched. Interdisciplinary/interdivisional committees and line organizations review the proposals. In addition, regulatory agencies must be apprised of plans to go into clinical studies and, finally, must approve its New Drug Application. Interim processes must be developed to produce the kilogram quantities required for safety testing in several species of animals over 2–3-year spans. Hundreds to thousands of patients must be treated with the new drug—some in double-blind comparison with the best therapy currently available.

Since our focus is on the chemist, we will look in a little more detail at the types of information he needs in the developmental steps. The first assignment will usually be for a few hundred grams (depending on toxicity—less toxic means more drug) of material for safety tests in acute–subacute studies. Unless the laboratory synthesis was pretty awful, this will usually be made by a simple "blowup" of the initial synthetic scheme.

However, a seasoned process chemist is quickly assigned to consider alternate routes. Cost and commercial availability of starting materials, yields, unusual equipment (pressure, high temperature), cycle times, labor requirements, solvents (recovery feasible?), and a host of other parameters must be considered for each alternative reaction pathway envisioned. Thus, the best details available in the literature on the reactants, products, and reaction conditions must be found. Once again, process patents will be crucial. In some countries, product patents are not issued and, hence, a patentable process should be developed.

The general need for process chemists to have access to reaction retrieval reminds me of a current deficiency in reaction retrieval capabilities. Computer systems (e.g., Molecular Design Ltd.) are now available with the capability to electronically perceive, capture, and store chemical structure information for reactants and products, along with descriptive information relative to reaction conditions, yields, limitations, etc. However, current reaction libraries are inadequate. Derwent Publications Ltd. (through Systems Development Corporation, Orbit IV) has a Chemical Reactions Documentation Service (CRDS) which, together with Volumes 1–20 of Theilheimer's *Synthetic Methods of Organic Chemistry*, gives a total of about 50 000 reactions. Retrieval is available through keywording and coding. However, the usual limitations of fragment coding apply here, and use by other than an experienced intermediary is frustrating at best. *Chemical Abstracts* indexing of organic reactions is quite general and is heavily linked to nomenclature. The Institute for Scientific Information has Current Chemical Reactions (CCR), but it is coded in Wiswesser Line Notation. Hence, there is need for a modern, computer graphics-based system using chemical connection tables with structure/substructure query, supplemented by mechanistic and other appropriate descriptors. Hopefully, data base builders and computer systems developers can manage the necessary compromises to bring this service to synthetic chemists.

Process chemists have to be concerned about what is left over after the reaction sequence leading to the new product is complete. Sometimes this is a matter of sheer economics. For example, in the early days of the manufacture of methyldopa (Aldomet), the resolution of the racemic mixture left half of the penultimate product as nonuseful. Since the daily dosage of Aldomet for hypertensives is 0.5–1.0 g/day, this meant that lots of expensive material was accumulating. Eventually, this led to a racemization, resolution and repeat of the cycle. Sometimes, the residuals are toxic and must be disposed of in accord with EPA/OSHA guidelines. Detection methods in effluent streams, chemical destruction (oxidation, reduction), combustion, or other methods may be required. Here, the chemist/engineer team must extract the necessary analytical or physical chemistry from the literature or derive new schemes.

Severe residue problems face the chemist in the veterinary product area. From the drug application (protection of workers) to the detection of residues on hides/wool, in edible carcass, in milk, and in eggs to the question of buildup in farm streams or ponds, the imagination, ingenuity of the chemist and literature resources are all taxed. Both laboratory (sophisticated equipment) and field test operations need to be included in the derivation of assays. Careful match to past

precedents from regulatory practice and from relevant literature is needed.

Market Support. Once the new product is launched, development chemists need to track market estimates of the projected material requirements. In the background, a team keeps their gray matter active to look for shorter routes, less expensive starting materials, and fine tuning of current production reactions, solvent recovery schemes, and waste disposal. These are important for reasons of cost containment as the rate of production goes up. For the longer pull, they anticipate that competition will follow with the possibility of price reductions. Both proprietary and published documents must be scrutinized to keep the product on course and to plan for the future.

During the market support phase, the discovery team is following the fate of the new product. How well did the laboratory methodology translate into human or animal medicine? Did the side effects noted in animal studies carry over to the clinical or field trials? Were there new ones? For chemotherapeutic agents, resistance development under hospital/field conditions must be monitored. This could mean a "back to the drawing boards" exercise. Coccidiosis and helminthiasis are well-known examples in the animal health area where resistance development is common. Competitive activities, as witnessed by foreign or U.S. patents, communications, meeting papers, etc., must all be monitored carefully by the discovery chemists.

Once a product is launched, there is need for publications describing the new drug. These will take the form of discipline-oriented papers (chemistry, biology, pharmacy, medicine) or may be published as interdisciplinary articles by the team. Here, chemists must integrate the data/information/knowledge generated from the project with that from the published literature.

SUMMARY

We have taken a cursory look at the several phases (discovery, lead development, product development, and market support) for the birth of a new drug for human and animal health. We have looked at a number of information needs of the chemists (organic, analytical, process, etc.). We have noted that information management is a major part of their daily lives. Both proprietary and published data, information, and knowledge must be collected, evaluated, and applied to the substances of interest. We have reviewed both the classical (screening) and newer trends, e.g., rational drug design using computer technology, as a means of discovering new chemicals with interesting biological properties. In short, chemists need access, preferably through structural and substructural pointers, to a chemical, its family relatives, its characterization, its synthesis, and its physicochemical and its biomedical properties. Also of keen interest is its metabolic fate and its propensity for buildup in biosystems, as well as its hazards for those who manufacture it or apply it in its clinical or veterinary role. We have noted a chance for improvement in the use of modern computer technology to organic reaction retrieval. On balance, the chemist in the pharmaceutical industry is well served through the available commercial data bases and information intermediaries.

An item causing serious concern now and likely to get worse is *information overload*. Though some dedicated chemists are trying to work smarter, others are throwing up their hands. Possible solutions include better definition of questions by the chemists, tougher "negotiations" by the information scientists, and increased use of computer-based systems for personal data/information management. We also need to provide user-friendly, direct-access methods based on structure/substructure for our chemists.

Finally, it is important that scientists publish their findings as data, information, or knowledge in the appropriate disciplinary or interdisciplinary journals. The chemist has drawn on the resources shared by others. He/she must in turn be willing to generate data, convert it to information by careful evaluation, and, finally, integrate it into the growing store of scientific knowledge. Through such activities the science of chemistry is perpetuated, and the information needs of future generations of chemists may be satisfied.

REFERENCES AND NOTES

(1) Skolnik, H. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 187–193.
(2) Conrad, C. C. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 197–201.
(3) Seltzer, R. J. *Chem. Eng. News* **1983**, *Aug 8*, 16.
(4) Gund, P.; Andose, J. D.; Rhodes, J. B.; Smith, G. M. *Science (Washington, D.C.)* **1980**, *208*, 1425–1431.
(5) Hansch, C.; Rockwell, S. D.; Jow, P. Y. C.; Staller, E. E. *J. Med. Chem.* **1977**, *20*, 304.
(6) Martin, Y. C. "Quantitative Drug Design"; Marcel Dekker: New York, 1978.