Chemistry by Computer

Test tubes and flasks no longer monopolize chemistry; computers have invaded the laboratory

BY LINDA GARMON

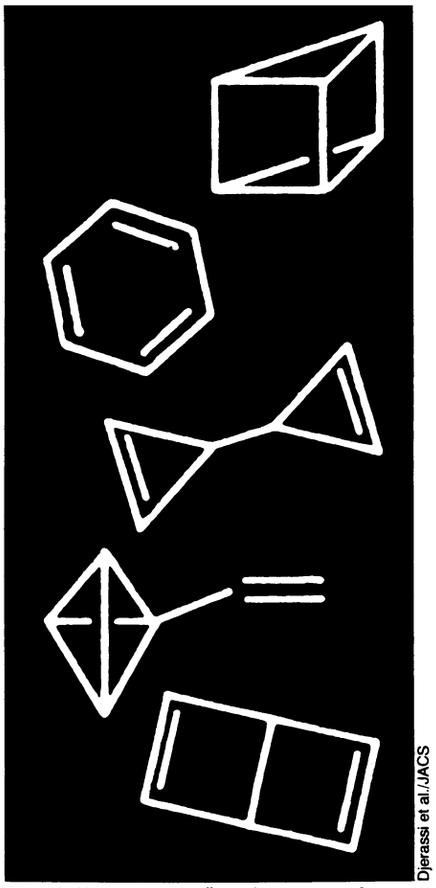
Like a pilot in a cockpit, Robert Langridge has an array of switches, nobs and joysticks at his command. This University of California chemist does not, however, manipulate these controls to guide the flight of an airplane; instead, he uses them to rotate models of molecules displayed on a computer screen.

The three-dimensional display of model molecules is the result of a computer program written by Michael L. Connolly and developed further by Langridge, both of the University's Computer Graphics Laboratory in San Francisco. The program, described in the Feb. 13 SCIENCE, is just one of the growing number of systems designed to do chemistry by computer.

The explosion of computer chemistry programs is due partly to laboratory economics. "Research is expensive," says computer chemist George T. S. Wolken Jr. of Battelle Memorial Institute in Columbus, Ohio, in the October INDUSTRIAL RESEARCH & DEVELOPMENT. "Manipulation of computers instead of chemical molecules can save valuable time and remove the necessity for making and testing of materials of questionable value." Because computer costs decrease annually about 30 percent, and because days of computer-assisted planning can save weeks of expensive laboratory trial and error, more chemists are turning to computers to save research dollars, Wolken explains.

But the computer's laboratory popularity is a consequence of more than dollars and cents; it also stems from the fact that computers can take chemists beyond the realm of bench-top experimentation. Using Langridge's computer graphics program, for example, chemists can "see" as never before how the thyroid hormone thyroxine fits onto its carrier protein prealbumin. By studying such molecular interactions, researchers hope to better understand the body's chemical feedback mechanisms — systems of two opposing but interacting "forces" whose balance depends on how much of the appropriate stimulator or inhibitor chemicals are present. Since the efficacy of a drug in part depends on its ability to interact with a specific receptor molecule, visualizing molecular interactions in three dimensions also has implications for drug design.

Previously the three-dimensional display of molecules was limited to wire structures and space-filling models that use spheres to represent atoms. These simpler systems, though, become mechanically impossible to use beyond a certain degree of molecular complexity. Langridge's computer graphics program, on the other hand, involves no wires that can bend out of shape or spheres that can detach; moreover, extremely complex molecules that would take a roomful of spheres and wires to represent are "sliced" into easy-to-view sections. With the flick of a switch or a shift of the joystick, chemists instantly can select a different "slice" of the model or rotate its bonds to view the display from a different angle.



The C_6H_6 "unknown" with a type of symmetry, at least three equivalent CH bonds and no unusually strained bonds is one of these structures, according to CONGEN.

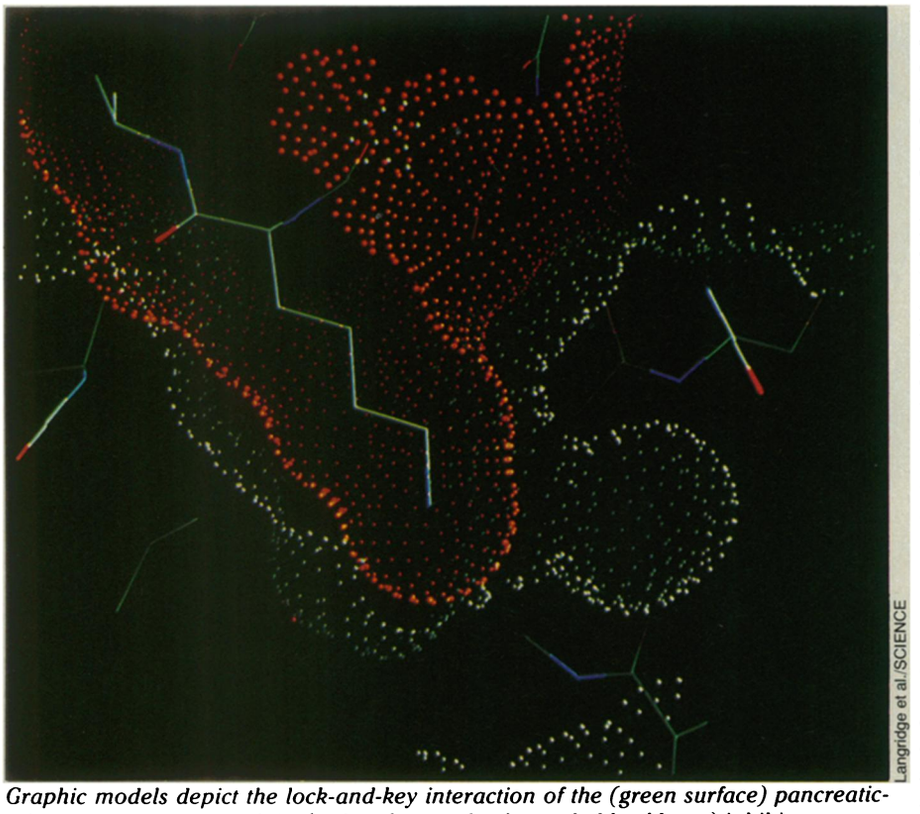
Now, in addition to controlling "slice" selection and rotation, chemists also can choose from a wheel of 64 hues to color-code their graphics displays. "A new command added to the list of display instructions sets the hues and saturation values so that different portions of the picture may be arbitrarily displayed in different colors," Langridge says. "The hue controls the relative amounts of red, green and blue primary colors in the picture, while different saturation levels are used to create 'pastels.'" Because the use of colors accentuates the depth and intermolecular contacts, Langridge describes his color addition as a "new dimension."

Color also enhances the information content of the recently reported microscopic image digital acquisition system (MIDAS) — a process that applies computerized digital imaging to chemical problems. Digital imaging analysis is the science of extracting information regarding the location and amount of specific components of a structure from pictures composed of thousands of grid-like elements, or pixels. "Imaging is here to stay; it's going great guns," says analytical chemist George H. Morrison of Cornell University in Ithaca, N.Y. "Gaining its initial impetus from the work of the Mariner Mission and, most recently, demonstrating its power in Voyager's fly by Saturn, it is routinely used in satellite reconnaissance," Morrison reports in the January ANALYTICAL CHEMISTRY.

In the December issue of that same journal, Morrison and colleagues explain MIDAS, or their use of digital imaging on the microscopic scale. MIDAS is designed to analyze images of a variety of solid surfaces, ranging from biological tissues to semiconductor materials. The sample surface first is blasted with an ion gun. The bombarding ions excite the top layer of atoms on the sample, causing them to "sputter off," Morrison explains. An ion micrograph, or ion photograph, of the sputtered layer is taken. While various conventional analytical techniques then can identify which elements are present on the sample, "now, more and more, people want numbers," says Morrison, so the next step of MIDAS is to *quantify* the *qualitative* information.

To quantify the information, Morrison and colleagues must digitize the ion image, converting each square micron (one-millionth meter) into a pixel. Quantitative information then more easily can be extracted from the resulting computer-digitized picture that differentiates features one micron apart. Morrison and co-workers have used such digital images to identify and locate "unknown" contaminants in semiconductor material.

Another computer program that aids in the identification of "unknowns" is CON-



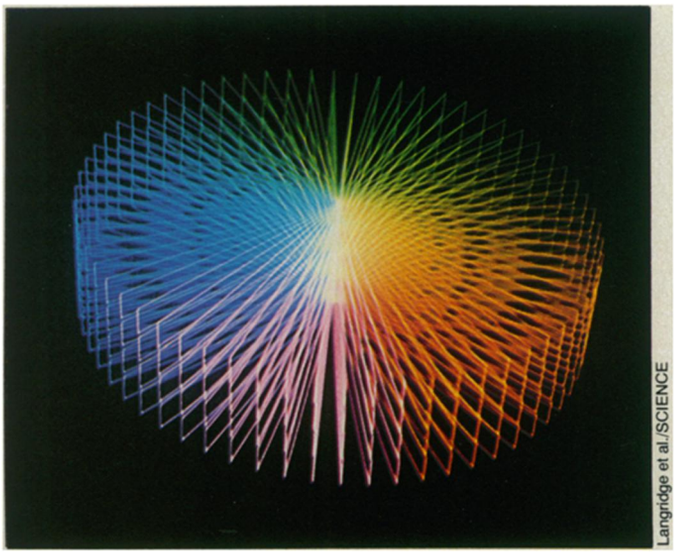
Graphic models depict the lock-and-key interaction of the (green surface) pancreatic-juice enzyme trypsin with its (red surface and color-coded backbone) inhibitor.

terials and various conditions. Computer chemists communicate with the CAMEO computer simply by drawing the starting materials onto a "computer tablet" with an electrostatic pen. "Input is so closely aligned with drawing chemical structures on a piece of paper that little instruction is required for one to attain proficiency," report CAMEO developers Timothy D. Salatin and William L. Jorgensen of Purdue University in West Lafayette, Ind., in THE JOURNAL OF ORGANIC CHEMISTRY (Vol. 45, No. 11, 1980). The CAMEO computer then looks at these structures, but before it does any chemistry, it recognizes the types of atoms and bonds and especially reactive centers present on the starting molecules. Details of this forward-moving computer chemistry program are described in a paper in press for THE JOURNAL OF CHEMICAL INFORMATION AND COMPUTER SCIENCE.

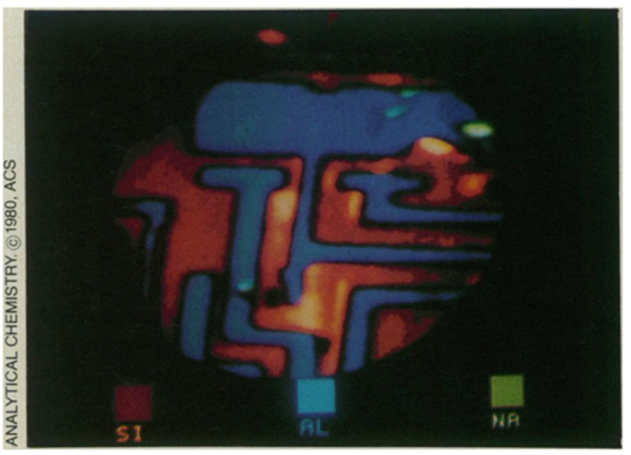
Moving backwards in a synthetic scheme is the focus of SECS — a computer program for the simulation and evaluation of chemical synthesis. While SECS operates in the opposite direction, it is like CAMEO in that it "is a tool to assist chemists in designing an organic synthesis and in determining what reactions or starting materials are appropriate," says the pioneer developer of SECS, W. Todd

GEN — a program now under the direction of Dennis H. Smith and Carl Djerassi of Stanford University in Palo Alto, Calif. CONGEN helps chemists determine the molecular structure of unknown compounds through the *con*strained *gen*eration of isomers, or compounds that have identical kinds and numbers of atoms but different structural arrangements of those atoms. In this program, operators give the computer the unknown's molecular formula, derived from analytical instruments such as mass and nuclear magnetic resonance spectrometers, and then ask it to list the possible structures. The computer is programed with certain constraints so that it rules out energetically or structurally undesirable possibilities. For instance, if a chemist asks CONGEN about the molecular formula C_6H_6, the computer answers by showing the 217 different structures possible, including the familiar benzene ring. But if the chemist restricts CONGEN's generation of isomers by asking for only those C_6H_6 structures that include a particular type of symmetry, at least three equivalent CH groups and no unusually strained bonds, then the computer shows or predicts only five different structures.

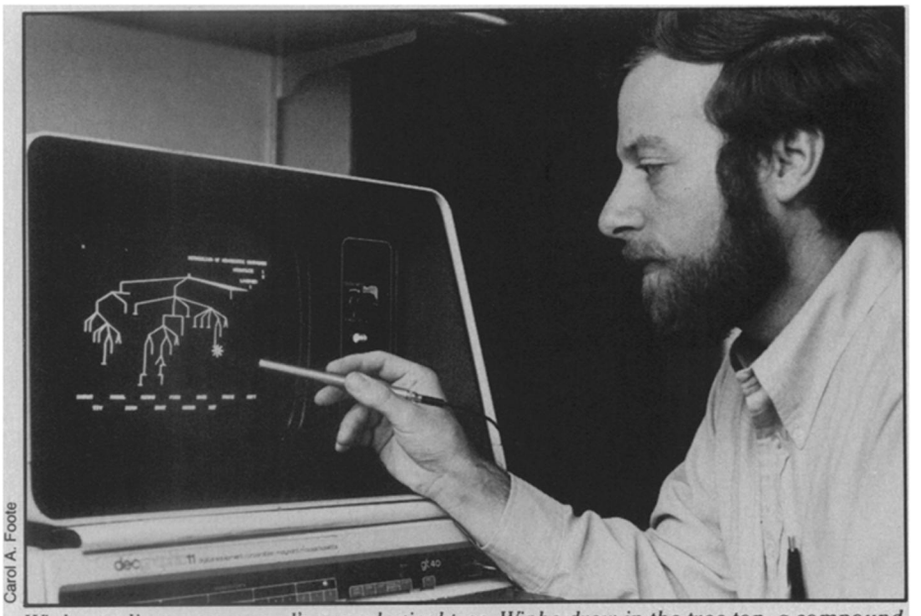
Other computer chemistry programs are being developed to predict whether individual chemical reactions in a proposed synthetic scheme are feasible. CAMEO, for instance, named for the computer-assisted mechanistic evaluation of organic reactions, now is being formulated to predict the products of organic (involving carbon-containing compounds) reactions given the starting ma-



The color wheel illustrates the depth cueing and 64 hue levels available in Langridge's computer graphics program.



The MIDAS system first displays one element at a time (Si, Al, then Na), stores them in its memory and then displays the color composite. In this case, Na (sodium) is an impurity in semiconductor material composed of Si (silicon) with Al (aluminum) overlay.



Carol A. Foote

Wipke studies a compound's genealogical tree. Wipke drew in the tree top, a compound, and the computer added the branches, possible synthetic routes to the top.

Wipke of the University of California at Santa Cruz. "Once you have decided on a molecule, it helps you determine how to make it."

SECS overcomes the prejudices, blocks and blind spots of scientists that often result in mistakenly ruling out feasible routes to a synthetic goal. Wipke says the program now is used worldwide to assist chemists in developing new or modified drugs and synthetic compounds modeled after naturally occurring substances.

But Wipke and colleagues continue to sharpen their tool. Programing a computer for SECS is harder than teaching one to play chess, Wipke says, "because chemistry is more complicated than chess." One SECS wrinkle the UCSC researchers are ironing out is the program's limited library of starting materials. In a recent test of SECS' ability, Wipke and co-workers asked it to show all of the possible synthetic paths to a particular insect pheromone — one that already had been synthesized a dozen different ways. While the SECS computer identified eight of the 12 existing paths and two more routes analogous to existing ones, it failed on two accounts due to the absence from its library of certain commercially available starting materials. "Now we're working on adding starting materials to the program's library," Wipke says.

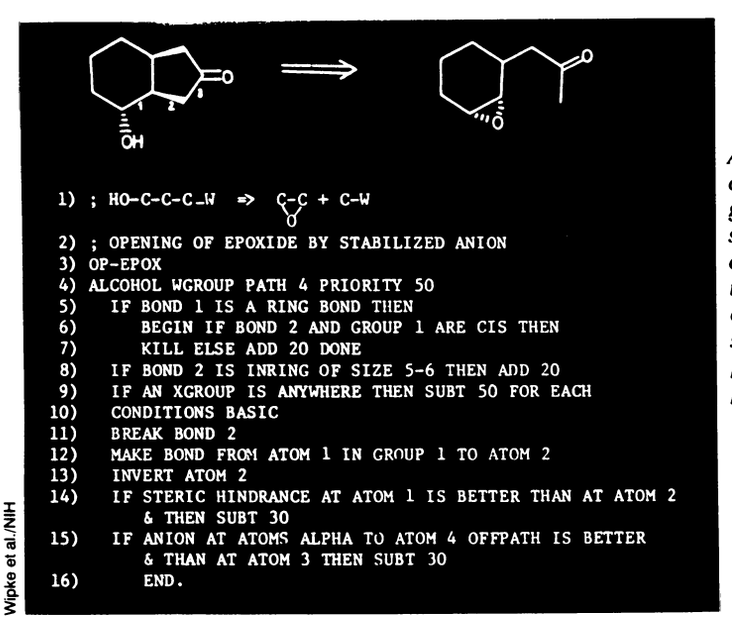
At the same time, Wipke and co-workers also are nurturing a second pet project — XENO, a computer program that can assist chemists in predicting the biological activity of xenobiotic (foreign) compounds, such as pesticides, by considering the compounds formed from the original xenobiotic after it undergoes metabolism in the body. "In many cases, the biological activity observed after administering a xenobiotic compound is really resulting from one or more metabolites and not directly from the original compound," says Wipke. XENO, therefore, attempts to correlate the activity not with the structure of the original compound, but rather with the metabolite structures.

XENO predicts the metabolic reactions after a chemist draws the xenobiotic compound on the computer screen. The chemist also must specify what animal species is involved. Thus far, the XENO library can generate lists of metabolites for xenobiotic reactions in mice, rats and human beings.

The generated list of metabolites, however, has "certain simplified features," Wipke says. The XENO computer cannot, for example, take into account that xenobiotic compounds often have to cross body membranes to reach crucial metabolic enzymes and that they sometimes are excreted from the body before crossing the internal divides. "The effect of not simulating transportation is that the computer generates some metabolites that probably would not be observed *in vivo*," says Wipke, "but at least we're not missing any that could occur *in vivo*." In other words, he explains, "We may overpredict, but we'll never underpredict."

XENO's prediction of xenobiotic metabolites is one of several computer programs related to a new trend in chemistry research — quantitative structure activity relationship (QSAR). The method of QSAR is described by Yvonne C. Martin of Abbott Laboratories in North Chicago, Ill., in the March JOURNAL OF MEDICINAL CHEMISTRY; it also was the topic of a recent meeting in Raleigh, N.C., sponsored by the Chemical Industry Institute of Toxicology, the U.S. Environmental Protection Agency, the National Institute of Environmental Health Sciences and The Burroughs Wellcome Co. Simply put, QSAR methods investigate the relationship between molecular structure and biological activity. The chemical world according to QSAR is one in which researchers should be able to predict the toxicity of a substance or the efficacy of drugs merely by looking at the structures of those compounds. Although that world still is in its primitive stages, "As with any tool," says Martin, "it is to be expected that QSAR will evolve."

The QSAR technique first involves assigning descriptors, or numerical codes, to those structures of compounds that are relevant to biological activity. Researchers then attempt to use the method to establish a quantitative relationship between the descriptors and the biological activity of the compounds. A computer program designed to automate such correlation is ADAPT — automatic data analysis using pattern recognition techniques. ADAPT, designed by Peter C. Jurs and co-workers of the Pennsylvania State University at University Park, attempts to separate carcinogens by recognizing the descriptors that code for pertinent structural features of the compounds under investigation.



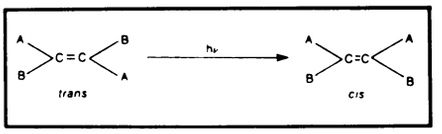
Wipke et al./NIH

A printout of the computer program SECS, designed to help chemists choose the correct strategy for successful synthesis of important compounds.

The program attempts such discrimination after studying an appropriate training set of "knowns."

Tests using those training sets indicate that ADAPT now accurately classifies only if the molecular pool is limited to one class of compound — polyaromatic hydrocarbons (PAH's) for example. Eventually, though, ADAPT may be a useful tool for setting toxicological testing priorities by screening diverse batches of suspect chemicals.

The screening potential of computer chemistry programs already is being exploited in a Battelle search for dyes that can store a sufficient amount of solar energy. The energy of light can cause the structure of those dye molecules to convert from the trans to the cis configura-



tion. Since it takes light energy to drive this transformation, the resulting cis configuration has stored energy. The Battelle researchers knew that a series of indigo dye derivatives could undergo light-driven, trans-cis conversions. What they did not know, however, was which dye derivatives in the cis form could store at least 15 kilocalories per mole (6.023×10^{23} molecules), the estimated minimum for profitable storage devices. "With molecules of this complexity, synthesis and characterization can be very difficult," says Battelle's Wolken. "In fact, for a typical compound in this series, approximately a man-year of effort could be required to make the compound to investigate its properties thoroughly."

To save those years of synthetic effort, the researchers decided to use a computer program to screen all of the potential dye derivatives — a procedure "far more economical than a purely experimental approach." Although the computer could not calculate the precise energy storage value for each dye, it was able to calculate trends that helped identify "the best prospects for success ... among the many possible structures."

In addition to enlisting the aid of the computer in their search for chemical solar energy collectors, Battelle researchers have used computers to study chemical corrosion and to calculate the heat of different reactions — a program that eventually may be used to analyze the many chemical steps that occur among short-lived molecules involved in the combustion of fuels.

"New uses for the computer are emerging," Wolken says. "In more and more laboratories, computers are being used to predict the properties of molecules before they have been produced." And although "much work has been done," says Wolken, "much work needs to be done." □

BOOKS

BOOKS is an editorial service for readers' information. To order any book listed or any U.S. book in print please remit retail price, plus 50¢ handling charge for each book to **Book Order Service,** Science News, 1719 N Street, N.W., Washington, D.C. 20036. All books sent postpaid. Domestic orders only.

ANIMAL PLAY BEHAVIOR — Robert Fagen. Addresses a major biological paradox. Why do young and old animals of many species spend time and energy, even risk physical injury, performing the apparently unproductive behaviors called play? What makes this "useless" activity so important that animals literally risk their lives for it? Presents an evolutionary approach to the development of behavior based on recent advances in theoretical biology. Oxford U Pr, 1981, 684 p., illus., $29.95, paper, $14.95.

ANNUAL REVIEW OF ANTHROPOLOGY, VOL. 9 — Bernard J. Siegel, Alan R. Beals and Stephen A. Tyler, Eds. An article by Cora Du Bois, "Some Anthropological Hindsights," introduces this volume. Annual Reviews, 1980, 646 p., illus., $20.

ANNUAL REVIEW OF PHYSICAL CHEMISTRY, VOL. 31 — B.S. Rabinovitch, J.M. Schurr and H.L. Strauss, Eds. Activity in diverse newer branches of physical chemistry has been matched by a resurgence of interest in some of the more traditional areas. This volume reflects progress across the broad spectrum of this field. Annual Reviews, 1980, 675 p., illus., $20.

BIGFOOT: A Personal Inquiry into a Phenomenon — Kenneth Wylie. For many years controversy has raged over the existence of Bigfoot, a two-legged creature of enormous size said to be roaming the backwoods of America. The author, a historian, anthropologist and amateur naturalist, follows the many dubious and not-so-dubious physical trails of the beast, surveys the extensive literature, consults scientists and interviews leading Bigfoot hunters. This is a history of a strange phenomenon together with studies of the people who maintain the pursuit. Viking Pr, 1980, 268 p., illus., $14.95.

THE COMPUTER: From Pascal to von Neumann — Herman H. Goldstine. A history of the computer concentrating on the ideas and the people who conceived them. This work is presented in three parts: pre-World War II; the war period, particularly at the Moore School of Electrical Engineering, University of Pennsylvania; and the post-war years at the Institute for Advanced Study in Princeton through 1957. Originally published in hardback in 1972. Princeton U Pr, 1980, 378 p., illus., $6.95.

A COUNTRY HERBAL — Lesley Gordon. A mix of folklore, history and fact tells the story of each herb — its physical description, medicinal purposes, household and cosmetic uses and cultivation. Beautiful illustrations enhance the descriptions of over 130 different herbs and spices, both common and rare, poisonous and useful. Mayflower Bks, 1980, 208 p., color/b&w illus., $19.95.

DISCOVERING THE HUMAN BODY: How Pioneers of Medicine Solved Mysteries of the Body's Structure and Function—Bernard Knight. Tells how the structures and function of the human body were discovered, by whom and how parts of the body were named. This historical background is given system by system. Lippincott & Crowell Pubs, 1980, 192 p., illus., $17.95.

INSECT WORLDS: A Guide for Man on Making the Most of the Environment—Lorus J. Milne and Margery Milne. Presents the fascinating world of insects, which have outnumbered all other forms of animal life for millions of years. Shows how insects make the most of the environment, whether it is the desert, the water surface or the forest canopy. Scribner, 1980, 274 p., illus., $12.95.

A MATTER OF LIFE: The Story of a Medical Breakthrough — Robert Edwards and Patrick Steptoe. The authors describe their lengthy struggles that led to the birth of the first baby conceived *in vitro*, the setbacks at every stage and the opposition that they encountered. The doctors discuss the ethical implications of their work and its long-term medical applications. Morrow, 1980, 188 p., illus., $9.95.

McGRAW-HILL ENCYCLOPEDIA OF ENERGY — Sybil P. Parker, Ed. The first section of this reference work features major signed articles on the conservation of energy, exploring energy choices, the risk of energy production, energy consumption, the outlook for fuel reserves and protecting the environment. The second section is devoted to 300 articles that examine the sources, development and distribution of energy. Explains air conditioning and air pollution to wet cells and wind power. McGraw, 2nd ed., 1981, 838 p., illus., $34.50.

THE RED LAMP OF INCEST — Robin Fox. A noted anthropologist traces the origins of the uneasiness we feel with this ancient taboo to find out what incest can tell us about who we are and how we evolved. Dutton, 1980, 271 p., $12.95.

SAVING THE TIGER — Guy Mountfort, foreword by Sir Peter Scott. A study of the tiger, how it hunts, mates, rears its young and marks its territories. Tells the story also of Operation Tiger, a successful campaign to save the tiger from extinction. Magnificent color photographs. Viking Pr, 1981, 120 p., color/b&w illus., $16.95.

TREES FOR EVERY PURPOSE — Joseph Hudak. Tells how trees grow and function, the basic requirements of trees, explains care for existing and transplanted trees and landscape guidelines. The main section is a "catalog of trees" that depicts 277 individual trees for landscaping. Each tree is described and illustrated and requirements as to appropriate climate, soil, light, exposure and care are given. McGraw, 1980, 229 p., illus., $21.95.