Transactions of the Illinois Academy of Science
(1987), Volume 80, 3 and 4, pp. 203-206

COMPUTER AIDED SELECTION OF ACTIVE THERAPEUTIC NATURAL PRODUCTS

C.B. Lewandowski, M.L. Quinn, and N.R. Farnsworth [1]
Program for Collaborative Research in the Pharmaceutical Sciences
College of Pharmacy
University of Illinois at Chicago
Chicago, IL 60612, USA

ABSTRACT

NAPRALERT, a specialized data base, can provide researchers with preliminary ethnomedical and scientific data to aid in selecting natural products with a potential for yielding useful biologically active agents. The data base is a compilation of published papers, coded and sorted for retrieval by a sophisticated computer system. A researcher can generate either a detailed profile or a simple bibliography based on interest or need. Examples of applications in various areas of medical research include use by the National Cancer Institute, the World Health Organization, UNESCO and others.

Mankind's development of medical science is richly laced with anecdotes of the "healing" potions of folklore becoming today's commonly accepted remedies. The Peruvian *Cinchona* bark drink, used by Indians as a treatment for fevers (Wheelwright, 1974), led to the discovery of quinine and quinidine, two widely used therapeutic agents. More recently, researchers reviewed 15th century literature and discovered a sweet plant that yielded hernandulcin (Compadre et al., 1985), a sweetening agent of great potential. There are literally hundreds of other examples.

Of the estimated 750,000 higher plants existing, a large number have been used in traditional medicine, and few of these have been scientifically evaluated. With the estimated loss of over 1,000 species per year due to destruction of habitat, it is most urgent to identify and protect those species which have the potential of supplying the world with new therapeutic agents.

[1]Presented at the 78th annual Illinois State Academy of Science meeting, April, 1985.

Traditional methods of identifying natural products with useful biological activity (phytochemical screening, pharmacological screening and ethnomedical reputation) are expensive, time consuming, and each has drawbacks. In today's world, researchers cannot afford the time or funds to pursue all these methods extensively for each species.

If a researcher could base selection of natural products to be screened on a solid search of ethnomedical and scientific material, the probability of isolating active therapeutic agents should be greatly enhanced. At the Program for Collaborative Research in the Pharmaceutical Sciences (PCRPS), University of Illinois at Chicago, a specialized data base has been developed to meet this need.

NAPRALERT, an acronym for NAtural PRoducts ALERT, is a specialized computer data base containing the world literature on natural product chemistry and pharmacology.

Material for entry in the NAPRALERT system is collected from thousands of current scientific journals that have been found to contain a majority of pertinent natural product information. International contacts and research activities have allowed access to important, but normally unavailable journals from many foreign countries. Articles containing information on all organism classes, the presence of secondary chemical constituents in natural products, or results of testing natural products for any type of biological activity are regularly included, as are folkloric and ethnomedical information. In addition to current literature (from 1975), retrospective literature searches on 2,000 individual topics (plant and animal genera, pure compounds and specific pharmacological activities) have been carried out and the data computerized. Some records from 1650 and earlier are contained in NAPRALERT.

Current data compiled from scientific articles, books, symposia, and abstracts are stored in four distinct record types - demographic, taxonomic, chemical, and pharmacologic. These record types are further classified under 70 field names or categories. Such a comprehensive information storage system allows for retrieval of even the most detailed, specialized format.

Selected articles are prepared for data entry by a four step process:

Step 1 - Demographic record - includes the full title, all authors, journal information, language in which the article is published, author address, secondary source information and other miscellaneous information. Each article also receives a unique NAPRALERT citation number.

Step 2 - Organism record - the taxonomic record includes organism class, family, genus, species, species authority, subspecies, subspecies authority, common name, taxon synonym, organism part studied, condition of part, and the country and geographic area where the organism was collected.

Step 3 - Chemical record - provides the compound name, method by which the compound was isolated and/or identified, and the compound code. The compound code is a unique alphanumeric NAPRALERT code which describes the chemical class, carbon skeleton, and functional groups present.

Step 4 - Pharmacology record - indicates the type of work (*in vitro*, *in vivo*, *in situ*, in human) pharmacological activity tested (over 2000 different activities are coded ranging from general system to specific biochemical effects), special experimental designs and modifications, type of extract used, mode of administration, animal model, if any, dose and dosing schedule, qualitative and quantitative results, and substrate employed. In addition, when pure compounds are isolated, the yield of each compound is also coded.

Individual records are tied together for each entry through the unique citation number assigned to each article, and an "occurrence" number is created for each record type. For rapid retrieval, data are stored using the indexed sequential access method (ISAM).

Each step in the coding process is supervised and reviewed by qualified experts on staff in the NAPRALERT program. Coded data are checked for accuracy and then entered in the data base. For example, the accuracy of natural product identification given in an article is verified by one of the taxonomists on staff. When a plant has been labeled with several Latin binomials or identified by more than one authority, a botanist will code the data indicating the recognized name and authority. Synonymous names are cross-referenced allowing data retrieval given a taxon synonym. Computerized data are proofread before posting to the main NAPRALERT system. There is also a proofreading phase built into the pre-posting process which enables the computer to call attention to possible coding and/or typographical errors and allows for a high degree of integrity in the data base.

Original articles on which the coding is based, are kept on file and can be referred to should questions arise regarding some part of the computerized information. Copies of articles can be provided, at cost, to users who request this service.

Formatted information on almost any subject concerning natural products can be retrieved, but the three most common formats requested by users are the ethnomedical, pharmacological and chemical profiles. An ethnomedical profile provides full organism information, including genus and species selected, synonymous names which alert the user to the fact of their existence, family name, organism part and its condition, and the geographic location of the sample. Common names (in several languages), and ethnomedical uses are also reported. A complete list of literature citations provides the user with the NAPRALERT citation number and a complete bibliography.

A pharmacological profile lists the full organism information mentioned above, type of pharmacologic activity tested, type of extract tested, mode of administration, animal model or substrate/tissue/cell/tumor system, dose employed, and qualitative and quantitative results. A list of literature citations provides the user with the complete bibliography,

The chemical profile details all secondary constituents reported to be present in or isolated from the plant, microbe, or animal in question, and provides the name of the constituent, major chemical class, part of the organism containing the constituent, percentage yield of constituent (if known), and geographic location of the sample. A complete NAPRALERT bibliography is again included.

There are a wide range of applications for a data base such as NAPRALERT, several examples of use in identifying active therapeutic natural products follow.

1). The World Health Organization (WHO) a)The Task Force on Plants for Fertility Regulation has used the NAPRALERT facilities to systematically search and analyze folkloric and scientific information using a point system based on several criteria to select the most promising indigenous plants useful for fertility regulation; b) The Task Force on Tropical Diseases has conducted a study of natural products which have several activities and may be used in preventing the spread of diseases such as malaria, leprosy, filariasis, and schistosomiasis in developing countries; c) The Traditional Medicine Program has sponsored numerous inquires for information to scientists interested in collecting and studying the practices of Traditional Medicine in Africa, South America, China, India, the U.S.A., and other countries. The development of useful inexpensive medicines to provide "health for all" by the year 2000 is the goal of WHO.

2). UNESCO — in an effort to make available to scientists in developing countries information on plants with potential therapeutic value, UNESCO has used NAPRALERT to compile all available folkloric and published information on monocots, dicots and gymnosperms. This project is in its final stages and should result in a reference source of world-wide importance.

3). National Cancer Institute — Since 1978, a regular surveillance of current literature for newly identified compounds and very current experimental data on extracts and pure compounds having antitumor, cytotoxic, and/or related activities has been conducted by the NAPRALERT system.

4). King Abdulaziz University has contracted with NAPRALERT for a complete book on the 600-700 plants used medicinally in Saudi Arabia and South Yemen. This book will contain full taxonomic descriptions, ethnomedical uses and biological and chemical activities for all native or imported plants known to be used in the region.

5). Other Users – requests from a variety of other users have been received including FDA, The London Times, several herbal companies, and individual researchers. These requests generally ask for complete information on specific plants, compounds or activities. The information provided has been used for television shows, popular articles, books, development of consumer products, research papers, and scientific presentations.

Past and current use indicates that NAPRALERT fulfills a vital role in the search for active therapeutic natural products.

ACKNOWLEDGEMENTS

The NAPRALERT data base is supported by grant BSR 8311297 from the National Science Foundation and from University of Illinois funds.

REFERENCES

Compadre, C.M., Pezzuto, J.M., Kinghorn, A.D. 1985. "Hernandulcin: An intensely sweet compound discovered by review of ancient literature." *Science* 227:417-419.

Farnsworth, N.R., Loub, W.D., Soejarto, D.D., Cordell, G.A., Quinn, M.L. and Mulholland, K. 1981. "Computer services for research on plants for fertility regulation" *Korean J. Pharmacognosy* 12:98-110.

Loub, W.D., Farnsworth, N.R., Soejarto, D.D. and Quinn, M.L. 1985. "NAPRALERT: computer handling of natural product research data." *J. Chem. Inf. Computer Sci.* 25(2):99-103.

Wheelwright, E.G. 1974. *Medicinal plants and their history.* Dover Publications, Inc. New York.