phospholipid content lend support to the electron microscopic evidence that this fraction is composed of smooth membranes. Electron microscopy of the pellet formed below the 2*M* sucrose layer showed that this was composed of free ribosomes (Fig. 1B). There was no apparent contamination by nuclei, mitochondria, or heavy membrane elements.

Deoxyribonucleic acid polymerase activity can be dissociated from the free ribosomes, but not the smooth membrane fraction, in the presence of 0.2*M* KCl or NH_4Cl. The ribosome and membrane-associated enzymes have been purified about 1300- and 400-fold, respectively, over the activity measured in the postmitochondrial supernatant solution (*7*). Gel-filtration on Sephadex G-200 indicates that the membrane-associated DNA polymerase is of relatively high molecular weight, whereas the ribosome-associated polymerase is of lower molecular weight. The latter behaves like a basic protein during chromatography on diethylaminoethyl cellulose and phosphocellulose, whereas the membrane-associated activity appears to be more acidic. Thus, the two fractions display differences in both physical and enzymatic properties.

There is still a question as to whether the distribution of these enzymes is an artifact produced during the fractionation. The ability to dissociate the enzyme activity from the free ribosomes by 0.2*M* KCl or NH_4Cl might be suggestive of nonspecific association (*9*), although this salt concentration has recently been used to dissociate protein factors from the bacterial ribosomes of which they are considered to be an integral part (*10*). Also, fractionation of normal rat liver with different homogenization and extraction conditions produced similar qualitative results (*7*). The requirement for exogenous DNA and the denatured DNA primer preference of the membrane-associated polymerase activity suggest that this enzyme differs from the mitochondrial membrane-associated polymerase reported by Schultz *et al.* (*2*). However, comparative studies of the physical and enzymatic properties of the enzymes reported here and those of rat liver nuclear and mitochondrial DNA polymerase are needed.

The results presented here do demonstrate that the DNA polymerase activity with a preference for denatured DNA primer, previously reported to be markedly increased in the postmicrosomal supernatant solution from hepatomas (*4*), is associated with the smooth membrane fraction. A similar but lower DNA polymerase activity is also associated with this subcellular fraction from normal rat liver. Deoxyribonucleic acid polymerase activity with a preference for native DNA primer is associated with the free ribosome fraction from both liver and hepatoma tissue.

The significance of the association of different DNA polymerase activities with these cytoplasmic fractions and the increased activity in hepatomas has not yet been defined. The alterations in the endoplasmic reticulum (*11*) and distribution of ribosomes (*12*) that occur in hepatomas could certainly influence this apparent association when the same methodological approach is applied to liver and hepatoma tissues. Furthermore, it is of interest that an association of DNA with microsomes and polysomes has recently been reported (*13*), although the function and origin of this DNA remains obscure at the present time (*14*).

EARL F. BARIL
MURPHY D. JENKINS
OLIVER E. BROWN, JOHN LASZLO
Department of Physiology and Pharmacology and *Department of Medicine, Duke University Medical Center* and *Veterans Administration Hospital, Durham, North Carolina*

References and Notes

1. R. Howk and T. Y. Wang, *Arch. Biochem. Biophys.* **133**, 238 (1969).
2. R. R. Meyer and M. V. Simpson, *Proc. Nat. Acad. Sci. U.S.* **61**, 130 (1968); G. F. Kalf and J. J. Ch'ih, *J. Biol. Chem.* **243**, 4904 (1968); S. R. Schultz and S. Nass, *Fed. Eur. Biochem. Soc. Letters* **4**, 13 (1969).
3. R. Mantsavinos, *J. Biol. Chem.* **239**, 3431 (1964); J. T. Bellair, *Biochim. Biophys. Acta* **161**, 119 (1968); Y. Iwamura, T. Ono, H. P. Morris, *Cancer Res.* **29**, 1557 (1969); P. Ove, M. Jenkins, J. Laszlo, *Biochim. Biophys. Acta* **174**, 629 (1969).
4. T. Ono and Y. Umehara, *Gann* **6**, 99 (1968); P. Ove, J. Laszlo, M. D. Jenkins, H. P. Morris, *Cancer Res.* **29**, 1557 (1969); P. Ove, O. E. Brown, J. Laszlo, *ibid.* **29**, 1562 (1969).
5. G. Blobel and V. R. Potter, *J. Mol. Biol.* **26**, 279 (1967).
6. R. K. Murray, R. Suss, H. C. Pitot, in *Methods in Cancer Research*, H. Busch, Ed. (Academic Press, New York, 1967), vol. 2, p. 239.
7. E. F. Baril, O. E. Brown, M. D. Jenkins, J. Laszlo, in preparation.
8. We thank Drs. L. Lessin and S. Quarfordt, Veterans Administration Hospital, Durham, North Carolina, for the electron microscopic analysis and the lipid analysis, respectively.
9. M. L. Petermann and A. Pavlovec, *J. Biol. Chem.* **236**, 3235 (1961).
10. J. Davison, L. M. Pilarski, H. Echols, *Proc. Nat. Acad. Sci. U.S.* **63**, 16 (1969).
11. H. C. Pitot, *Ann. Rev. Biochem.* **35**, 335 (1966).
12. T. E. Webb, G. Blobel, V. R. Potter, *Cancer Res.* **25**, 1219 (1965).
13. H. E. Bond, J. A. Cooper, II, D. P. Courington, J. S. Wood, *Science* **165**, 705 (1969); W. C. Schneider, E. L. Kuff, *J. Biol. Chem.* **244**, 4843 (1969); E. Bell, *Nature* **224**, 326 (1969).
14. D. Fromson and M. Nemer, *Science* **168**, 266 (1970).
15. We are grateful to Dr. H. P. Morris of Howard University for providing the tumor-bearing rats used in this study. Supported in part by PHS grant CA-08800-04.

2 March 1970; revised 23 April 1970

Drug Effect Prediction by Computer

Abstract. *The mass of information available about effects of chemical substances (drugs) on behavioral, biochemical, and physiological systems of living organisms is so extensive as to defy traditional methods of analysis. A procedure that provides automated, computerized searches for patterns among these effects has been developed and has been applied to a data base constructed of medical and chemical information from* The Merck Index. *One promising result is the development of new hypotheses about mechanisms of drug action.*

A considerable amount of information about the behavioral effects of chemical substances is, in a practical sense, buried in medical literature and is therefore neglected. Several factors contrive to create this situation. (i) Knowledge of the effects of drugs, based upon accumulated medical experience, is so vast as to be unmanageable without systematic means for accessing (or retrieving) it. (ii) Traditional methods of search, even though systematic, require expenditures of time and money that are frequently unreasonable. (iii) Behavioral effects are often codified in terminology that lacks precision in terms of psychologically meaningful dimensions. (iv) New knowledge about drugs tends to be related to specific problems with little opportunity for generalizing. (v) A number of scientific disciplines are involved in drug analysis, and interpretations across multidisciplinary lines are difficult.

Advances in computer technology have increased the feasibility of developing automated search routines (software packages) that permit gathering information from large bodies of data without an intermediate encoding process. The costs of accessing such natural-language data bases are modest enough to justify conducting inquiries on a broad scale, even in cases where payoff potential may be low. In fact, the new computer systems have again

made inductive research strategies profitable. Instead of formulating a specific hypothesis and then gathering data that would validate or invalidate this hypothesis, there is an opportunity to begin with a comprehensive group of data and to examine it repeatedly and interactively with new hypotheses.

A source of comprehensive knowledge about drugs, suitable for construction of a natural-language data base, is *The Merck Index* (*1*), which offers a broad listing of chemicals, together with their characteristics and uses. Each of these chemicals is identifiable in the Index by its common name and by synonyms that include all chemical and trade names by which the chemical has been referred to in the literature. If a chemical has a medical use (about 4000 in the current 8th edition do), the entry under the common name includes medical information that ranges from one word (for example, "analgesic") to a sentence or more, often including the dosage. Information about toxicity and side effects is also provided. The behavioral effects for which we search are implicit, if not explicit, in the medical use. If these effects are implicit and perhaps loosely defined, a careful analysis of the similarities and differences among medical use terms (semantic analysis) can be employed to extract the appropriate information.

The fact that *The Merck Index*, in its several editions, presents a historical record of advances in pharmacology is another feature that recommended it for our use. By comparing the 6th (1952) edition with the 8th (1968) edition, it becomes possible to trace the growth of knowledge about psychoactive drugs. In 1952, drug research was on the threshold of discoveries that have brought far-reaching changes in psychiatry. The development of psychoactive drugs has had a profound influence on work in neurochemistry and has served to generate many new questions.

First estimates indicated that the desired entries from both the 6th and the 8th editions totaled approximately 3 million characters. In order to use information under "Medical Use" as source data for semantic analyses, we were obliged to copy this material verbatim. In the interest of economy and manageability, chemical information beyond the common name, the synonyms, and the formula was excluded. The structural formula can easily be retrieved from the Index volume itself after the common name has been found (*2*). It was discovered in pilot runs with sample data that the Geneva system (*3*) names, which are provided for each compound, are a valuable source of chemical information in and of themselves. By practicing economy, the data files actually contain only 2.2 million characters.

In order to use search routines available in present-generation software packages, it was necessary to begin each line of copy with the common chemical name or, in instances in which the names became too lengthy, an abbreviation of the name. The program commands used in searching typically retrieve a single line of copy and, for the purposes of our work, each line must have a chemical tag. Savings that could be realized from a special number code were deemed too great a price to pay for giving up the convenience of natural-language tags. To facilitate the retrieval of different kinds of information, four separate files were established. Two of these consist of medical-use entries from the 8th and the 6th editions of the Index, the third is a chemical formula file, and the fourth and largest file contains all of the synonyms. For use with a standard teletypewriter, subscripts and superscripts have to be written on the line, only capital letters can be used, and names of Greek letters have to be spelled out with the English alphabet. Only one kind of parenthesis is available on the teletypewriter keyboard, and differences between lower case *l* and the numeral *1*, and between upper case *O* and the numeral *0* (zero), must be scrupulously observed.

Four college girls punched the 2.2 million characters on paper tape with model 33 teletypewriters. The paper tapes were read with a rapid optical tape reader and transferred to magnetic tape, which is our data storage medium (*4*). Editing the data base for errors, of which there were relatively few, was done on-line, using an "Editor" program provided by Tymshare, Inc.

Accessing is accomplished after the data have been transferred from tape to temporary disk storage. Search routines are now generally available from

Table 1. Medical use of indole nucleus compounds (alphanumeric string: indo). Med, medical; Vet, veterinary; and G.I., gastrointestinal.

Compound	Medical use
	Physiologically recognizable adrenergic effects
Carbazochrome salicylate	Med and Vet use: systemic hemostatic for capillary bleeding characterized by increased capillary permeability
Emex	Hemostatic agent
	Biochemically recognizable adrenergic effects
Benanserin hydrochloride	Formerly tried as serotonin antagonist
Mebhydroline	Antihistaminic
Medmain	Experimental serotonin antagonist
10-Methoxyharmalan	Experimental serotonin antagonist
	Behaviorally recognizable adrenergic effects
Adrenochrome	Has been used experimentally to produce psychic effects
Adrenoxine	See Adrenochrome
Bufotenine	Hallucinogenic
Chlorthalidone	Diuretic, antihypertensive. Dose: oral, 50 or 100 mg/day. Side effects: weakness, dizziness, and G.I. disturbances may occur Vet use: Diuretic
Chlorisondamine chloride	Med and Vet use: antihypertensive, ganglion-blocking agent
Clorexolone	Hypotensive; diuretic
Harmaline	Amine oxidase inhibitor; CNS stimulant
Harmalol	Seeds have been used as anthelmintic, narcotic
Harmine	CNS stimulant
Indomethacin	Anti-inflammatory, analgesic, antipyretic agent. Dose: oral, 25 mg. Side effects: headache, dizziness, and G.I. disturbances may occur
Molindone	Hydrochloride has antipsychotic properties
Oxypertine	Tranquilizer
Solypertine	Adrenergic blocking agent
	Nonadrenergic effects
Etryptamine	Formerly as antidepressant
Indocyanine green	Diagnostic agent
Indoxole	Proposed anti-inflammatory, antipyretic agent
Methisazone	Antiviral agent
Oxyphenisatin acetate	Laxative. Dose: oral, 2 to 5 mg
Triacetyldiphenolisatin	Laxative: Dose: oral, 10 to 20 mg
L-Tryptophan	Nutrient
Tylocrebrine	L-Form as experimental antileukemic agent

commercial time-shared computer services. These routines locate alphanumeric strings (a group of letters, numerals, punctuation, and spaces) specified by the user and actuate a printout of each line that contains the string in a specified file of the data base.

The Geneva system names included in the synonym file provide access for the generation of lists of medical uses on the basis of structural modules that are common to the compounds. A group of compounds that contain the same structural module often has so many different descriptors that recognition of similar medical uses is obscured. This is why searching for medical use alone, without the additional sorting capability provided by the presence of chemical information in the data base, will not yield the desired information. The fact that these structural modules (identical fragments of different molecules) can be searched for in plain English is a welcome innovation arising from this study. Results of two searches for such modules are reported here.

A trial run was initiated by a search command that looks for the alphanumeric string "phenethylamine." The occasion for the search was a question about the mechanism of action of mescaline (*5*). Mescaline itself is listed only as "Exptl psychotomimetic agent," which affords little clue as to its mode of action. The search, however, associates mescaline clearly with sympathomimetics and with other substances whose effects depend on stimulation of the sympathetic nervous system, such as anorexigenics (agents that control the appetite). On-line time for this search was 1 hour and 38 minutes. Similar compounds that were not sorted out by the computer, because the Index does not name them as phenethylamines, include phenylephrine hydrochloride and amphetamine phosphate, both of which are also sympathomimetics.

Table 1 contains a list of compounds sorted out by a search command that looks for the alphanumeric string "indo." Of the 27 compounds found in this search, 19 have adrenergic effects (*6*). These, in contrast to the eight compounds that do not have adrenergic effects, are so substituted that they can undergo quinone methide formation and can thus act as adrenergic agonists (*7*). The *Rauwolfia* alkaloids, which also contain the indole nucleus, were not sorted out by the computer because the Index did not name them as indole derivatives. Some medical descriptors for adrenergic effects which appear in Table 1 and which have behavioral implications are: psychic effects, hallucinogenic, antihypertensive, hypotensive, central nervous system (CNS) stimulant, narcotic, analgesic, antipsychotic, tranquilizer, adrenergic-blocking agent, and ganglion-blocking agent.

The drugs listed in Table 1 show a continuity of effects, from those that are physiologically recognizable to those that are behaviorally recognizable. These drugs also show, as an intermediate stage, effects that are biochemically recognizable. The existence of this continuity can be used for research planning. For example, patients receiving prenylamine, dopamine, racefemine, or tyramine (phenethylamines for which only physiologically recognizable effects are listed) should show definite behavioral effects. The rest of the amphetamines that appear in the phenethylamine search may, in large doses, exercise some useful physiological effects. The existence of this continuity also implies that a separate mechanism need not be sought for the action of psychoactive drugs. The implication is that drug action (for example, adrenergic effects by compounds that contain indole rings) is effected by the same mechanism (in the example, lowering of capillary permeability) everywhere in the body, but that some organs (in the example, the brain) have lower response thresholds.

Each of the four files in the data base is a collection of words that forms a semantic universe. The 11 behaviorally recognizable adrenergic effect descriptors listed in the previous paragraph were used as alphanumeric strings in searches of the medical-use file. It is apparent from these searches that the semantic universe represented by the medical-use file contains a number of overlapping clusters. For example, 59 antihypertensives were found. Of these, eight are also described as ganglion-blocking agents, five as adrenergic-blocking agents, and one as a tranquilizer (*8*).

Another cluster centers around analgesics and narcotics. Of 173 analgesics, 33 are also described as narcotics. Correspondingly, 33 out of 60 narcotics are also described as analgesics. Learning more about the nature of these semantic clusters and their overlapping makes it possible to (i) construct a picture of the points of contact between medical uses and structural modules of molecules, (ii) identify sources of confusion in the medical terminology (semantic noise), and (iii) find new cues to possible mechanisms involved in drug effects.

In summary, a technology has been developed for systematizing a large and presently confusing mass of information in an area (psychopharmacology) which has important implications for application as well as for basic knowledge. Procedures, employing the capabilities of the modern computer, are described which (i) store the wealth and variety of information about reported effects of drugs (chemicals) on behavioral, biochemical, and physiological properties of living organisms; and (ii) search for patterns among the drugs in terms of molecular structures and of medical uses. One important result of this pattern recognition is the potential that it has for the generation of hypotheses that can be put to test in the current effort to explore the mechanisms of drug action. Only construction of the data base and two examples of its use are reported here. The two examples do, however, confirm that no separate mechanism for the action of psychoactive drugs need be invoked. Rather, a continuity of somatic and psychic effects of drugs is evident.

EUGENE E. GLOYE
RUDOLPH J. MARCUS
Office of Naval Research,
1030 East Green Street,
Pasadena, California 91101

References and Notes

1. P. G. Stecher, *The Merck Index* (Merck, Rahway, N.J., 6th ed., 1952; 8th ed., 1968). We thank Mr. Stecher for permission to use the Index for this study, and for his cooperation.
2. The staff of *Chemical Abstracts* has developed a method for coding structural formulas for computer storage and retrieval, but it has not been practical for us to adopt this system.
3. The Geneva system of nomenclature has been universally used by chemists since 1864 and is kept up-to-date as new nomenclatural exigencies arise. In the Geneva system, each syllable uniquely defines a structural module of a molecule, and the position of the syllable in the name tells how the modules fit together in the molecule.
4. We acknowledge with thanks the generous efforts of Robert Iverson and his staff at the Jet Propulsion Laboratory, Pasadena, in accomplishing our media transfer.
5. The search on mescaline was suggested by Dr. G. C. Tolhurst of the Office of Naval Research, Washington, whom we thank for his support and encouragement.
6. The search for indole-ring compounds was suggested by Dr. Ernest S. Barratt of the University of Texas Medical Branch at Galveston, whom we thank for helpful correspondence.
7. A. A. Larsen, *Nature* **224**, 25 (1969).
8. Most of these descriptors have come into use since the publication of the 6th edition (1952) of *The Merck Index.*

27 April 1970