Translated from *Uspekhi Khimii*, **57**, 1565-1586 (1988) U.D.C. 541.69

Future Outlook for the Construction of Biologically Active Compounds

O.A.Raevskii and A.M.Sapegin

The state of the art in regard to studies on the relation between structure and properties expressed by chemical compounds, especially biological activity, has been assessed, the determination of structure-activity relationships being the only feasible alternative to costly empirical screening. It has been shown that in this area of research, which embraces the fields of chemistry, physics, mathematics, and biology, further advances will only be possible by means of a compatible language for the computer-aided description of structure, and methods for revealing structure-property relationships and quantitative criteria of properties.

CONTENTS

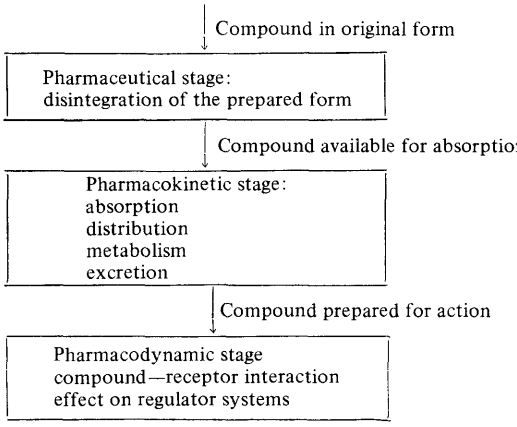
I. Introduction 897
II. Molecular recognition, the platform of the structure-activity relationship 898
III. Correct computer-aided description of structure for establishing a structure-activity relationship 899
IV. Methods for determining structure-activity relationships 901
V. Software packages for determining structure-activity relationships 904
VI. Conclusions 905

I. INTRODUCTION

Progress in science and technology is concomitant with the growth in number of chemical compounds used in industry, agriculture, medicine, and domestic life. On account of the virtually limitless scope for synthesizing new compounds, chemistry ensures a constant improvement in the exploitable properties of available substances. At present there are approximately ten million known chemical compounds. However, together they represent a negligible portion of the inifinite number of total possible compounds. This is especially evident from the data presented in communication.[1] If merely fifty substituents well characterised in organic chemistry are used for substitution in various combinations into seven free positions of naphthalenecarboxylic acid alone, 7.8×10^{12} chemical compounds may be obtained. It clearly follows that there is a high probability that compounds may be synthesized which have markedly superior properties to those of existing chemicals. In this regard the question of what strategies to adopt in order to obtain highly effective compounds in the shortest possible time with minimal effort is of major importance. The use of comprehensive tests for properties (i.e. total screening) in order to find new preparations is a very expensive excercise (with this approach in order to introduce one new compound, 12 000–15 000 chemical compounds must be synthesized and thoroughly examined). For this reason it is necessary to elaborate sound principles for the construction of compounds with the desired properties.

According to references 2, 3 the development of concepts on chemical structure was closely linked to the identification of an interrelationship between structure and the properties expressed, initially at the qualitative level[3,4] but subsequently with the discovery of the periodicity of the elements there was a gradual transition to the quantitative determination of structure-activity relationships.[5]

The problem of establishing structure-activity relationships becomes of particular interest and significance in the search for biologically active compounds. Conventional approaches such as isolation of biologically active compounds from natural sources, synthesis of analogues of natural compounds, study of folk medicine, and extensive biological trials cannot keep pace with the rapidly increaseing demand for new preparations. For this reason from the middle of the nineteenth century a new approach to the search for biologically active compounds began to develop, i.e. determination of the relationship between the structure of a compound and the biological activity expressed by the compound (structure-activity relationship). Justification of this approach was demonstrated by the work of Richardson, who for the first time was able to demonstrate qualitatively the relationship between the structure of certain organic compounds and their physiological activity,[6] and also by the studies of Crum-Brown and Fraser, who made the first attempt to construct a mathematical model representing the relationship between structure and activity.[7] The basic principles for establishing quantitative structure-activity relationships have now been defined (see reference[3] for the historical aspects of the determination of structure-activity relationships). The following schematic illustrates current concepts on the mode of action of biologically active compounds based on quantitative structure-activity relationships:[8]



The interrelationship between the structure of a compound and its biological effect following the introduction of the compound into the body becomes specific and fully expressed only at the pharmacodynamic stage. As regards other processes associated with the transformation of compounds, enzyme inhibition, diffusion of compounds in different organs, etc., information (although steadily growing[9]) is sometimes extremely limited and therefore it is difficult

to determine quantitative structure-activity relationships. Thus the problem of the quantitative structure-acitivity relationship is a special case of the structure-property relationship. However, on account of the complexity of the application of the structure-activity relationship to biological specimens, it is necessary to go beyond the traditional framework of the structure-activity problem and to focus attention not only on methods for detecting relationships and on the development of experimental investigations into structure and the form of the description but also on the quantitative determination of biological activity and on the production of a detailed description of the biological processes themselves.

There are currently thousands of publications devoted to anlalyses of quantitative structure-activity relationships. In a specialised journal (Quantitative Structure-Activity relatinships) commenced publication in 1981 in addition to original articles a large number of abstracts of studies on quantitative structure-activity relationships published in other issues including monographs and reviews are cited.[10–20] However, in most cases these studies are either specialised and focus on a single approach or problem (i.e. regression analysis,[18] the use of computers,[19] application of quantitative structure-activity relationships to the analysis of pesticide activity[20]) or mainly concentrate on a description of mathematical methods for analysis of quantitative structure-activity relationships.[16,17]

In the present review emphasis is place on the basic principles of the construction of compounds with desired properties[21,22] which constitute the physicochemical foundation for the determination of quantitative structure-activity relationships. In this case the research worker should have reliable quantitative information on the level of biological activity that the chemicals possess.[18] When the objective is to establish the quantitative structure-activity relationship there is a tendancy to construct a model which describes the interaction between compound and receptor and to use a method for determining the relationship between the correct (within the limits of the method used) representation of the structure of the compound and the activity which it expresses.

II. MOLECULAR RECOGNITION, THE PLATFORM OF THE STRUCTURE-ACTIVITY RELATIONSHIP

It has now become evident that, when the wide range of properties possessed by chemical compounds is analysed, intermolecular interactions with potential partners should be examined. The so-called "problem of recognition in chemistry" is based on a large amount of information which indicates that a particular molecule A_0 (the recogniser) reacts or forms a complex in solution only with one particular moleculer effector B_0 in a mixture of sometimes a very large number of possible effectors B_i.[23–25]

Historically it was the analytical chemist who first encountered the qualitative aspect of recognition in the search for compounds which effectively bind specific ions in multi-component solutions. With advancements in chemistry the problem of recognition has extended into organic and physical chemistry. In this reagard it is relevant to mention the fundamental studies of van't Hoff concerning the development of the principles of the theory of solutions[26] which may be regarded as the transition to a study of the quantitative aspect of recognition. As mentioned in[25], when these processes are described quantitatively, the accuracy of the recognition of chemical compounds increases by an order of magnitude in sequence from the selective donor-acceptor interaction of H-bonding of molecules containing a single electron donor and acceptor centre, and subsequently to the selective complex-formation of metal ions with polydentate ligands, and to selective catalytic reactions and ultimately to enzyme reactions. In the same sequence there is an increase in the size and complexity of the structure of molecules which are involved in the recognition process.

The methods currently adopted to investigate recognition in chemistry usually involve systems of selective complex formation between macrocyclic polyethers and cryptands (see Refs.27, 28]. However, even for these relatively simple compounds complex stability is dependent on many factors, and it is therefore difficult to establish the principles of molecular recognition. It is therefore essential to establish a clear definition of the basic principles involved in the recognition process.

According to[23–25] recognizer A_0 should bind with a particular type of molecule B_0, i.e. with the correct effector from a solution containing N types of molecule: B_0, B_1, $B_2 \ldots, B_{N-1}$ of random concentration. The number of possible false recognition events is defined as P_f. Then the degree of recognition may be expressed as the number of correct recognitions which is dependent on the total number of possible effects N and the proportion of false recognition P_f.

In the simplest case formation of complex A_0B_i may be represented by the equation:

$$A_0 + B_i \rightleftarrows A_0B_i$$

with a reaction constant of

$$\frac{[A_0B_i]}{[A_0]\,[B_i]} = K_{0i} = \exp(-\Delta G_{0i}/RT);\ i = 0, 1, \ldots, N-1$$

and the number of possible false recognitions P_f will be equal to

$$P_f = \frac{\sum_{i=1}^{N-1} [A_0B_i]}{[A_0B_0] + \sum_{i=1}^{N-1} [A_0B_i]} = \frac{\sum_{i=1}^{N-1} [B_i]\,K_{0i}}{[B_0]\,K_{00} + \sum_{i=1}^{N-1} [B_i]\,K_{0i}},$$

from which it follows that recognition becomes more accurate the greater is the value of the ratio of the correct recognition binding constant to the false recognition binding constant, or the greater the difference between the free energies ($\Delta\Delta G$) of formation of complexes A_0B_i (ΔG_{0i}) and A_0B_0 (ΔG_{00}) respectively:

$$\Delta\Delta G_i = \Delta G_{0i} - \Delta G_{00}.$$

Thus highly specific recognition demands a relatively large difference between the capacity of the recognizer A_0 to interact with the correct effector B_0 and with a false effector. When $\Delta\Delta G$ values are sufficiently high, recognizer A_0 is able to distinguish one effector B_0 from N other effectors B_i. Thus a targeted search for compounds with the desired properties could tend towards the construction of molecules the structure of which ensures maximum binding with the recognizer responsible for the property under investigation.

These aspects of the question of recognition with respect to biological specimens were further developed in studies by Fischer,[29] who by 1894 had already established that spatial compatibility affects the interaction between substrate and enzyme. This formed the basis of the so-called "lock and key" principle or "complementarity principle". This principle has now been changed to the generally-accepted concept that formation of enzyme-substrate complexes, which play an important role in enzyme catalysis, is determined by the fine

structure of active and effective centres and the unique structure of the overall enzyme molecule which ensure high catalytic activity and the specific action of the biocatalyst.[30]

Another important contribution to the development of the structure-activity concept is associated with Hansch, who in the sixties proposed an empirical model for biological activity based on the linear relationship between the free energy of the process under investigation and physicochemical parameters of the compounds. This model is the most widely used instrument in studies concerning optimization of the biological activity of chemical compounds.[1,31]

Development of the ideas of Fischer and Hansch has so far introduced two independent approaches to structure-activity relationship determinations associated with the selection of common structural elements in different compounds with one particular type of activity and with the optimization of biological activity in a homologous series of chemical compounds. It will later be demonstrated that, in order to make a transition to a qualitatively new level of interpretation of the problem of quantitative structure-activity relationships, these approaches should be unified within the same systemic study. Definition of the problem in this manner implies that the structure of the compounds has been adequately determined and correctly described, including the special features of the electron-topographic and conformational structure of the molecules.

III. CORRECT COMPUTER-AIDED DESCRIPTION OF STRUCTURE FOR ESTABLISHING A STRUCTURE-ACTIVITY RELATIONSHIP

Before commencing a study on the relationship between structure and activity, a method for presenting the original data should be selected.[14] A great deal of effort has been devoted to the selection of a system of parameters which may be either determined experimentally or by calculation and which would fully describe the structure. In the present review four main approaches for describing the structure of compounds are considered.

1. Fragment Code for Substructure Superimposition

A fragment code for substructure superimposition (FCSS) was proposed at the beginning of the seventies.[32] In the analysis of chemical structure based on this strategy potentially active centres named "descriptor centres" are identified and atoms containing mobile p- and π-electrons or a full electrostatic charge are related to these centres. They are all heteroatoms, nitrogen, oxygen, sulfur, phosphorus, halogens, and metals, and also rings and pairs of carbon atoms linked by multiple bonds. A detailed list of descriptor centres is presented in Ref.33. Initially for each compound a set of linear descriptors is produced on the basis of the structural formula, and each descriptor is represented by a seven-digit number (two of the numbers were designated to the descriptor code and the chain length. The property of conjugation was represented by one figure, which was unity in the presence of conjugation and zero when there was no conjugation). A calculation of the characteristics of the cyclic fragments when the compounds were coded for analysis of their similarity is presented in Ref.34. FCSS has been found to be useful in statistical studies on the large amount of information available on biologically active compounds. A method was subsequently proposed for coding the structure of a chemical compound by using the descriptor centre concept while retaining the topology of the molecular structure as a single entity in the form of a connection matrix for descriptor centres (CMDC).[33] In this matrix diagonal elements represent descriptor centre codes and non-diagonal elements may represent the distances between respective descriptor centres (in the case of a matrix for descriptor centre geometry) or the number of linkages between them. This approach has been used in a number of studies[16,17] for identifying the common elements in the structures of compounds possessing a specific type of biological activity.

2. Topological and Geometric Descriptor Systems

The advent of systems of topological and geometric descriptors is the natural development of a language for the description of chemical structures. The Wiswesser system has become the most widely used of the line-formula notation methods.[35] However, in view of the limitations inherent to linear coding systems, the main limitation having already been mentioned, nowadays the coding of structures with the aid of a connection matrix with a wide set of molecular descriptors is preferred. The latter include mainly topological descriptors, fragment descriptors, substructure descriptors, connection descriptors, and environment descriptors, etc. They are all designed to obtain a more accurate description of the vicinity of active centres in molecules.[14] As with FCSS, a deficiency of this coding method is that it is not the structure of the chemical compounds but their structural formula which is coded. This method for describing structure may be somewhat improved, if geometric descriptors are used in addition to topological descriptors. Geometric descriptors, which describe the general form and size of the molecules, imparts a three-dimensional structure to the compound diagram. Conformational calculations of equilibrium distances between the active centres by molecular mechanics techniques provide a more complete description of the structure of compounds.[36] However in this case also it is not possible to reveal the electronic properties of the active centres, which determine their ability to undergo intermolecular interaction. This circumstance significantly affects the identification of real interrelationships between the structure of chemical compounds and the activity which they express.

3. Quantum Chemical Methods Used for Describing Structure

Quantum chemical calculations of molecular structures have been used in structure-property relationship determinations since the time when they first appeared, and also at present as a result of progress in computer technology these methods are becoming more widely used in studies on the relationship between structural parameters of molecules and their properties, especially their biological properties.[19,21,37–39]

Currently-available computer programmes cover virtually the entire range of quantum chemical calculations from elementary schemes for the complete neglect of differential overlap (CNDO) to strict non-parametric approaches: CNDO/2, CNDO/BV, CNDO/S, INDO, INDO(S), MINDO/2, MINDO/3, MNDO, VRDDO, and the ab initio STO-39 calculation. However, the attraction of quantum chemical calculations for structure-activity relationship determinations is mainly that the most stable molecular conformations are determined in which the active centres are accessible for interaction with a receptor or that molecular physicochemical characteristics associated with the transport of the compound to the bio-target are evaluated. By transferring the "centre of gravity" of the

quantum chemical calculations to the electronic properties of the molecules, it has been possible to attempt identification (construction) of compounds with the desired properties within essentially new classes of compounds. The above-mentioned approaches do not offer such possibilities.

It is evident that the use of quantum chemical calculations for structure-activity determinations is a promising approach. Further development of computer technology and transition to non-empirical calculations[40] will undoubtedly lead to an increase in the role of quantum chemistry in the determination of structure-property relationships. Nowadays calculation of the electronic structure of large numbers of chemical compounds will be difficult to achieve even with the use of semi-empirical methods, on account of the huge demands on computer time. In this regard it will be necessary to devise a system of calculation which would be relatively simple and also which would effectively provide the correct description of the electronic structure within the limits of structure-activity relationship determinations. An attempt has been made[41] to fill the connection matrix of descriptor centres with a quantum chemical content, which takes into account the electronic structure of the molecules.

However, one should note that quantum chemistry is by no means a natural source of information on the electron-topographical structure of molecules. Moreover, in view of the above-mentioned difficulties, information obtained as a result of quantum chemical calculations is not always sufficiently reliable and accurate. Hence in the future development of methods for determining structure-property interrelationships, in addition to these calculations, physicochemical approaches to structure-activity relationship determinations will be elaborated and such approaches will be directly associated with experimental studies on the physicochemical properties of molecules or their fragments.

4. The Physicochemical Approach to Structure-Activity Relationship Determinations

Extensive use of physicochemical properties in structure-activity relationship determinations began with the studies of Hansch et al.[1,5,31] The complete set of diverse parameters used in calculations according to the Hansch equation have been described.[18,42] All of these parameters may be separated into three groups. The first group includes parameters associated primarily with hydrophobic properties, the logarithm of the partition coefficient ($\log P$), the lipophilicity parameter (π), thin-layer chromatography parameter ($\log K'$), parachor (Pr), molar volume (Vm), molecuar weight (M), logarithm of partition in water ($\log C_{aq}$), partial pressure (p), the Hildebrandt solubility parameter (σ_s^*), molar attraction constant (F_α), entropy constant (σ_s^{**}) and others. The second group includes various induction and resonance constants (σ^*, σ^0, σ^I, σ^R, etc), radical substitution constants (E_R), Swain and Lupton parameters (F, R), polarizability (α), molecular refraction (MR), dipole moment (μ), ionization potential (IP), spectral frequencies (ν), ionization constants (pK_a, pK_b), and other characteristics associated with electronic properties. Finally the third group consists of parameters characterising steric effects of intramolecular interactions, Taft's steric parameters (E_s and E_s^0), van der Waals radii (r_v), molecular binding potential (χ), etc. Although many of the above-mentioned parameters do not reflect in the pure state only hydrophobic, only electronic, or only steric properties, the separation of these parameters is extremely useful, for example in relation to the problem of multi-collinearity in the selection of parameters for a model in accordance with the scheme for the action of biologically active compounds.

Although the Hansch equation has been effectively used on many occasions, one should however emphasise that this approach and various modifications of this approach "operate" only within the narrow range of related compounds. Replacement of only one active centre in the molecule leads to a change in the nature of intermolecular interactions and consequently to the perturbation of previously-established correlations between physicochemical parameters describing the compound and the property expressed. In this regard the previously described approaches in which common structural elements may be selected would seem to be preferable, since a broad range of compounds could be studied. The question consequently arises, can the Hansch approach be extended for analysis of different orders and classes of compounds by combining this approach with other methods of differentiating common structural fragments. This concept described in Ref.43 was realised for the first time in structure-activity determinations in which the following cases were examined. In the traditional descriptor analysis of the similiarity of structures a set of descriptor centres with their classification in terms of the types of atoms and nature of the linkages is used. In an intramolecular environment the boundaries separating the electronic properties of the descriptor centres may become diffuse and this will complicate the separation of activities. It is therefore evident that descriptor centre classification should be based not on the type of atom but on its ability to participate in intramolecular interactions.

The enthalpy of complex formation (ΔH), which is proportional to the free energy (ΔG) of the interaction between molecules containing one electron-donor or acceptor centre, may provide a quantitative measurement of this interaction.[25] A wide range of experimental data on ΔH values has now been obtained by means of which it is possible to predict the thermodynamic properties of a whole variety of donor-acceptor pair combinations, and therefore one might suggest a new classification of the descriptor centres. For this purpose it is necessary either to have experimental data on ΔH values for all possible donor molecules with a given acceptor or to find a system of empirical parameters with the aid of which the ΔH value for the interaction between any donor–acceptor pair may be evaluated. In the first variant (see for example the donor numbers of Gutman[44]) there are significant limitations such as inadequate solubility of certain complexes or the possibility that an irreversible reaction may occur. The second variant has much more potential, and the two equations presented in[45,46] are extensively used for achieving this objective. In the latter modification of the equation[45] the enthalpy of complex formation is expressed as the sum of three paired products of six types of parameter. The formula based on the classification of donor–acceptor interactions and described in report[43] is much simpler.

According to report[46] the enthalpy of complex formation is proportional to the product of two dimensionless values, the i-acid (acceptor P_i) and j-base (donor E_j) factors.

$$\Delta H_{ij} = -\Delta H_{11} P_i E_j, \qquad (1)$$

where H_{11} is a coefficient determined according to a particular standard complex. In the development of this approach[43] a common scale of donor and acceptor factors has been proposed which is constructed on the basis of the following modification of equation (1):

$$\Delta H_{ij} = |\Delta H_{11}| E_i E_j,$$

where $E_i = P_i$. In later studies[21,22,47] not only neutral but also charged donors and acceptors were included in this approach. A descriptor centre classification based on this common scale has been proposed[48] for the analysis of structure-activity relationships. A list of E_i values for 70 extensively used acceptors and E_j values for 230 diverse electron-donor compounds is presented in communication.[22] Even with this relatively small group of data enthalpy values of complex formation may be calculated for more than 16 000 complexes.

From the analysis of a large group of experimental data on complex formation, the relationship may be established between $E_i(E_j)$ values and substitution constants characterising induction effects and the conjugation effect.[49,50] In this approach it was essential to calculate the values for these constants for complex functional groups. In this regard empirical equations have been derived[52,53] which describe the transfer of the induction effect via carbon, oxygen, and sulfur atoms. The creation of an automated system for evaluating $E_i(E_j)$ values should be regarded as the final stage in the introduction of donor–acceptor factors into the actual construction of biologically active compounds. Recently the authors of this review proposed and derived an algorithm on a NORD-10 minicomputer for calculating the value $E_i(E_j)$ with the automatic determination of the Taft constant of complex functional groups according to the structural formula programmed into the computer and the described types of atoms.† The mathematical software was thus created for providing a physicochemical approach to the description and classification of compounds in terms of electronic properties of descriptor centres used for selecting the structural elements responsible for a particular type of biological activity. Clarity and the high content of easily interpreted principles of the classification are features of this approach, and this classification is sufficiently straightforward for widespread use and it is already used at the qualitative level for constructing compounds with the desired properties.[54]

In addition to the correct evaluation of the electronic properties of descriptor centres, of no less importance is the accurate description of the spatial structure of the molecules, taking into account the dynamics of intramolecular motion, tautomeric and conformational transitions, and large- and small-amplitude oscillations. In this case it is evident that inclusion of descriptor centres in the previously-described connection matrix as non-diagonal elements in the values for the number of linkages between the respective descriptor centres is not consistent with current concepts on the spatial structure of molecules. Inclusion of equilibrium distances[17,33] in the connection matrix and the use of moments of inertia and molecular volumes[21] as descriptors is naturally more appropriate, but does not overcome the problem of correctly describing the spatial structure of the molecules. The use of the concept of the spatial structure of a molecule in the form of a radial distribution plot, as practised in gas chromatography, has been proposed[55] as a simple method for describing molecular structure which in general satisfies the requirements of the "ideal" method. An interesting attempt to use an analogous transformation of molecular structure in order to separate tranquilisers and sedatives[56] is noteworthy, and this method was found to be much more effective than the use of fragment descriptors with this material.[57]

The methods described for representing molecular structure reflect to a satisfactory extent the structure of the compounds, when a computer is used for determining structure-activity relationships.‡ Moreover, selection of the form of description of a molecular structure is closely linked with the particular objective of the study (classification of compounds in terms of the type of bio-activity expressed, an information search for related compounds, identification of common structural elements, a quantitative prediction of the order of bioactivity, etc.) and also with the method used for this purpose. From this point of view the question of optimal conformity between the form of description of a structure and the method selected for treatment of the structure information is very important.

IV. METHODS FOR DETERMINING STRUCTURE-ACTIVITY RELATIONSHIPS

The numerous methods available for determining structure-activity relationships can be divided into the following two major groups according to the type of problem in quesion:[21] methods directed towards the optimisation of properties in a series of related compounds and methods designed for the identification of common structural fragments in molecules of chemicals exhibiting a certain type of bio-activity.

1. Optimisation of Properties in a Series of Related Compounds

This group of methods involves the use of different multi-parameter regression models linking the structural parameters of the compounds with their bio-activity.[58–62] The main strategies for processing structural information used in structure-activity determinations for optimising the properties in a series of related compounds are presented in Table 1.

The successful use of the Hansch model described in the previous section has led to the appearance of many generalisations and modifications for "screening" of different sets of physicochemical and structural parameters and regression models. Even as early as the seventies more than 2000 regression dependences had been derived for more than 20 000 chemical compounds,[63] relating hydrophobic, electronic, steric and other properties of substituents to biological activity expressed by the compound. Certain of these relationships were used both for predicting efficacy[64–66] and also for determining the mode of action of biologically active compounds.[18,67] However, many of these, according toHansch,[68] were "astonishingly deformed". They included equations which were unrealistic in the physical sense or ineffective for predicting biological activity but which had high correlation coefficients.

The existence of these equations indicates that the high values of the correlation coefficient R, the square of which is numerically equal to the contribution of the bio-response function variance ($\sigma_y^2 - \sigma^2$) "explained" by the regression dependence

$$R^2 = (\sigma_y^2 - \sigma^2)/\sigma_y^2,$$

where σ_y is the bio-response function variance, and σ is the residual variance, is an essential condition but would not ensure a good

†A.M.Sapegin, O.A.Raevskii, V.V.Chistyakov, and I.V.Martynov, Khim.-farmatsevt. zhurn., 1987, **21**(9), 1098.

‡O.A.Raevsky, QSAR in Drug Des. Toxicology, D.Hadzi et al., (editors), Elsevier, Amsterdam, 1987, p.31.

regression model. This may be attributed to the fact that for a regression analysis there are certain limitations, and that infringement of the latter leads to "mishaps" with regression coefficients,[69] which characterise the contribution of each of the structure parameters of the compound to the bio-activity value.

Table 1. Main Methods for Optimising Properties in a Series of Related Compounds.

Form of analysis	Method of processing information	Field of use	Refs.
Multidimensional statistical	Calculation of the statistical characteristics of the distribution of random values, means, scatters, etc.	Omission of incorrect data and analysis of "rejections"	[72-76]
Multidimensional scaling	Reduction of the scale of multidimensional space with the object of omitting data and variables which do not provide sufficiently informative data	Selection of compounds which provide the greatest amount of information	[73-75]
Correlative	Identification of relationships between variables based on a correlation matrix	Testing the hypothesis concerning the independence of the parameters in the model	[78, 119]
Factorial	"Reduction" of information by introducing new variables representing linear combinations of the original data	Determination of the "structure" of the data, the extent to which the items and parameters in the model may be grouped	[79-82]
Regression and dispersion	Construction of a functional relationship between independent variables and the response (dependent variable)	Determination of the quantitative relationship between structural parameters and the bioactivity expressed by the compounds	[31, 61-68] [85-87]

For example in communication[70] attention is drawn to the irrational use of the Hansch model proposal concerning the normal rule for the distribution of the variables included in this model. Analysis of fourteen possible distribution rules for the variables in the Hansch model showed that, regardless of the hydrophobicity characteristics of the biologically-active compound, the best model for describing the transport of a compound to the bio-target is the exponent of a power rule for the size distribution of the hydrophobicity parameter with a clearly pronounced assymetry. The structure of this distribution indicates that the probability of the substrate reaching the site of action increases in proportion to its hydrophobicity in accordance with the power rule and decreases in accordance with the exponential rule. The presence of asymmetry indicates that random events in a given system are not always mutually-independent. The presence of a correlation between "independent" parameters is one of the most widespread reasons for the low predictive capacity of the derived relationships. The emergence of models, which do not effectively predict quantitative structure-activity relationships, in addition to "technical" features of the regression analysis, is associated with a number of fundamental difficulties: a) in establishing the uniformity of the mode of interaction between the group of compounds under investigation and the same receptors;[1] b) in selecting a regression model conforming to the mode of action of the compound; c) in selecting variables which adequately describe the property under analysis; d) in the use of bioactivity data obtained by different research workers (the problem of accuracy and reproducibility of a biological experiment); e) in establishing the relationship of the resultant biological activity of a compound, which affects many physiological processes occurring in the body.

On account of the above-mentioned difficulties and because there is currently no alternative to a regression analysis for obtaining a quantitative prediction of the magntitude of bioactivity, an alternative proposition would be to construct regression models after a preliminary qualitative analysis of the available data had been carried out. In this case an analysis of a quantitative structure-activity relationship could have the following sequence. In the first stage of the preliminary or, as otherwise termed, the "exploratory"§ analysis, the general structure of the data is established. For this purpose one may use graphic display of the multidimensional space of the data, and also automated methods of classification and statistical treatment of the information. At this stage within the limits of the structure-activity determination two problems may be resolved, i.e. to separate classes of related compounds, and to convert to a set of a significantly smaller number of informative parameters describing the structure of the compounds. In the following stage the hypothesis concerning the independence of parameters in the model may be checked with the aid of correlation analysis methods.[78] If necessary, by using factorial analysis methods, factors which represent a linear combination of the original parameters in the model may be introduced into the regression model.[79-82] It has been recommended[18] that the following rules should be obeyed, when the final regression model is being constructed: 1) at the preliminary analysis stage as many independent parameters as possible should be examined as model variables; 2) to select from the various combinations of variables those which contain parameters related to the different types of interaction (for example $\log P$ and σ [83]); 3) to disregard regression coefficient variables with a low level of significance; 4) to select a model with the lowest standard deviation, when the values of the correlation coefficients are similar; 5) in order to exclude random correlations, at least five–six results of the biological tests for each parameter tested in the model should be included;[84] 6) other conditions being equal to give preference to simpler models; 7) with the objective of reducing and simplifying the calculations and avoiding fundamental errors to take into account the reality of the proposed model from the point of view of the phenomenon under examination.

One should also take into account the potentialities of modern regression analysis in regard to the evaluation of the quality of the regression dependences obtained.[85-87] By correct construction of the regression dependence, a model describing the action of the biologically active compound may be obtained which has high predictive capacity and it is thereby possible to resolve the problem of optimising properties in a series of related compounds. However, it remains an open question whether it is possible to predict bioactivity of compounds which differ in structure. Resolution of this problem is dependent on the identification of common structural elements.

2. Identification of Common Structural Elements in Biologically Active Compounds

The suggestion that the biological activity of a compound results from the presence of certain molecular substructures (pharmacophores) patterns the basis for the search for common structural

§P.F.Velleman and D.C.Hoaglin, "Applications, Basics and Computing of Exploratory Data Analysis", Duxbury Press, Boston, 1981, p.354.

elements in compounds. A typical feature of this approach is the use of pattern recognition techniques (PRT)[87-89] for dividing compounds into active and inactive species according to a set of structural parameters. In these terms the Free-Wilson model,[90] based on the additive contribution of substructures in substituents to the bio-activity value, occupies an intermediate position between quantitative regression models and qualitative models of pattern recognition techniques. With this simple model it is possible to narrow down significantly the area in which searches are made for effective biologically active compounds. For example, when testing compounds in which substituents may occupy four positions or when the number of the substitutions is changed from three to five, use of the Free-Wilson method reduces the number of bio-tests required from 120 to eleven.[18] However, since information is used at the qualitative level concerning the presence or absence of certain types of pharmacophores in the structural formula of compounds, a quantitative prediction of the bio-activity value is often found to be unreliable. Thus attempts have been made to improve this model by linking the latter with the mode of action of compounds[91-94] or by including the mutual effect of the substituents in the examination.[95] However, in the latter case there is a sharp increase in the number of variables in the model and the model becomes difficult to operate.

In the case of items which may be described by a large number of parameters, pattern recognition techniques based on the automatic classification of items according to similarity or lack of similarity are found to be more effective.[96] Methods related to the three types of analysis (Table 2) are most frequently encountered in the literature concerning structure-activity determinations.

Table 2. Main Methods for Identifying Common Structural Elements and Classification of Compounds.

Form of analysis	Method of processing information	Field of use	Refs.
Discriminant	Construction of separating hyperplanes according to an instruction selection consisting of two classes of items	Selection of informative parameters of the model for describing the structure and classification of the compounds into active and inactive	[14, 77, 97-101]
Linear instruction machine	Non-parametric discriminant analysis, based on the correction of an incorrect classification by means of the inverse relationship in the iterational process	Ditto	[14, 88]
Logical-structural	A form of discriminant analysis based on the use of the laws of inductive logic	Ditto	[17]
Cluster	Classification of items according to similarity or differences with the aid of optimisation of the efficiency function	Classification of compounds according to types of activity	[102-107]
Modelling of spatial structure properties of the molecule	Calculation and construction of three-dimensional images of molecules by quantum chemistry and molecular mechanics methods with the aid of molecular graphics	Determination of the steric specificity of substrate-receptor interaction	[110-116, 119]

Discriminant analysis[97-101] is based on the division of compounds into active and inactive species by constructing a separating hyperplane or some other surface in the space of the parameters of the compounds which describe their structure. In the limits of this analysis a selection of the most informative parameters from the point of view of separating the compounds is presented, a separating surface is constructed, and the effectiveness of the resultant discriminant model is tested on instruction and examination selections of compounds. Certain data cited in the literature[99-101] serve as examples of the successful use of discriminant analysis.

Logical structural analysis[17] uses algorithms of discriminant analysis methods in which information is presented in a natural form for the research worker. Many procedures for processing information are directed towards the dialog mode of computer operation.

Cluster analysis[102] involves the grouping of items as concentrations of points in multidimensional space which describe these items of the parameters. This kind of analysis is clearly associated with multidimensional scaling techniques (mapping of the points of multidimensional space on a surface of smaller dimensions) and numerical taxonomy (separation of items into pre-determined classes). The absence of empty clusters is a distinctive feature of cluster analysis techniques. Cluster analysis methods may be used in structure-activity determinations either independently[104-106] or in combination with other methods.[77,107,108]

In all automatic classification methods the studied topological and topographic features of molecular structure are represented in the form of a set of indexed parameters, geometric descriptors which are coordinates of the compounds like points in the multidimensional space of the parameters. In recent years with the development of computer graphics it has become possible to produce a visual three-dimensional image of the molecular structure on the graphic display screen of a computer.[108-110] This has initiated yet another approach in structure-activity relationship studies, the molecular modelling of the spatial structure of compounds.

The most productive approach was found to be the use of methods for modelling the three-dimensional structure of molecules in order to study the spatial specificity of the substrate-receptor interaction by the "lock and key" principle,[110] especially when the electronic pattern of the molecule is represented as surfaces surrounding a nuclear skeleton of the molecular conformation[27] obtained with the aid of a quantum-chemical calculation[111] or the molecular mechanics method.[36] By using this approach it is possible to determine the similarity between the electron-topographical properties of the molecules which is not discernable with the other analytical methods. The quantitative aspect of the calculation of steric specificity of the substrate-receptor interaction has been developed in a regression analysis,[112-115] although a trend towards the combined use of molecular modelling techniques and regression analysis has recently become apparent[115-116] and this should ultimately lead to a system approach to the determination of structure-activity relationships.

3. The System Approach to the Analysis of Structure-Activity Relationships

The concept of the system approach to the analysis of structure-activity relationships involves the creation of a single model for the search for biologically active compounds covering both qualitative and quantitative and also a priori information on the class of compounds studied within the limits of the expert system, which includes a corresponding group of methods for determining structure-activity relationships. The usefulness of this approach is dictated by the need to take into account all information available to the research

worker and also by the selection of the optimal method for processing the information. In fact, if for example the problem of binding between receptor and substrate is defined as a problem of finding the distribution density function of the probability of complex formation according to a set of parameters characterising the structure of the substrate, it becomes apparent that there is a link between this problem and that of the reduction in the dependences under conditions of the inverse incorrect problem.[87] From the theory for the resolution of similar problems[117] one might conclude that the quality of the result obtained is to a marked extent dependent on the degree of detail in which a specific problem is investigated and the selection of the corresponding method for resolving the problem. Since at the various levels and stages of the analysis of structure-activity relationships the research worker presents information differing in quantity, quality and form in which it is presented and usefulness, there obviously is a need to use not one method for determining structure-activity relationships but a range of methods in a sequence which would ultimately lead to a quantitative prediction of the magnitude of the biological activity possessed by an unexamined compound on the basis of parameters characterising the structure of the compound.

The first stage towards combining qualitative and quantitative methods of analysis of structure-activity relationships within a single model was the elaboration of a physicochemical strategy similar to that described in Section III.

The potential of a system analysis of structure-acitivity relationships in this approach was linked with the formation of a unified description of the electron-topographical properties of molecules based on a single donor-acceptor model representing the substrate-receptor interaction, in which the model parameters could be extended as indicator variables at the automatic compound classification stage and as quantitative variables on reduction of the structure-activity regression dependences. This poses the problem of optimising the properties in the series of compounds with different structural formulae including those which belong to different classes but which have similar donor-acceptor and topographic properties of the descriptor centres.

The present tendency to integrate the methods and approaches in the structure-activity relationship determination is also reflected in the packages of programmes currently being elaborated.

V. SOFTWARE PACKAGES FOR DETERMINING STRUCTURE-ACTIVITY RELATIONSHIPS

In order to perform the methods and approaches considered for the analysis of structure-activity relationships, relevant software packages must be produced. Currently-available software packages differ both in function and in the extent to which they cover the problem to be resolved, and also in the class of computer in which they may be used. Software packages designed according to the modular principle with a dialog operation regime are most convenient. Such packages are usually designed for use in mini-computers, and on account of the ease with which new programmes can be incorporated into the package, they undoubtedly have an advantage over other software packages in regard to the production of an expert system for determining structure-activity relationships based on a laboratory mini-computer.

By way of an example four of these packages will be examined in which previously-described approaches to the analysis of structure activity relationships are performed: automatic classification based on methods for recognition of samples (ADEPT), logical-structural method (TOPLOG), molecular modelling (MOLY), and physicochemical (RASTR) approaches.

The ADAPT software packages[14] based on processed information contains the structural formulae of the compounds. Any of them may be entered into the computer by building its image on a graphic display monitor with the aid of a special set of commands in the interactive mode. The structural formulae entered are stored in the form of a connection matrix on discs (disc drives). It is envisaged that there could be facilities for the addition, exclusion, alteration, and recalling of the stored structures of compounds. Structure-activity relationship analysis is based on the methods of parametric and non-parameteric statistics and recognition of the items: linear regression, discriminant analysis; a linear instruction machine, and classification according to the number of close neighbours. The ADEPT system operates on the basis of a 16-bit MODCOMP II/25 minicomputer and together with the data bank occupies approximately 2 Megabites of memory on magnetic discs.

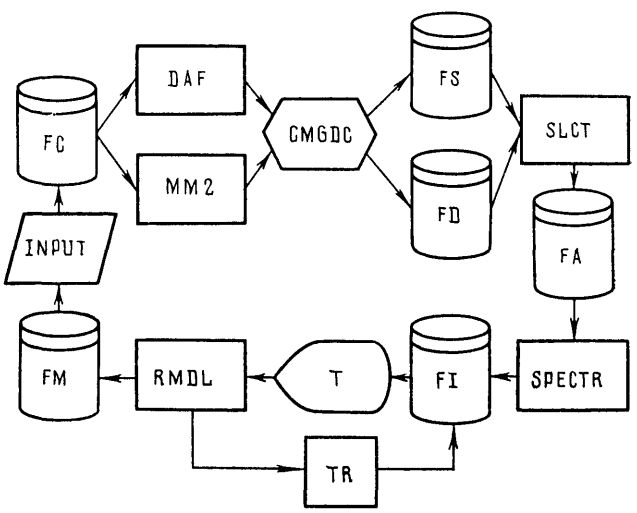
The software package TOPLOG[17] offers the user a more advanced system of input, editing, and information conversion to a form which is acceptable to the research worker. This is reflected in the file system of the package which counts 14 types of files with different functions. The package includes a procedure for achieving a logical-structural approach like that described in reference 17. A distinctive feature of the TOPLOG package is the inclusion of an algorithm for selecting common structural fragments of molecules based on an analysis of graphic representations of structural formulae of the compounds under investigation. The package consists of 14 programmes with a total capacity of approximately 8000 fortran operators and it may be run in an HP-1000 minicomputer.

In contrast to the packages described in which the main type of information is the structural formula of compounds, in the MOLY package[10,118] in order to construct a biologically-active compound a stereoscopic three-dimensional image of the molecule is used in which its properties are represented (for example surfaces of electrostatic potential). With the aid of molecular modelling techniques it is possible to obtain parameters of the molecule structure which reflect the spatial specificity of the interaction between substrates and receptors and which may be used for constructing regression models of quantitative structure-activity relationships. The MOLY package contains approximately 15 000 fortran operators and occupies 1 Megabite of memory.

In the RASTR (recognition of active structures)¶ software packages procedures are included which provide a physicochemical approach for the detection of common fragments of biologically active compounds according to the electron-topographic properties in structure-activity determinations. A feature of this package is that a system approach to the analysis of structure-activity relationships may be achieved by using a set of previously-examined methods at different stages of the study (Tables 1 and 2). This is achieved by means of a unified description of the electron-topographic properties of the molecules from the position of a single donor-acceptor model of the substrate-receptor interaction. For example one possible operational variant with the RASTR package may be as follows. The structural formula of a compound and

¶O.A.Raevskii and A.M.Sapegin, Khim.-farmatsevt. zhurn., 1987, **21**, 1338; A.M.Sapegin, A.N.Razdol'skii, V.V.Chistyakov, and O.A.Raevskii, Khim.-farmatsevt. zhurn., 1987, **21**, 1341.

information on the type of atoms contained in the formula are introduced from the computer terminal. With the aid of an INPUT programme (Fig.1) these data are converted into a connection matrix and are entered into the compound file on a magnetic disc. Information from this file is used for calculating a descriptor centre geometry matrix, a physicochemical "portrait" of the molecule describing donor-acceptor and topo-graphic properties of the active centres in the substrates. Diagonal elements of the descriptor centre geometry matrix are coded on the basis of calculated values of donor-acceptor factors in the descriptor centres in accordance with a unified scale of a donor-acceptor interactions. An algorithm for calculating the donor-acceptor factor of the descriptor centre (DAF programme) is constructed in order to analyse the tree of a graphic representation of the compound's structural formula using a data bank of Taft constants for different types of atoms and correlation equations for calculating the Taft constant of complex substituents and the donor-acceptor factor of the descriptor centre.[50,52-53] Extra-diagonal elements of the descriptor centre geometry matrix contain information on the topography of the descriptor centres in the form of the distance between respective descriptor centres for equilibrium conformations calculated by the molecular mechanics method[36] (programme MM2). The descriptor centre geomtry matrix is formed by the CMGDC programme which operates in two modes: analysis of structure-activity relationships, and in a mode for logging the descriptor centre geometry matrix into the data bank.



Operating configuration for the main nucleus of the RASTR package Programmes: INPUT, Input; DAF, calculation of donor-acceptor factors; MM2, calculation of equilibrium conformations of molecules by the molecular mechanics method; CMGDC, formation of descriptor centre geometry matrix; SLCT, selection of common structural fragments; SPECTR, construction of the spectrum of interatomic distances; RMDL, construction of regression models; TR, transformation of regression models of the pharmacophores into the spectrum of interatomic distances; files: FC, compounds; FS, structures; FD, structures in data bank; FA, activities (of pharmacophores); FI, images (spectra of interatomic distances); FM, regression models.

With the aid of the SLCT programme common structural fragments are selected for the compound under investigation (FC file) and compounds stored in the data bank (FD file). The procedure for searching for common structural fragments is analogous to that used in the TOPLOG package.[24] It differs only in that in this instance the common structural fragment is considered to be not a strict identity of the elements in the structural formula of the compound but their closeness with respect to electron-topographic properties which may be regulated by the research worker. The isolated fragments are potential pharmacophores and are entered into the biologically active structure file (FA file).

The descriptor centre geometry matrix of common structural fragments in the biologically active compounds which are contained in the FA file may be represented (with the aid of a SPECTR programme) as a range of interatomic distances, which is convenient for a visual comparison of the structures of the compounds or as a set of structural parameters for constructing regression models of structure-activity relationships. Interatomic distance spectra are stored in the "Images" file (FI file).

Thus at the stage described above the problem of identification of potential pharmacophores is qualitatively resolved, i.e. on the basis of indicator parameters of the donor-acceptor model a concept can be formed concerning the electronic and spatial properties of the active centres in substrates responsible for a particular kind of biological activity.

The next important step is to determine the quantitative relationship between structural parameters of the compound and the magnitude of the biological response. In the RASTR package examined the original information for the construction of regression models (RMDL programme) is the information from the FI file. The search for the optimal regression dependence is based on the principle of the structural minimisation of mean empirical risk.[87] The predictive capacity of the model is assessed according to a test selection of data.

Results obtained at the regression analysis stage may be repeatedly used in the RASTR package for designing a more accurate model or for constructing a new model. These data are stored in the FM file. As a result of the file organisation for transfer of information between programmes in the package, the system is highly flexible as regards the inclusion of standard or original methods for processing information at any stage in the study, and may accept all kinds of quantitative structure-activity relationship analysis, the main condition for constructing an advanced expert system.

All programmes in the package have been written in FORTRAN-77 language and are used in a NORD-10 minicomputer (OZU capacity, 128 kilobites).

The programme packages examined indicate that nowadays the software covers a wide range of algorithms for processing structural information and thereby creates the basis for a system approach to the analysis of structure-activity relationships. This objective could be achieved in practice by elaborating a problem-orientated expert system for the construction of compounds with desired properties.

VI. CONCLUSION

In order to search for effective biologically active synthetic compounds in accordance with the basic principles of constructing compounds with desired properties, the following problems must be resolved: 1) identification of the compound interacting with the

receptor; 2) determination of the electron and spatial structure of the compound; 3) the correct description of the structure within the limits of the computer used; 4) derivation of statistically significant data on the biological activity of the compounds; 5) optimisation of a method or group of methods for resolving specific problems related to the determination of structure-activity relationships.

Table 3. Examples of correct predictions of structure-activity relationships.[119]

Type of compound	Form of bioactivity
Benzothiadiazines	Hypotensive action
Propinylamines	Monoamine oxidase inhibition
N-(Phenoxyethyl)-cyclopropyl-amines	Monoamine oxidase inhibition
β-Carbolines	Monoamine oxidase inhibition
Sulphonamides	Carboanhydrase inhibition
Carbanoylpiperidines	Choline esterase inhibition
Ketones	Chymotrypsin inhibition
Benzylpyridinium ions	Inhibition of complement
Steroids	Binding with progesterone receptor
Esters	Papain hydrolysis
Dopamine esters	Hydrolysis of arylamidases
Peptidyl-*p*-nitroanilines	Subtilisin hydrolysis
Polycyclic aromatic hydrocarbons	Hydroxylation of microsomal oxidase
Isatin-β-isothiosemicarbazones	Antiviral carcinogenicity
Pyrazoles	Ditto
Cytidine derivatives	Cytostatic, immunodepressive
Nitrosourethanes	Cytostatic
Mitomycins	Ditto
Hydantoin derivatives	Ditto
Rifamycin amides	Antibacterial
Sulphonamides	Ditto
Erythromycins	Ditto
Quinoxaline-1,4-dioxides	Ditto
Phenanthrenes	Antimycoplasma activity
Phenylamidinourethanes	Ditto
Naphthoquinones	Ditto
Tuberines	Ditto
Methoxychlorines	Insecticidal
N-Hydroxypyridones	Fungicidal
Pyridine derivatives	Ditto
Trifluoromethane sulphonamides	Herbicidal
Alkylarylcarboxylic compounds	Ditto
Aromatic esters	Ditto
Piperidinoacetanilides	Phytoactivity
Pyrimidones	Inhibition of the Hill reaction
Urethanes	Herbicidal
Phenylacetamides	Hypnotic
Pyridylmethanes	Spasmolytic
Pronazines	Neuroleptic
Aromatic and aliphatic acids	Anti-inflammatory
Benzothiopine derivatives	Neuroleptic
Carboxylated esters	Antithrombic
Acrylates	Toxicity
Copper chelates	Cytotoxicity
Pyramine amines	Immunodepressive
Azapurine-6-ones	Immunodepressive
Phenyloxazolidines	Radioprotectant
Carbonoylethylamines	Cholinolytic
Triazines	Dehydrofolate reductase inhibition
Benzdiazepines	Anti-pentylene tetrazolium activity
Cyanocyclohexylamines	Fungicidal
Piperidinoacetanilides	Anti-serotonin activity
Somatostatin analogues	Antidiabetic
Steroids	Progesterone activity
Azacycloalkanes	Analgetic
Organosilicon amines	Fungicidal
Inda-1,3-diones	Toxicity
Sulphones	Antibacterial
Alkobenzylalkylamines	Toxicity
Spirobarbituric acids	Sedative
Adenosine analogues	Hypotensive
Trimethoprim analogues	Antibacterial

An incorrect approach to the resolution of any of the above-mentioned problems could impede the determination of the "true" structure-activity relationship or even lead to the proposal of a false relationship. For example identification of the substrate with a substance introduced into the body (i.e. disregard of the possibility that chemical and biochemical transformations of the chemical may occur during transportation from the site of administration to the bio-target) sometimes causes complications in the detection of common structural fragments among compounds expressing a particular kind of biological activity. Approximation of a model describing the electronic and spatial structure of compounds automatically affects the predictive potential of the model. The efficiency of the latter may also be determined by selecting a method for processing the information, i.e. its compatibility with the form of the presentation and level of information on the structure and properties of the molecules.

There are now many examples in the literature concerning the correct prediction of the biological activity of compounds (Table 3).[19] However, most of these studies are concerned with the optimisation of properties within a series of related compounds. A qualitatively new level in quantitative structure-activity relationship determinations, which would lead to the identification of essentially new classes of effective compounds, would only be attainable by using methods for the selection of common structural fragments within a diverse range of compounds with subsequent optimisation of the useful properties which they express.

This leads to the need to develop a system approach to the determination of structure-activity relationships which should include, in addition to the previously-considered questions, an extended section of chemical data[120] and artificial intelligence.[121] The ultimate stage in a system investigation of a structure-activity relationship could include a plan for synthesis and technological schemes for producing the new compounds.[122,123] Development of this strategy should eventually transform an empirical approach for structure-activity relationship determinations into a fundamental scientific approach exploiting the advances in chemistry, physics, biology, and information technology. A powerful platform will thereby be created for the cost-effective construction of compounds with desired properties for the needs of industry, agriculture, and medicine.

---o0o---

While this review was being prepared for publication a significant number of papers appeared in the literature on studies dealing with virtually all aspects of the question of targeted screening and the construction of effective biologically-active compounds. Worthy of mention are a paper presented at the Sixth European Symposium on Quantitative Structure-Activity Relationships[124] and a monograph[125] in which the fundamentals of the construction of systems for determining the main kinds of bioactivity for large numbers of compounds are presented.

The approach concerned with the three-dimensional perspective of molecular structure by molecular graphics techniques[126–131] based on empirical[132,133] and quantum-chemical[134,135] calculations of molecular conformations or the corresponding X-ray data[136,137] is progressing in an interesting manner. Considerable attention is being focused on the elaboration of models for describing structure with the aid of sets of descriptors[138–149] and on the visual representation of results obtained from the statistical analysis of structural information.[150–153]

Closer examination of the physicochemical content of quantitative structure-activity relationship models[154–157] including the radical changes in programming technology in expert systems[165–171] together with the pharmacodynamic and pharmacokinetic stage[158–164] should be included in the new trends in structure-activity determinations.

The creation of an expert system for quantitative structure-activity relationships could become essential for the formation of a unique "programme" theory for the targeted search and construction of biologically active compounds.

REFERENCES

1. K.Khench, Khim.-farmatsevt. zhurn., 1980, **14**(10), 15.
2. A.S.Glasser, Meth. Find.: Exptl. Clin. Pharmacol., 1984, **6**(10), 563.
3. E.E.Balobanov, Development of Quantitative Concepts Concerning the Relationship Between Structure and Biological Activity of Chemical Compounds, Institute of the History of Natural Science and Technology, Ruk. dep. All-Union Institute of Scientific and Technical Information, No.1639-84, Moscow, 1984.
4. J.Blake, Arch. Gen. Med. III Ser., 1839, **6**, 289.
5. C.Hansch, Drug Metabolism Rev., 1984–85, **15**(7), 1270.
6. B.W.Richardson, Brit. Assoc. Rep., 1864, **34**, 120; 1867, **37**, 47; 1868, **38**, 179.
7. R.Crum-Brown and T.Fraser, Trans. Roy. Soc., 1869, **25**, 151; 1869, **25**, 693.
8. E.J.Ariens and A.M.Simonis, Top. Curr. Chem., 1974, **52**, 1.
9. M.A.Landau, Molecular Mechanisms of the Action of Physiologically Active Compounds, Nauka, Moscow, 1981.
10. "Computer-Assisted Drug Design" (Eds. E.S.Olson et al.), Washington, Chem. Soc., 1979.
11. "Strategy in Drug Design" (Ed. I.A.K.Buisman), Elsevier, Amsterdam, 1982.
12. "Quantitative Structure-Activity Relationships of Drugs" (Ed. J.G.Topliss), Acad. Press, New York, 1983.
13. "QSAR and Strategies in Design of Bioactive Compounds" (Ed. J.K.Seydel, UCH Publ., New York, 1985, 442 p.
14. E.Styuper, U.Bryugger, and P.Dzhurs, "Computer Analysis of the Relationship Between Chemical Structure and Biological Activity" (Translation from English), Mir, Moscow, 1982, p.235.
15. "Drug Design" (Ed. E.J.Ariens), Acad. Press, New York, 1979, 420 p.
16. V.E.Golender and A.E.Rozenblit, "Computer Methods for Drug Design", Zinatne, Riga, 1978, p.232.
17. A.B.Rozenblit and V.E.Golender, "Combinational Logic Methods in Drug Design", Zinatne, Riga, 1983, p.352.
18. S.V.Nizhnii and N.A.Epshtein, Uspekhi Khimii, 1978, **47**(4), 739.
19. A.A.Ovchinnikov and A.I.Boldyrev, Uspekhi Khimii, 1986, **55**, 539.
20. L.N.Balueva, V.M.Zatsepin, and V.K.Promonenkov, "The Use of Mathematical Methods for the Analysis of the Relationship Between Molecular Structure and Pesticidal Activity, Series; Chemical Agents for Plant Protection", Scientific-Research Institute of Technical Chemistry, Moscow, 1985, p.56.
21. O.A.Raevskii, "Introduction to the Design of Biologically Active Compounds", Moscow Chemico-Technological Institute, Moscow, 1984, p.80.
22. O.A.Raevskii and V.Yu.Grigor'ev, Problems in the Course of the "Design of Biologically Active Compounds", Moscow Chemico-Technological Institute, Moscow, 1986, p.80.
23. Z.Simon, Angew. Chem. Int. Ed., 1974, **13**(11), 719.
24. R.C.Hayward, Chem. Soc. Rev., 1983, **12**(3), 285.
25. O.A.Raevskii, Teoret. eksperim. khimiya, 1986(4), 450.
26. V.A.Kritsman, Zhurn. Vsesoyuz. khim. o-va im. D.I.Mendeleeva, 1975, **20**(6), 612.
27. J.M.Lehn, Recherche, 1981, **127**, 1213.
28. R.M.Kellog, Recherche, 1984, **156**, 819.
29. E.Fischer, Chem. Ber., 1894, **27**(3), 2985.
30. T.T.Bertsov and B.F.Korovkin, Biological Chemistry, Meditsina, Moscow, 1983, 137.
31. C.Hansch and T.Fujita, J. Amer. Chem. Soc., 1964, **86**, 1616.
32. V.V.Avidon and L.A.Leksina, NTI, Ser.2, 1974(3), 22.
33. L.S.Gitlina, V.E.Golender, V.V.Drboglav, et al., "Methods of Presenting and Treatment of Structural Information for the Analysis of Structure-Activity Relationships", Preprint, Institute of Organic Synthesis, Academy of Sciences of the Latvian SSR, Riga, 1981, p.74.
34. V.V.Avidon, S.P.Kozlova, and V.S.Arolovich, NTI, Ser.2, 1974(12), 21.
35. W.J.Wiswesser, "A Line-Formula Notation", T.J.Cro. Co., New York, 1954.
36. U.Burket and N.Ellinger, "Molecular Mechanics", Translation from English, Mir, Moscow, 1986, p.364.
37. L.B.Kier and L.H.Hall, "Molecular Connectivity in Chemistry and Design" Acad.Press, New York, 1976, 257 p.
38. J.J.Kaufman, Int. Quant. Chem., 1979, **26**, 221.
39. V.B.Luzhkov and G.N.Bogdanov, Uspekhi Khimii, 1986, **55**(1), 3.
40. J.S.Rollett, J. Mol. Struct., 1985, **130**, 1.
41. A.S.Dimoglo, Khim.-farmatsevt. zhurn., 1985(3), 438.
42. I.Motoc, Lecture Notes in Chemistry, 1980, **15**, 2.
43. O.A.Raevskii and V.P.Novikov, Khim.-farmatsevt. zhurn., 1982(5), 583.
44. K.Burger, Solvation, Ionic Reactions, and Complex Formation in Non-Aqueous Solutions, Translation from English, Mir, Moscow, 1984, p.256.
45. M.K.Kroeger and R.S.Drago, J. Amer. Chem. Soc., 1983, **103**, 3250.
46. A.V.Iogansen, Teoret. i eksperim. khimiya, 1971(7), 302, 312.
47. O.A.Raevskii, V.Yu.Grigor'ev, and V.P.Solov'ev, Khim. -farmatsevt. zhurn., 1984(5), 578.
48. O.A.Raevskii, V.V.Avidon, and V.P.Hovikov, Khim.-farmatsevt. zhurn., 1982(8), 968.
49. I.V.Martynov and O.A.Raevskii, Vestn. Akad. Nauk SSSR, 1983(7), 99.
50. I.V.Martynov and O.A.Raevskii, Dokl. Akad. Nauk SSSR, 1982, **265**(3), 664.
51. I.V.Martynov and O.A.Raevskii, Zhurn. Vsesoyuz. Khim. o-va im D.I. Mendeleeva, 1983, **28**(6), 716.
52. I.V.Martynov and O.A.Raevskii, Dokl. Akad. Nauk SSSR, 1980, **254**(6), 1404.
53. I.V.Martynov and O.A.Raevskii, Dokl. Akad. Nauk SSSR, 1985, **283**(2), 417.
54. L.A.Tyurina, V.A.Semenov, Ch.Sh.Kadyrov, et al., Agrokhimiya, 1986(1), 99.
55. V.P.Novikov and O.A.Raevskii, Khim.-farmatsevt. zhurn., 1982(5), 574.
56. L.J.Soltzberg and C.L.Wilkins, J. Amer. Chem. Soc., 1977, **99**, 439.
57. A.J.Stuper and P.C.Jurs, J. Amer. Chem. Soc., 1975, **97**, 182.
58. H.Kubinyi, Arzneim-Forsch, 1979, **29**(8), 1067.
59. C.Hansch and J.M.Clayton, J. Pharm. Sci., 1973, **62**, 1.

60. L.Szentpaly, J. Amer. Chem. Soc., 1984, **106**, 6021.
61. C.Hansch, Drug Intell. Clin. Pharm., 1982, **16**(5), 391.
62. P.P.Mager, Med. Res. Rev., 1983, **3**(4), 435.
63. C.Hansch, J. Med. Chem., 1976, **19**, 1.
64. W.H.Dekker, H.A.Selling, and J.C.Overeem, J. Agr. Food Chem., 1975, **23**, 785.
65. W.P.Pursell, Eur. J. Med. Chem., 1975, **10**, 335.
66. P.G.Goodford and F.E.Norrington, "Quant. Struct. -Act. Relat." (Ed. M.Tichy), Akad. Kiado, Budapest, 1979, p.79.
67. V.C.Marchevka, N.A.Ebner, K.D.Sehon, and P.E.Hanna, J. Med. Chem., 1985, **28**(1), 18.
68. J.A.Montgomery, J.G.Mayo, and C.Hansch, J. Med. Chem., 1974, **17**, 477.
69. F.Mosteller and G.T'yuki, Analysis of Data and Regression, Translation from English, Finansy i Statistika, Moscow, 1982, Ch.1, 2.
70. G.T.Frumin and V.K.Mukhomorov, Khim. -farmatsevt. zhurn., 1985(3), 180.
71. A.Barth, R.Franke, and M.Orlick, "Quant. Struct. -Activity Analysis" (Eds. R.Franke et al), Akad. -Verlag, Berlin, 1978, p.395.
72. W.P.Dillon and M.Goldstein, "Multivariate Analysis. Methods and Applications", Wiley, New York, 1984.
73. S.Dove, W.J.Streich, and R.Franke, J. Med. Chem., 1980, **23**, 1456.
74. S.Dove, E.Coats, P.Scharfenberg, and R.Franke, J. Med. Chem., 1985, **28**(4), 447.
75. W.J.Streich, S.Dove, and R.Franke, J. Med. Chem., 1980, **23**, 1452.
76. R.Wootton, J. Med. Chem., 1983, **26**, 275.
77. S.Wold and W.J.Dunn, J. Chem. Inf. Comput. Sci., 1983, **23**(1), 6.
78. Yu.A.Zhdanov and V.I.Minkin, Correlation Analysis in Organic Chemistry, State University, Rostov on Don, 1966.
79. K.Iberla, Factorial Analysis, Translation from English, Statistika, Moscow, 1980.
80. M.L.Weiner and P.H.Weiner, J. Med. Chem., 1973, **116**, 655.
81. R.Franke, Farmaco, 1978, **34**, 545.
82. R.Franke, S.Dove, and R.Kühne, Eur. Med. Chem., 1979, **14**, 363.
83. P.N.Craig, J. Med. Chem., 1971, **14**, 680.
84. J.G.Topliss and R.J.Costello, J. Med. Chem., 1972, **15**, 1066.
85. G.Seber, "Linear Regression Analysis", Translation from English, Mir, Moscow, 1980, p.456.
86. I.Moriguchi, K.Komatsu, and Y.Matsushita, Quant. Struct.-Act. Relat., 1984, **3**(3), 106.
87. V.N.Vapnika (Editor), "Algorithms and Programmes for the Reduction of Dependences", Nauka, Moscow, 1984, p.816.
88. P.Dzhurs and T.Aisenauer, Sample Recognition in Chemistry, Translation from English, Mir, Moscow, 1977, p.230.
89. D.R.Henry, P.C.Jurs, and W.A.Denny, J. Med. Chem., 1982, **25**, 899.
90. S.M.Free and J.W.Wilson, J. Med. Chem., 1964, **7**, 395.
91. T.Fujita and T.Ban, J. Med. Chem., 1971, **14**, 148.
92. A.Cammarata and S.J.Yau, J. Med. Chem., 1970, **13**, 93.
93. J.M.Clayton and W.P.Pursell, J. Med. Chem., 1969, **12**, 1087.
94. P.N.Craig, J. Med. Chem., 1972, **15**, 144.
95. J.Kopecky and K.Bocek, Experientia, 1967, **23**, 125.
96. S.A.Aivazyan, Z.I.Bezhaeva, and O.V.Staroverov, Classification of Multidimensional Observations, Statistika, Moscow, 1974, p.238.
97. Y.C.Martin, J.B.Holland, C.H.Jarboe, and N.Plotnikoff, J. Med. Chem., 1974, **17**, 409.
98. D.R.Henry and J.H.Block, Eur. Med. Chem., 1980, **15**, 133.
99. R.S.Nigmatullin, A.L.Osipov, A.P.Puzatkin, and V.A.Koptyug, Khim.-farmatsevt. zhurn., 1985(2), 179.
100. P.C.Jurs, T.R.Stouch, M.Czerwinski, and J.N.Narvaez, J. Chem. Inf. Comput. Sci., 1985, **25**, 296.
101. V.S.Shul'man, Khim.-farmatsevt. zhurn., 1984(9), 1087.
102. B.Duran and P.Odell, Cluster Analysis, Translation from English, Statistika, Moscow, 1977, p.128.
103. C.Hansch, S.H.Unger, and A.B.Forsythe, J. Med. Chem., 1973, **16**, 1217.
104. G.N.Adamson and D.Bawden, J. Chem. Inf. Comput. Sci., 1981, **21**, 204.
105. S.Wold, Pattern Recognition, 1976, **8**, 127.
106. P.Willett, J. Chem. Inf. Comput. Sci., 1984, **24**(1), 29.
107. M.Randic, Int. Quan. Chem. Quan. Biol., 1984(11), 137.
108. "Advances in Drug Research", Vol.14, (Ed. B.Testa), Acad. Press, New York, 1985, 339 p.
109. M.Andy, Comput. Graph. Int. Conf. Proc. L, Pinner, 1983, 435.
110. "Macromolecular Structure and Specificity: Computer Assisted Modelling and Applications" (Eds. B.Kataraghaven et al.), Acad. Sci., New York, 1985, 209 p.
111. A.J.Hopfinger, Pharm. Int., 1984, **5**(9), 224.
112. A.K.Chose and G.M.Crippen, J. Med. Chem., 1985, **28**, 333.
113. C.R.Beddel, Chem. Soc. Rev., 1984, **13**(3), 279.
114. A.J.Hopfinger, J. Med. Chem., 1983, **26**, 990.
115. C.D.Selassie, Z.-X.Fang, R.-L.Li, and C.Hansch, J. Med. Chem., 1986, **29**, 621.
116. V.E.Kuz'min and S.V.Krutius, Khim.-farmatsevt. zhurn., 1986(7), 791.
117. A.N.Tikhonov, A.V.Goncharskii, V.V.Stepanov, and A.G.Yagola, Regularising Algorithms and A Priori Information, Nauka, Moscow, 1983, 200.
118. T.M.Dyott, A.J.Stuper, and G.S.Zander, J. Chem. Inf. and Comput. Sci., 1980, **20**(1), 28.
119. R.Franke, "Theoretical Drug Design Methods", Akad.-Verlag, Berlin, 1984, 415 p.
120. J.E.Rush, J. Chem. Inf. Comput. Sci., 1985, **89**, 4426.
121. E.J.Clementi, Phys. Chem., 1985, **89**, 4426.
122. A.P.Johnson, Chem. in Brit., 1985, **21**(1), 59.
123. M.S.Zefirov and E.V.Gordeeva, Uspekhi khimii, 1987, **56**, 1753.
124. "QSAR in Drug Design and Toxicology" (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, 375 p.
125. G.M.Barenboim and A.G.Malenkov, "Biologically Active Compounds. New Principles for Searching", Nauka, Moscow, 1986, p.363.
126. M.N.Liebman, J. Mol. Graphics, 1986, **4**(1), 61.
127. W.C.Ripka, Nature, 1986, **321**, 93.
128. G.D.Rose and R.H.Lee, Biophys. J., 1986, **49**(1), 83.
129. D.N.J.White and J.E.Pearson, J. Mol. Graphics, 1986, **4**(3), 134.
130. C.Hansch and T.E.Klein, Acc. Chem. Res., 1986, **19**(12), 392.
131. E.F.Meyer and W.Bode, "QSAR in Drug Des. Toxicol" (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.247.
132. J.Tomasi, "QSAR in Drug Des. Toxicol" (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.269.
133. J.Labanowski, I.Motoc, C.B.Naylor, et al., QSAR, 1986, **5**, 138.

134. P.L.Cummins, A.P.L.Rendell, D.J.Swanton, et al., Int Rev. Phys. Chem., 1986, **5**(2/3), 139.
135. Q.Kikuchi, QSAR, 1987, **6**, 179.
136. T.Fujita, QSAR, 1987, **6**,(2), 54.
137. W.G.J.Hol, Angew. Chem. Int. Ed. Engl., 1986, **25**(9), 767.
138. L.B.Kier and L.H.Hall, "Molecular Connectivity in Structure-Activity Analysis", Wiley, New York, 1986, 262 p.
139. L.B.Kier, QSAR, 1986, **5**(1), 7.
140. L.B.Kier, QSAR, 1987, **6**(1), 8.
141. L.B.Kier, QSAR, (3), 117.
142. A.Verloop, "The STREIMOL Approach to Drug Design", Marcel Dekker, New York, 1987.
143. C.R.Kramer, Z. Phys. Chem., 1986, **267**(2), 277.
144. N.Trinajstic, D.J.Klein, and M.Randic, Int. J. Quant. Chem. Quant. Chem. Symp., 1986, **20**, 699.
145. W.J.Dunn III and S.Wold, Trends Anal. Chem., 1986, **5**(3), 53.
146. P.M.King (Editor), "Chemical Applications of Topology and the Theory of Graphs", Translation from English, Mir, Moscow, 1987, p.560.
147. N.Ramaswamy, Chem. Inf. Comput. Sci., 1987, **27**(2), 82.
148. A.K.Ghose and G.M.Crippen, J. Comput. Chem., 1986, **7**(4), 565.
149. G.Young, E.J.Lien, and Z.Guo, QSAR, 1986, **5**(1), 12.
150. J.W.McFarland and D.J.Gans, J. Med. Chem., 1986, **29**(4), 505.
151. M.P.Derde and D.L.Massart, Anal. Chim. acta, 1986, **191**, 1.
152. "Advances in Multivariate Statistical Analysis. 3" (Ed. A.K.Gupta), Reidel, Dordrecht, 1987, 402 p.
153. J.W.McFarland and D.J.Gans, "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.25.
154. R.L.Lopes de Compadre, R.A.Pearlstein, et al., "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.149.
155. L.H.M.Janssen, "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.191.
156. J.J.Morris, L.R.Hughes, and P.J.Taylor, "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.204.
157. A.Goldblum, "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.207.
158. A.Bruns, W.Wahnsiedler, P.Weitzel, et al., "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, 103.
159. M.Tichy, "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.109.
160. G.L.Grunewald, D.J.Sall, and J.A.Monn, "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.118.
161. E.A.Peace, M.G.Rord, R.Greenwood, et al., "QSAR in Drug Des. Toxicol." (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.331.
162. J.E.Leysen and W.Gommeren, Drug dev. Res., 1986(8), 119.
163. S.Balaz, E.Sturdic, and J.Augustin, Biophys. Chem., 1986, **24**, 135.
164. S.E.Jakes and P.Willett, J. Mol. Graphics, 1986, **4**(1), 12.
165. "Artificial Intelligence Applications in Chemistry" (Eds. T.H.Pierce et al.), Amer. Chem. Soc., Washington, 1986.
166. P.B.Ayscough, S.J.Chinnick, R.Dybowski, and P.Edwards, Chem. Ind., 1987, **15**, 515.
167. M.P.Derde, L.Buydens, C.Guns, D.L.Massart, and P.K.Hopke, Anal. Chem., 1987, **59**(14), 1868.
168. E.Darvas, "QSAR in Drug Des. Toxicol., (Eds. D.Hadzi et al.), Elsevier, Amsterdam, 1987, p.37.
169. G.Elty and M.Coombs, "Expert Systems. Concepts and Examples", Translation from English, Financy i Statistika, Moscow, 1987, p.191.
170. P.M.Forsyth (Editor), "Expert Systems. Principles of Operation and Examples", Radio i svyaz, 1987, p.224.
171. F.Kheies-Rot et al.(Editors), "Design of Expert Systems", Translation from English, Mir, Moscow, 1987, p.441.

Institute of Physiologically Active
Compounds, Academy of Sciences
of the USSR, Chernogolovka,
Moscow District