Correlation of Chemical Structures and Biological Data[1]

By PAUL N. CRAIG
Smith Kline & French Laboratories, Philadelphia, Pa.
Received September 21, 1961

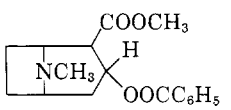
The classical approach to the development of new drugs may be expressed by a "Structure-Function" formula

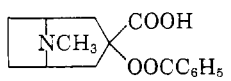
$$A \xrightarrow{1} (B \underset{3}{\overset{2}{\rightleftharpoons}} C)n \longrightarrow X$$

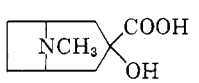
(a) A "lead" compound (A) of known chemical structure has a given biological activity. (b) The medicinal chemist plans and prepares (step 1) new compounds (B) related structurally to A. (c) Results (C) of biological testing of B are used by the chemist to guide his choice of which additional compounds (B_n) next to prepare (Steps 2 and 3). (d) Compound X, an outgrowth of steps $(B \rightleftharpoons C)_n$, is found to have optimal properties, and is selected for clinical evaluation.

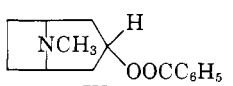
As a result of both positive and negative biological test data, the medicinal chemist gradually relies less on intuition and more on these positive and negative experiences to guide his selection of new compounds to be prepared.

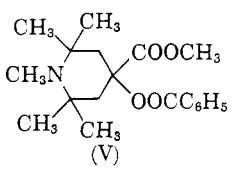
This allows a definition of the term "structure-function correlation" as *the process of utlilizing the experiences gained in steps* $(B \rightleftharpoons C)_n$ *as a guide for the design of new drugs.* Although this "Definition" is rather vague, consideration of several examples will better illustrate what is meant.

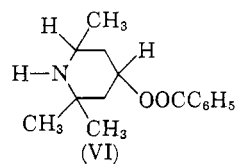

Cocaine (I)

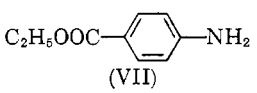

(II)

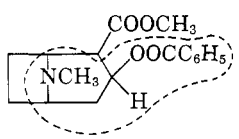
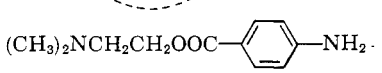

(III)


(IV)


(V)


(VI)

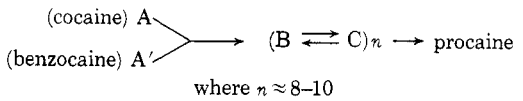

(VII)



Procaine (VIII)

As soon as the structure of cocaine (I) was elucidated, one of the first medicinal chemical studies was begun to determine those structural features which are necessary for local anesthetic activity. Hydrolysis of I to form the ester-acid II, or to form the alcohol-acid III, both led to complete inactivation. However, decarboxylation of II gave IV, which retained local anesthetic activity.

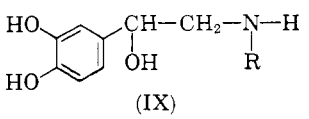
The tropane ring system was shown to be unnecessary by the good activity found for V and VI.

An independent study meanwhile had established that moderate local activity was shown by benzocaine, VII. Einhorn reasoned by analogy that structural features of both series of local anesthetics could be combined, and prepared procaine, VIII, in 1909. This research story[2] may be considered in light of the suggested structure-function formula

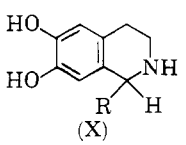


Another illustration of the classical approach is to be found in the development of bronchodilator drugs. Epinephrine (IX, R = Me) has been used for many years as a bronchodilator agent. Konzett[3] studied isoproterenol (IX, R = i-Pr) and found it to have certain advantages over epinephrine in the treatment of asthma.

R = H-(arterenol), CH_3-(epinephrine), $(CH_3)_2CH$-(isoproterenol).

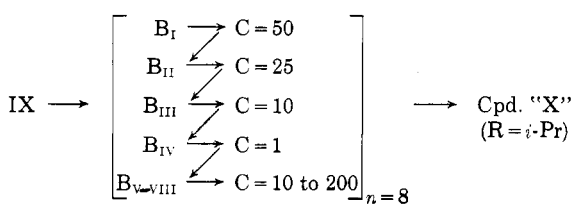

(IX)

R = CH_3- → 50
C_2H_5- → 25
n-C_3H_7- → 10
i-C_3H_7- → 1
n-, *i*-, *sec*- and *t*-C_4H_9 → 10-200


(X)

(numbers refer to dose in mg./kg., i.p. required to protect guinea pigs from aerosol. histamine).

Applying this knowledge that variation of the alkyl group "R" in epinephrine resulted in a better agent, the alkyl group "R" of a weak bronchodilator agent, X, R = Me, was varied as shown (X, R = Et, *n*-Pr, *i*-Pr, . . .).[4] From this study, it was found that optimal bronchodilator activity occurred in this series when R = *i*-Pr. The application of the structure-function formula here may be shown



Often information gained from one structure-function study is successfully applied in pursuit of a new study which may not even be in the same biological area as the first. Thus, structure relationships learned in the antihistamine area were found to be of value in the development of the potent tranquilizers. Many cross-overs similarly have been found between antispasmodic, analgetic and local anesthetic compounds.

Although many examples of the classical medicinal

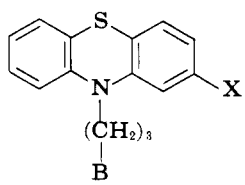
chemistry approach to the development of new drugs result in new and improved agents, there are a number of drugs which remain as the optimum compounds despite frequent attempts to improve upon their biological activity by structural modifications. Among these "durable" drugs are aspirin, digitoxin, morphine and atropine. Some of the newer synthetic derivatives of these drugs have demonstrable superiority over the prototypes in laboratory tests, but have failed to gain acceptance in medicine for various reasons.

Correlation of chemical structures with biological activity in animal tests is only worthwhile in medicine to the extent that such data can be carried over into human therapy. The jump from animal laboratory data to successful clinical validation in man is perhaps the worst link in the entire chain, partially because so few compounds are studied in man that he remains one of the least-studied test organisms. One of the best studies of the correlation of animal data with utility in humans has been found in the series of phenothiazine tranquilizers. In Table I are shown comparisons between animal activity (conditioned escape response data obtained in rats) and clinical activity (in terms of recommended daily dosage) in man (schizophrenic patients) for a series of marketed phenothiazine drugs.

Such correlations are rare, because in most cases only one or two of a series of chemically related compounds are tested in man, and therefore, insufficient data are available for comparison. Such heartwarming results as those shown in Table I help bolster confidence in the classical medicinal chemistry approach to the development of drugs.

With this rough definition of structure-function correlation, and with the admission of occasional failures to

TABLE I
Activity of Phenothiazine Tranquilizers in Animals and Humans



X	B	C.R. ED_{50}[a] mg./kg.	Rank	Recommended daily oral dose (mg.)	Effective initial[b] oral dose	Approximate clinical rank in terms of dosage
CF_3	—N⟨piperazine⟩N—CH_3 (trifluoperazine)	0.8	2	3–30	2–80	1
CF_3	—N⟨piperazine⟩N—CH_2CH_2OH (fluphenazine)	0.5	1	6–60	2–20	1
Cl	—N⟨piperazine⟩N—CH_2CH_2OH (perphenazine)	1.1	3	8–64	4–96	2
Cl	—N⟨piperazine⟩N—CH_2CH_2OAc (thiopropazate)	1.5	4	15–30	—	3
Cl	—N⟨piperazine⟩N—CH_3 (prochlorperazine)	3.7	5	50–150	10–150	4
CF_3	—$N(CH_3)_2$ (triflupromazine)	4.1	6	100–300	30–400	5
Cl	—$N(CH_3)_2$ (chlorpromazine)	9.9	7	200–600	30–1200	6
H	—$N(CH_3)_2$ (promazine)	20	8	400–1000	50–1500	7

[a] Oral dose in mg./kg. which blocks conditioned avoidance response in 50% of rats. [E. Weidley, L. Cook, *Ann. N. Y. Acad. Sci.*, **66**, (1957)]. Data calculated for the free base. Unpublished data obtained in these laboratories by Drs. D. H. Tedeschi, R. E. Tedeschi, L. Cook and co-workers.

[b] F. Ayd, Jr., *J. Med. Soc. N. J.*, **57**, 4 (1960).

come up with useful drugs by this approach, what can be done to try to improve upon this method for the development of new medicinal agents?

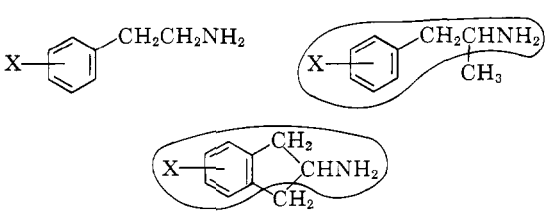
First of all, one can utilize the best available methods to try to ensure that all possible pertinent data are considered by the medicinal chemist in his planning for new compounds of type B. Rather than attempting to use machine correlation to replace the medicinal chemist, let us consider how best to use machine aids for quick and comprehensive retrieval of pertinent information to help the medicinal chemist make his decisions.

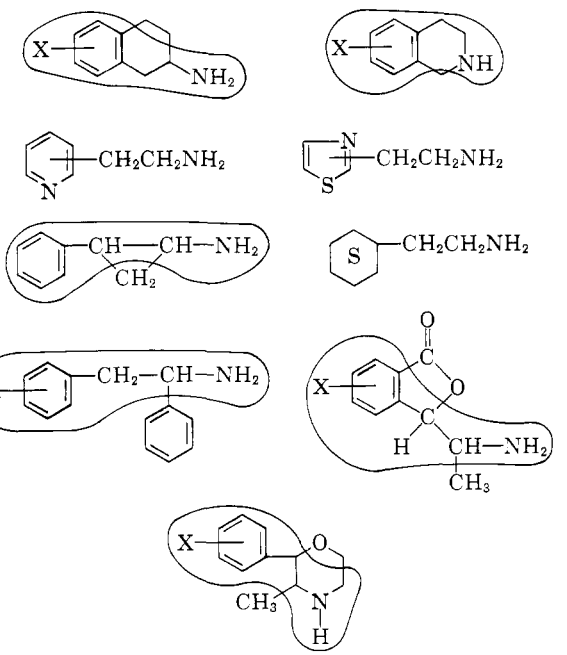
In large research organizations many unpublished internal biological test data have been accumulated on thousands of compounds. The bulk of this internal data is effectively buried, not only to outsiders in other research organizations or in academic life, but also to the scientists within the same organization, unless adequate methods of information retrieval are being applied.

The following discussion outlines the methods currently in use to handle the internal chemical and biological data at Smith Kline and French Laboratories. In order to enable quick retrieval of all generically related compounds on a structural basis, the 13,000-plus internal compounds have been coded by a modified CBCC structure code, and these codes are punched onto IBM cards, in a system similar to that described by Dr. Geer.[5] Any Smith Kline and French medicinal chemist or pharmacologist may request generic searches of the file. The problem of searching for individual compounds, of course, is adequately handled by a molecular formula file.

The ready availability of lists of compounds (by code number or by structural formulas on 3 × 5″ cards) to the medicinal chemist or pharmacologist must be matched by ready availability of the biological test results. At Smith Kline and French this is now accomplished by microfilming the central biological data file and making the microfilm copies available at strategic locations throughout the R & D Division. Thus any scientist can have the desired chemical and biological data at his fingertips with a minimum of effort and storage space.

The biological counterpart of the chemical IBM card system is of great utility but is much more difficult to set up, and will not be discussed further here. The system as described, involving IBM cards to select compounds on a chemical structural basis, followed by ready access to the original biological data for these compounds, has been found to make the work of the individual medicinal chemist and pharmacologist more efficient by eliminating needless duplication and by stimulating his imagination in ways he might not have considered. For example, in answer to a request for compounds related to β-phenethylamine, a search of the IBM cards may retrieve these types of structures





Some of these variations on the phenethylamine theme may not have occurred to the medicinal chemist, and in this manner, the machine system can help him in planning new compounds to be prepared. The medicinal chemist then can check quickly the microfilm records to see what biological activity data are already available for these particular compounds. In many cases he can learn enough from previous biological tests to arrive at a decision concerning the desirability of this type of structural change on the particular bio-activity under consideration. This saves both the chemists and biological scientists the time and effort which would have been spent in preparing a related compound and obtaining test results. Again, one must learn to be very careful not to extrapolate too far, since minor structural changes often play major roles in changing biological activities. It is axiomatic that such a system as this is much more safely applied in a positive sense (*i.e.*, suggesting new or related compounds) than in a negative sense (*i.e.*, don't test this compound).

As a further step in an effort to make the available scientific data as accessible and useful as possible to the laboratory scientists, several chemists and pharmacologists with laboratory backgrounds form the Structure-Activity Correlation group (SAC), which is part of the Smith Kline and French Science Information Department. This information group acts as a buffer between the operating mechanics of input and output (*via* machine methods) and the laboratory scientists. Thus, free from any need to understand details of the system, the laboratory scientist finds it easy to discuss his information request with the SAC group representatives. These representatives handle the programming of the machine request and also screen the initial output results for pertinency. The object of this group is to get into the hands of the laboratory scientists as much pertinent internal data as possible on a desired subject.

Finally it is worthwhile to analyze why this field of structure-function correlation is in its infancy. Many experiences have shown the lack of precise correlations between chemical structural changes and the resulting

changes in biological activity. The main reason for this is the difficulty (really the impossibility!) of altering only one variable at a time in a biological test system of multiple variables. For illustration, changing "R" in a homologous series from CH_3- to CH_3CH_2- to $CH_3CH_2CH_2$-, *etc.*, which on the surface appears to be the simplest possible change, simultaneously affects, to varying degrees, all physical and chemical properties, among which are these

Physical	Chemical
1. Molecular shape and volume.	1. Acid or base strength. (organic ion formation).
2. Electrical properties such as dipole moment, inductive effects, polarizability, etc.	2. Internal or external hydrogen bonding.
3. Liquid-water distribution ratio (solubility phenomena)	3. Reactivity in displacement reactions.
	4. Ease of oxidation (attack by radicals).

Complicate these problems with more significant structural changes, such as substitution of halogen atoms for hydrogen, *etc.*, and there remains no reason for surprise at frequent failure to obtain such correlations. Indeed, one should be gratified to find as many successful correlations as have been found, since the variability of biological determinations due to individual variations in test organisms also is a severe handicap.

Few cases where structure-activity correlations break down have been examined in sufficient detail to determine precisely the point of failure. Among the best of the continuing studies of structure-activity relationships is that of Beckett[6] on potent analgetic agents. By careful study of the effect of changes in structural size and shape, absolute configuration, and dissociation constant, on the analgetic activity of potent analgetic agents, Beckett has been able to determine the necessary absolute configuration, permissible range of base strengths, and cross-sectional size limits required for strong analgetic activity. Those who wish to design novel potent analgetics can consider Beckett's results as a guide; this may enable them to arrive at active agents more directly than before. It is obvious that such increasing knowledge requires years of basic research for each specific case.

Such detailed studies as these will take many years before a sufficient amount of knowledge becomes available to allow the advance design of optimally active drugs. Efforts toward this end are frustrated by the lack of sufficient test data. Such basically important properties as dipole moment, dissociation constant, molecular volume, *etc.*, have been determined for only a few of the drugs in current medical use. Indeed, even in such intensely studied series as the antihistamines, local anesthetics, tranquilizers, *etc.*, it is difficult to find complete biological or chemical measurements "across the board" in more than two or three test systems. The published data obtained for purposes of characterization, such as melting point, analytical and spectral data, *etc.*, are useless for this purpose.

This lack of data has led some to consider the possibility of filling in the "holes" of missing data by means of establishing relationships which may be evident to a computer in a multi-dimensional matrix study, but which are beyond the scope of human recognition. This possibility appears to offer no immediate solution to this problem. Thus we are left with the random screening and the above described empirical approach as the only alternatives, at present, for the development of new drugs. It is gratifying to look back at the past 25 years and to realize that these methods have paid off in many therapeutic advances. At the same time, it is by the slow accumulation of empirically-obtained data that sufficient clues ultimately will become available to enable structure-function correlation to become more precise.

Acknowledgment. The author wishes to acknowledge the ground work done by Dr. George P. Hager, who modified the CBCC chemistry code and initiated the work of the SAC group at SK&F. Helpful discussions with Drs. Alfred Burger and Maxwell Gordon are also gratefully acknowledged.

REFERENCES

(1) Presented at the Symposium on Information Retrieval and Analysis, sponsored jointly by the Divisions of Medicinal Chemistry and Chemical Literature, 138th National Meeting of the American Chemical Society, New York, September 15, 1960.
(2) A. Burger, in "Medicinal Chemistry," 2nd Edition, Interscience Publishers, New York, N. Y., 1960, pp. 444-446.
(3) H. Konzett, *Arch. F. exper. Path. u. pharmakol.*, **197**, 27 (1941).
(4) P. N. Craig, F. P. Nabenhauer, P. M. Williams, E. Macko and J. Toner, *J. Am. Chem. Soc.*, **74**, 1316 (1952).
(5) H. A. Geer and C. C. Howard, *J. Chem. Doc.*, in press.
(6) A. H. Beckett, *J. Pharm. and Pharmacol.* 8, 848 (1956).