Russian Chemical Reviews **68** (6) 505 – 524 (1999) © 1999 Russian Academy of Sciences and Turpion Ltd

UDC 54.02 + 519.852.6

Molecular structure descriptors in the computer-aided design of biologically active compounds

O A Raevsky

Contents

I.	Introduction	505
II.	Classification of descriptors	506
III.	Element-level descriptors	506
IV.	Structural formula descriptors	506
V.	Descriptors of the electronic structure of molecules	509
VI.	Molecular shape descriptors	509
VII.	Intermolecular interaction descriptors	511
VIII.	Indicator descriptors	515
IX.	Modern computer software for the calculation of descriptors	515
X.	The necessary properties of descriptors	517
XI.	Conclusion	518
XII.	Appendix	518

Abstract. The current state of description of molecular structure in computer-aided molecular design of biologically active compounds by means of descriptors is analysed. The information contents of descriptors increases in the following sequence: element-level descriptors – structural formulae descriptors – electronic structure descriptors – molecular shape descriptors – intermolecular interaction descriptors. Each subsequent class of descriptors normally covers information contained in the previous-level ones. It is emphasised that it is practically impossible to describe all the features of a molecular structure in terms of any single class of descriptors. It is recommended to optimise the number of descriptors used by means of appropriate statistical procedures and characteristics of structure – property models based on these descriptors. The bibliography includes 371 references.

I. Introduction

The empirical regularities in the variation of properties of compounds as functions of their structure established by the middle of the XIX century underlay the concept of the structure – property relationship for chemical compounds. The Periodic Law discovered by D I Mendeleev can be considered as the scientific base of the structure – property concept, and the prediction of physical properties and existence of new elements, unknown at that time, is the first stable predictive model based on this relationship.[1]

Systematic studies on the relationship between the structure and the biological activity of organic compounds started at the end of the XIX century[2–4] and by now, they have become widespread. The recognised founder of the research on the quantitative structure – activity relationship (QSAR) carried out nowadays is C Hansch, who reported in 1964, together with T Fujita, a relation that has found the most extensive use in the studies on the structure – biological activity correlations.[2] However, whilst acknowledging the merits of this scientist, one should also remember his numerous predecessors. In particular, it was noted[3] that as early as 1944, N V Lazarev proposed[4] that the partition coefficients of chemical compounds in oil – water systems be used to establish the quantitative structure – activity relationships. The intense progress in computational hardware and its adoption for research practice, including that in the field of science in question, promoted the formation of a border line of research — computer-aided design of biologically active compounds. Many concepts of this rapidly progressing field are being modified and improved. Therefore, it is necessary to analyse at intervals the current state of studies dealing with this subject. The previous review on this topic[5] was published in this journal in 1988.

With allowance for the structure – activity relationship concept, three main stages can be distinguished in the computer-aided design of biologically active compounds.

(1) Formation of a trial set of compounds having a specified property (activity) or a set of properties;

(2) description of the molecular structure of compounds in this set;

(3) elucidation of the structure – biological activity relationship followed by creation of a stable prognostic model suitable for predicting the properties of new compounds.

This review analyses the modern approaches to the quantitative description of molecular structures of chemical compounds within the framework of computer-aided design of biologically active compounds in terms of independent variables characterising structural features of molecules or their parts (substituents, fragments). These variables are called descriptors.

The progress in computational hardware has a crucial influence on the development of this line of research. It should be noted that computer-aided molecular design faces the problem of processing the data on molecular structures in a unified form with their information content being retained as fully as possible after the data have been entered in the computer.

O.A.Raevsky Institute of Physiologically Active Compounds, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russian Federation. Fax (7-095) 913 21 13. Tel. (7-096) 524 50 62.
E-mail: raevsky@ipac.ac.ru

Received 16 November 1998
Uspekhi Khimii **67** (6) 555 – 576 (1998); translated by Z P Bobkova

Complex natural processes are studied by developing their models. Advantages of these models over the original are that they are easier to investigate, can be better manipulated and monitored and that the investigations are cheaper.[6] The models can be classified into iconic (ensuring similarity regarding the form), analog (ensuring similarity regarding the form and functions) and abstract (symbolic, conceptual) models, which provide the correlation with the original by virtue of symbols.

The modern views on the molecular structure are certainly not exhaustive and, in the future, they will be improved with the aid of various models. Here, it is pertinent to quote M Dewar, who said that we don't know and, perhaps, shall never know what molecules actually are. Our understanding of molecules is based on models, which reproduce their properties sufficiently well to be useful.[6]

Description of models, which are used in chemical research, can be found in the literature.[6–8] In this review, molecular structure descriptors are considered briefly with allowance for the above classification of models.

II. Classification of descriptors

Various types of classification of descriptors are possible. Three types of classification are considered in a paper by Charton.[9] According to the first type, descriptors are classified into simple (describing one effect of interatomic interactions) and composite ones (describing two or more effects). Composite descriptors, in turn, can be divided into unicomposite (describing effects of the same type) and multicomposite descriptors (reflecting the sum of effects of various types).

The second classification is based on the way in which descriptors are estimated, *i.e.*, experimental and theoretical ones.

The third classification takes into account the sort of effect described by a given descriptor. Three sorts of effects are distinguished — electronic (electric), steric and intermolecular effects. A specific case of using this type of classification is provided by the classical Hansch equation,[2,10–14] which represents biological activity as a function of Hammett constants (electronic effects), Taft constants (steric effects) and hydrophobicity constants (transport effects arising due to intermolecular interactions).

For visualisation of the description of chemical structures, Testa and Kier[15] used geometrical figures corresponding to the level of structure being considered: element level → two-dimensional structure → three-dimensional structure → bulk properties → stereodynamic structure → stereoelectronic structure → interactions with the environment. The information content of each subsequent level embraces the information content of the preceding one. Apparently, a similar approach in a somewhat different form could be used in this review to discuss descriptors. Figure 1 shows a chart for several levels of descriptors in which each subsequent level covers the information content of the preceding one. In this representation, it is assumed that structural formula descriptors carry the whole body of information contained in the element-level descriptors; electronic structure descriptors carry the information contained in the structural formula descriptors and descriptors of intermolecular interactions cover the whole body of information provided by all of the preceding levels. In some cases, descriptors do not bear the information of the previous level (*e.g.*, some descriptors of intermolecular interactions such as polarisability and hydrophobicity are often calculated without allowance for the three-dimensional structure), but in general, this approach to the classification of descriptors is visual and reflects the degree of their complexity and, thus, it is convenient for the analysis. In this review, each class of descriptors is considered. It should be noted that a descriptor of any level can characterise both a molecule as a whole and its part (fragment, screen, functional group, substituent).

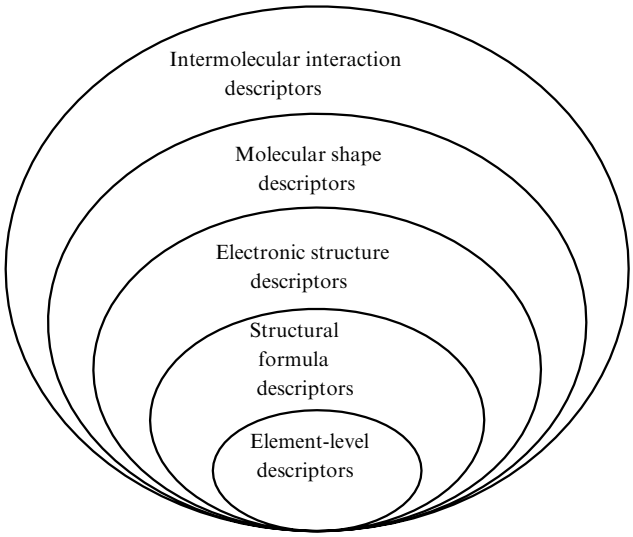


Figure 1. Information contents of molecular structure decriptors.

III. Element-level descriptors

The gross formula bears the information about sorts of atoms that make up the molecule and the number of atoms of each sort. The only property that can be predicted precisely based on the gross formula is molecular weight (MW), which can be regarded as an element-level descriptor. The number of atoms of a particular sort is sometimes taken as another descriptor of this level.[16] An example of using such descriptors can be found in the Appendix [see Eqn (A.1)].

A legitimate construction of the structure–property correlation should be carried out for related compounds. Therefore, the structure variation within a series of compounds being considered reduces to variation of substituents. Hence, the properties of substituents can be regarded as structural descriptors. In particular, the atomic weights of substituents can also be considered to be element-level descriptors. Generally, descriptors of this level carry too little information on the molecular structure and, hence, they cannot be used alone for elucidation of the real structure–property relationships.

IV. Structural formula descriptors

The structural formula of a molecule is the most widespread way of describing a chemical compound. As a model of molecular structure, this formula contains elements of iconic, analog and abstract models[6] and provides the basis for the construction of many types of descriptors, most of all, topological descriptors (Fig. 2).

1. Topological indices

Topological descriptors (indices) are calculated based on the description of a structural formula of a compound by a molecular graph G, which is a two-dimensional representation of the molecule (the vertices correspond to atoms and the edges match the chemical bonds).[17] Normally, the skeleton atoms and the bonds between them are considered, hydrogen atoms being erased. The matrix form of the graph is used to estimate the topological indices. The adjacency matrix $A(G)$ and the distance matrix $D(G)$ are used most often. The adjacency matrix elements a_{ij} are equal to either unity or zero depending on whether or not vertex i of graph G is incident to vertex j. The number of units in an ith line or jth column of the matrix is equal to the vertex degree.

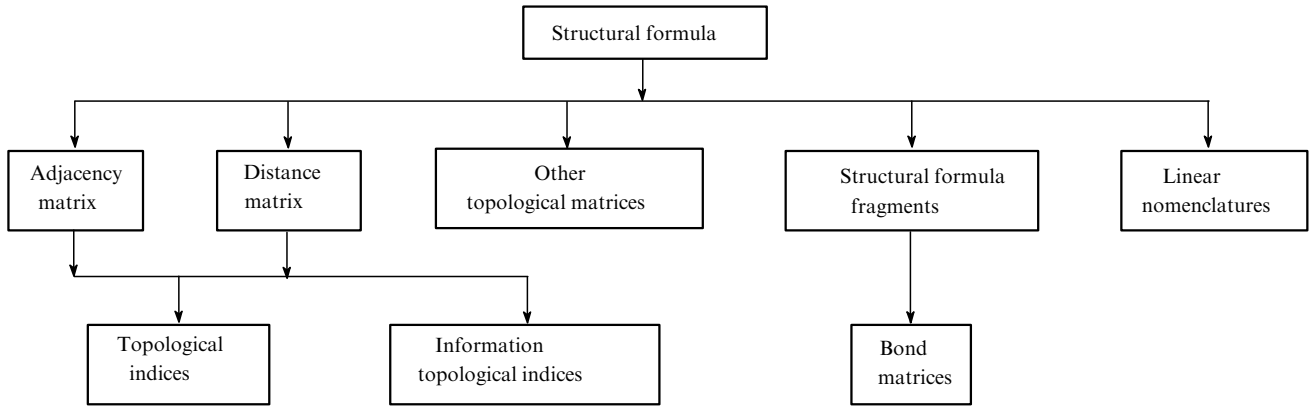
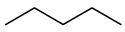
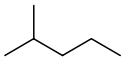


Figure 2. Elements of molecular structure description based on the structural formula.

Each element in the distance matrix $D(G)$ represents the number of edges connecting vertex i to vertex j *via* the shortest route and is denoted by d_{ij}. The adjacency and distance matrices for pentane, 2-methylpentane, cyclopentane and furan are shown in Scheme 1. Note that the adjacency and distance matrices for the two last-mentioned compounds are identical because the molecular graphs are identical.

Scheme 1

Compound	Molecular graph	Adjacency matrix	Distance matrix
$CH_3(CH_2)_3CH_3$		1 2 3 4 5 1 0 1 0 0 0 2 1 0 1 0 0 3 0 1 0 1 0 4 0 0 1 0 1 5 0 0 0 1 0	1 2 3 4 5 1 0 1 2 3 4 2 1 0 1 2 3 3 2 1 0 1 2 4 3 2 1 0 1 5 4 3 2 1 0
$CH_3CH(CH_3)(CH_2)_2CH_3$		1 2 3 4 5 6 1 0 1 0 0 0 0 2 1 0 1 0 0 1 3 0 1 0 1 0 0 4 0 0 1 0 1 0 5 0 0 0 1 0 1 6 0 1 0 0 1 0	1 2 3 4 5 6 1 0 1 2 3 4 2 2 1 0 1 2 3 1 3 2 1 0 1 2 2 4 3 2 1 0 1 3 5 4 3 2 1 0 4 6 2 1 2 3 4 0
Cyclopentane		1 2 3 4 5 1 0 1 0 0 1 2 1 0 1 0 0 3 0 1 0 1 0 4 0 0 1 0 1 5 1 0 0 1 0	1 2 3 4 5 1 0 1 2 2 1 2 1 0 1 2 2 3 2 1 0 1 2 4 2 2 1 0 1 5 1 2 2 1 0
		1 2 3 4 5 1 0 1 0 0 1 2 1 0 1 0 0 3 0 1 0 1 0 4 0 0 1 0 1 5 1 0 0 1 0	1 2 3 4 5 1 0 1 2 2 1 2 1 0 1 2 2 3 2 1 0 1 2 4 2 2 1 0 1 5 1 2 2 1 0

The most commonly used topological indices calculated from the adjacency matrix are the Platt index $F(G)$ (the sum of degrees of each edge in graph G); the Gordon–Scantlebury index $Y(G)$ (the number of paths of length 2); the full adjacency index $A'(G)$ (the sum of all the non-zero elements in the adjacency matrix); the Randic connectivity index $\chi(G)$ (it will be discussed below); and the Zagreb group indices $M_1(G)$ and $M_2(G)$. The indices that should be mentioned as the descriptors calculated from the distance matrix are the Wiener index $W(G)$, the Hosoya index $Z(G)$, the polarity number $P(G)$, the distance sum index $S(G)$ and the index of distances between the vertices VDI(G). The formulae for the calculation of these widely used descriptors and other topological indices estimated based on the adjacency and distance matrices as well as examples of using these descriptors in the structure–property correlations can be found in original publications.[18–20]

The above matrices and topological indices reflect only topological features of molecules without allowance for the types of atoms and the real distances between them. The imperfection of this description of the structures of compounds becomes clear if one recalls that the adjacency and distance matrices (and, hence, all the topological indices mentioned above) for cyclopentane and furan coincide.

In addition to the adjacency and distance matrices described above, the extended distance matrix E,[21] connectivity matrix C[22] and Wiener matrix W[23] are used.

Many researchers use various descriptors estimated based on connectivity matrix C; therefore, we shall briefly consider them here. In 1975, Randic proposed that the connectivity index $\chi(G)$ be determined in the following way:[24]

$$\chi(G) = \sum(v_i v_j)^{-1/2}, \quad (1)$$

where v_i and v_j are the degrees of vertices i and j in graph G, respectively. Summation is carried out over all edges of graph G.

Kier and Hall[25] have employed the Randic scheme as the basis for the development of a general method for the description of structures of organic compounds and proposed the concept of molecular connectivity for this purpose.[25] The first equation for a topological index was as follows:

$$^1X = \sum(\delta_i \delta_j)^{-1/2}, \quad (2)$$

where δ corresponds to the number of neighbouring atoms (with hydrogen atoms being omitted). This equation is identical to Eqn (1). However, the researchers made this approach more versatile by extending it not only to bonds but also to atoms ($^0X = \sum(\delta)^{-1/2}$) and to fragments with several bonds. For example, when a molecular graph is cut into two-bond fragments, the index $^2X = \sum(\delta_i \delta_j \delta_k)^{-1/2}$ is estimated, when it is cut into three-bond fragments, the index $^3X = \sum(\delta_i \delta_j \delta_k \delta_l)^{-1/2}$ is determined and so on. In addition, calculation of these indices for molecules containing heteroatoms was started. The hydrogen atoms attached to heteroatoms were ignored (like those attached to carbon).

Later, Kier and Hall[26] proposed a general equation for the estimation of δ for any atom

$$\delta_v = Z_v - h, \quad (3)$$

where Z_v is the number of valence electrons with allowance for the electrons of lone electron pairs and h is the number of hydrogen atoms attached to the given atom (which are present in the structural formula but lacking in the molecular graph). Thus the δ_v value for a carbon atom of benzene is equal to three (the atom has four valence electrons and is linked to one hydrogen atom); the δ_v value for the oxygen atom in alcohols is equal to five (the atom

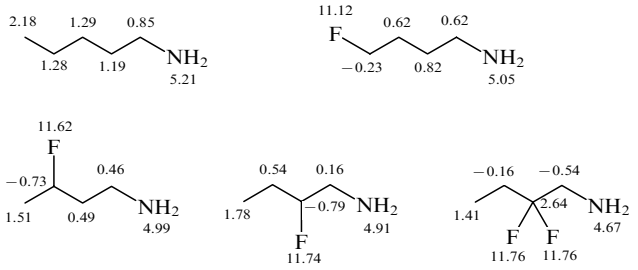
has six valence electrons and is linked to one hydrogen atom) and that for the oxygen atom in ethers is six (the same number of valence electrons but no hydrogens at the heteroatom).

In the subsequent studies, Kier and Hall [27–30] proposed using electronic state indices (S) to describe quantitatively the mutual influence of atoms. The S value for any atom is estimated by the formula

$$S = I + \Delta I, \tag{4}$$

where $I = (\delta_v + 1)/\delta$ and $\Delta I = (I_i - I_j)/r_{ij}^2$ (r is the number of bonds between the two atoms considered). The indices of the electronic topological state for some substituted amines are presented below (Scheme 2).

Scheme 2



The Kier and Hall indices have been used rather widely in the search for quantitative structure–property relationships [see, for example, the Appendix, Eqns (A.2) and (A.3)]. These indices became very popular because they can be easily calculated and provide clear representation of the structural formulae of compounds. Several recent publications should be noted.[31–39]

However, description of the electronic structure by the number of valence electrons and description of interatomic distances by the number of bonds between the atoms are too simplified to ensure elucidation of the true molecular structure. This accounts for the criticism levelled at the attempt of using these indices for establishing the real structure–biological activity correlations.[40]

A commonly used index calculated from the distance matrix is the Wiener index $W(G)$, which is the first topological index used in chemistry. It is defined as the half-sum of the elements of the distance matrix D. The Wiener index and related topological indices are often used in the structure–property studies to characterise the compactness of a molecule.[41–43] However, it should be borne in mind that in matrix D, the distance between atoms is expressed as the number of bonds separating these atoms. In order to bring the structure description in terms of the distance matrix closer to the real situation, it has been proposed to use averaged (typical) distances between atoms in molecules placed in a three-dimensional grid [44–47] or distances found from experimental data or by quantum-chemical calculations [48, 49] as the off-diagonal matrix elements.

Molecular graphs serve as the source of diverse topological indices. Thus Kier [50–54] proposed various kappa-indices for coding the shape or configuration of molecules. Description of some other new indices can be found in several publications.[55–64] Particular correlations of these topological indices with various physical properties and biological activity have been considered.[65–72]

Information topological indices have also been used for structure description. The estimation of these indices is based on the separation of the set of graph vertices into equivalence classes taking account of the closest environment of the vertices.[18, 73] Within the framework of this approach, a number of topological indices of the theoretical-information type have been devised, namely, ^{1}IC (information content), ^{1}SIC (structural information content), ^{1}BIC (binding information content) and ^{1}CIC (complementary information content), which are calculated from the following equations:

$$^1\mathrm{IC} = -\sum \frac{n_i}{n} \log_2 \frac{n_i}{n}, \tag{5}$$

$$^1\mathrm{SIC} = \frac{^1\mathrm{IC}}{\log_2 n}, \tag{6}$$

$$^1\mathrm{BIC} = \frac{^1\mathrm{IC}}{\log_2 q}, \tag{7}$$

$$^1\mathrm{CIC} = \log_2 n - {}^1\mathrm{IC}, \tag{8}$$

where n_i is the number of vertices in the ith equivalency class, n is the number of graph vertices for the molecule and q is the number of graph edges. Examples of using the information topological indices for describing structures of organic compounds in elucidation of structure–property correlations including structure–biological activity correlations have been reported.[74–80]

2. Encoding of structural formulae by linear nomenclatures and bond matrices

The necessity of automating the processing of information on the structures of chemical compounds, the amount of which has sharply increased in recent years, initiated the development of some new methods for representation of the structural formulae of chemical compounds, namely, various linear nomenclatures and bond matrices, which can be regarded as abstract models of molecules.[6]

According to the linear nomenclature, a structural formula is encoded by a string of characters, which, in turn, encode particular fragments of the structure. Each linear nomenclature is characterised by a definitely specified sequence of characters. The character set of a linear notation usually includes standard characters. The initial stage of development of linear nomenclatures is associated with the name of Wiswesser. As an example, we present the Wiswesser codes for the structures of some compounds:

$CH_3-CH_2-O-CH_2-CH_2-CH_3$ (3O2),

$H_2N-CH_2-CH_2-C(O)OH$ (Z2VQ),

$HOCH_2CH(NH_2)CH_2CH_2CH(CH_2NH_2)CH_2OH$ (Z1Y1Q2YZ1Q).

The Wiswesser linear nomenclature has been described in detail.[16]

In Russia, a similar approach is practised by Avidon,[82, 83] who uses the SSFC language (substructure superposition fragmental code).

At present, the most widespread linear nomenclature used to represent structural formulae of organic compounds is contained in the SMILES system,[84–86] which is used, in particular, to reveal the structural similarity of compounds.[87] A linear nomenclature suitable for the description of not only the structural formula but also the stereoisomerism also deserves attention.[88] In this nomenclature, a compound is represented by a canonical line code consisting of two parts; the first part describes the structural formula and the second part characterises the stereochemical features of the molecule.

When a structural formula is described using a connectivity matrix, each atom, each bond and each type of atoms are coded directly. Connectivity matrices are similar to adjacency matrices and distance matrices. The main diagonal of a connectivity matrix normally includes codes for the atoms in the structural formula, while off-diagonal elements are the codes for the types of bonds. These matrices were used rather widely in the 1970s and 1980s; they have been described in a monograph.[16] A connectivity matrix of descriptor centres (CMDC) has been proposed;[89] in this matrix, the codes of atoms are replaced by codes of functional groups (descriptor centres) and the off-diagonal elements correspond to the numbers of bonds between the descriptors instead of the bond codes. In order to perfect the language of structure description in terms of CMDC, later it has been proposed to use computable distances between the descriptor centres.[90]

3. Structural fragments as descriptors

The topological indices and various matrices described above provide certain information on a chemical compound as a whole. However, the crucial role in the interaction of a compound with other chemical and biological objects is sometimes played by separate fragments of molecules; therefore, in the elucidation of the structure–property and structure–biological activity correlations, the structures of compounds of a given series are often described fragment-by-fragment. A structural fragment can comprise one atom, a substituent, a functional group or a particular combination of atoms (with indication of the hybridisation types).[91] Each structural fragment can be used as an independent descriptor. Further detailing can be attained by specifying the local characteristics of atoms forming the fragment (see, *e.g.*, classification of the descriptor centres in terms of their donor–acceptor properties).[92,93] Numerous approaches to the computer-aided construction of new substances based on structural fragments of compounds having a common property have been developed. Various substructure analysis descriptors,[94–97] topological pharmacophores in the LOGANA programme package procedure,[98–100] substructure descriptors,[101,102] descriptors of the structural language of the DARC system,[103,104] chemical functionality descriptors [105] and *de novo* design descriptors [106] can be cited as examples.

V. Descriptors of the electronic structure of molecules

Quantum-chemical calculations for complex molecules make use of atomic and molecular quantum-chemical descriptors. The methods for calculation, the potential and limitations of these descriptors have been described in detail in a review.[107] In this Section, we consider briefly only quantum-chemical descriptors related to intramolecular electronic properties.

Atomic charges are quantum-chemical descriptors used most often to elucidate the structure–property correlations. If chemical interactions are divided into electrostatic and covalent components, atomic charges characterise the electrostatic component. The full charge on an atom can be regarded as a non-directional descriptor, whereas the σ- and π-electron densities characterise the possible orientation of chemical interactions; therefore, they can be regarded as directional descriptors.[108] Examples of successful use of these descriptors can be found in the literature.[109–111] When the charges on atoms are known, a new set of descriptors of this type can be used in each particular study.

The energies of the highest occupied (E_{HOMO}) and lowest unoccupied (E_{LUMO}) molecular orbitals are descriptors that are frequently used in calculations of electronic structures of molecules.[112–114] It was found that in some cases, the E_{HOMO} values are directly correlated with the ionisation potential (I) and characterise the susceptibility of the molecule to electrophilic attack. The E_{LUMO} values correlate with the electron affinity (A) and characterise the susceptibility of the molecule to nucleophilic attack. Note that I and A are also used as independent electronic structure descriptors.[115,116] The $E_{HOMO} - E_{LUMO}$ difference is believed to reflect the molecule's stability.[117] The notions of softness and hardness of molecules are associated with exactly these electronic structure descriptors.[118] The molecular hardness descriptor (η)

$$\eta = \frac{I - A}{2} \qquad (9)$$

has also found use in the structure—property correlations (see, for example, Ref. 119).†

† Hardness as a measure of energy stabilisation has been defined by some researchers as the half-sum of the E_{HOMO} and E_{LUMO} values [120] or as the difference between E_{HOMO} and E_{LUMO} taken with a minus.[121,122]

Dipole moment and related descriptors. The polarity of a molecule is described by its dipole moment (μ). The magnitude of this descriptor can be determined experimentally or calculated using a quantum-chemical method. This descriptor is used rather widely in elucidation of structure–property correlations.[12,112,123—125] However, the vector summation of the dipole moments of several polar groups in a molecule can result in a small value. Therefore, in addition to μ, the greatest difference between the charges of two atoms, the topological electronic index, the local dipole index, the quadrupole moment tensor, *etc.* are sometimes used as polarity descriptors.[107]

Examples of using electronic structure descriptors in structure–biological activity relationships are given in the Appendix [Eqns (A.4) and (A.5)].

VI. Molecular shape descriptors

It is now obvious that features of the three-dimensional structure of molecules need to be taken into account in elucidating the structure–property relationsips. Suffice it to say that almost all of the modern automated systems for the search for new compounds contain, as obligatory elements, various procedures for comparison of the spatial structures of molecules.[126,127]

The description of the three-dimensional structure of a molecule by a single quantitative descriptor (or a set of descriptors) is a fairly complicated task. Let us consider the simplest situation when the compounds under consideration have the same framework and different substituents at the same position. In 1976, Verloop and coworkers proposed a set consisting of five steric constants (STERIMOL parameters) for 243 different substituents; the constants were calculated based on standard bond lengths and bond angles.[128] These simple parameters have found fairly wide use in structure–property studies.[129–131] A similar approach based on estimation of shape parameters of substituents has been proposed in another study.[132]

Amoore [133] was the first to analyse the projections of molecular structures onto the three orthogonal planes in a study dealing with the relationship between the structure and odour of compounds. An approach describing the shape of a molecule in terms of six descriptors $S_1 - S_6$ has been proposed.[134]

The main problem arising in the comparison of 3D structure parameters of various molecules is associated with conformational mobility. The appearance of an additional internal rotation axis results in an exponentially increased number of possible conformations; therefore, certain restrictions are required as well as special procedures for selecting real conformations and comparing their shapes.

Several methods for the estimation of molecular shape descriptors currently exist. These methods are briefly described below.

The distance geometry approach has been developed since the late 1970s, being constantly modified by Crippen *et al.*[135–141] The consecutive steps of the procedure used to compare the 3D molecular shapes can be outlined as follows.

1. Generation of 3D structures for the compounds considered; the initial structure is based on the bond lengths and angles found in an X-ray diffraction experiment.

2. Ascribing of definite physico-chemical properties to atoms in the molecules.

3. Elucidation of low-energy regions in the conformational space for each of the compounds.

4. Determination of the geometrically possible superpositions of molecules onto the reference structure. First, all combinations of three atoms in the reference structure and in the molecule being analysed are considered. After the best correspondence of distances between some three-atom groups in the pair of compounds being compared has been found, superposition of these three pairs of atoms is fixed and the process of spatial superposition is continued for other atoms.

5. Analysis of a great number of 'good' superpositions of compounds on the reference structure makes it possible to compare the molecular shapes of the compounds; as this is done, the concept of the molecular shape descriptors is visualised and the shape of the binding site in a potential receptor is evaluated.

The binding site approach.[142–144] This approach represents the binding sites of a potential receptor as non-overlapping areas covering the whole space rather than as points (as in the previous method). With such representation of a spatial structure, each atom should be located in one (and only one) area and the binding site model should include the list of areas (descriptors) in which each atom is located. Fig. 3 *a* shows the binding site model calculated for compounds exhibiting competitive properties in relation to the protein isolated from mouse liver,[145] and Fig. 3 *b* presents the optimum binding model for benzopyrene.

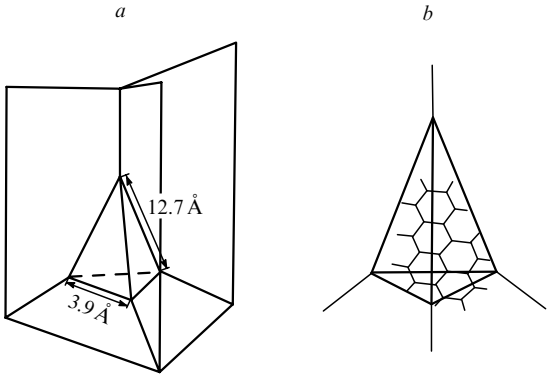


Figure 3. Binding site model for compounds exhibiting competitive properties towards a protein (*a*) and optimum binding model for benzopyrene (*b*).[145]

Molecular shape analysis. This approach has been developed during the last 15 years in the studies of Hopfinger *et al.*[146–155] The procedure for the quantitative comparison of the spatial structures of the molecules of various compounds includes six consecutive stages.

1. Conformational analysis with allowance for the internal rotation about all the single bonds and calculation of the energy for each 10° increment. This is done using methods of molecular mechanics with a force field comprising contributions of dispersion, steric and electrostatic interactions and, whenever possible, hydrogen bonds.

2. Selection of individual conformations. The real set of possible conformations for each compound is determined by the chosen difference between the conformation considered and the global energy minimum.

3. The choice of the reference compound and the active conformation by pairwise comparison of the molecular shapes in the series of compounds under consideration.

4. Superposition of each molecule in the series of compounds under consideration onto the molecule of the reference compound.

5. Quantitative comparison of molecular shapes. The following descriptors characterising the molecular shape similarity have been proposed:

the common overlap steric volume (V_0) determined from the following equation

$$V_0 = V_u \cap V_v, \tag{10}$$

where V_u and V_v are the space occupancy by the molecules of the compound considered and the reference compound, respectively;

the relative measure of the degree of similarity defined as the function

$$f_0 = \frac{V_u \cap V_v}{V_u}. \tag{11}$$

These descriptors do not contain information regarding the intramolecular stability of each conformation of the pair of compounds. Therefore, one more descriptor, the shape commonality index (I_c), has been proposed:

$$I_c = S_{(u,v,w)} - \frac{\Delta E_u}{[\Delta E_u^*(\Delta E_u^* + \Delta E_v^* + \varepsilon)]^{1/2}} + \tag{12}$$

$$+ \frac{\Delta E_v}{[\Delta E_v(\Delta E_u^* + \Delta E_v^* + \varepsilon)]^{1/2}},$$

where $S_{(u,v,w)}$ is wV_0 or wf_0, w is the shape similarity parameter, ΔE is the difference between the conformation energies, $\Delta E_v^* = 1$ kcal mol^{-1}, ΔE_u^* is taken to be 1, 3 and 6 kcal mol^{-1} in three separate calculations and ε is a correction factor.

The hypothetical active site lattice.[156–160] This method is based on the assumption that a molecule can be described by descriptors characterising the space that it occupies and the nature of atoms that occupy this space. Three-dimensional space is defined as a set of regularly arranged points in a three-dimensional lattice. Translation of the molecule to the set of these points occurs with retention of those points of the lattice that fall within the van der Waals radius of each atom in the molecule. The resulting molecular lattice consisting of points carries information on the type and properties of atoms in the molecule. The type of atoms is defined simply: the value +1 is attributed to atoms rich in electrons, the value −1 is attributed to atoms poor in electrons and electrically neutral atoms are denoted by 0.

Estimation of molecular positions.[161–163] In this approach, the comparison of spatial shapes of molecules also starts with the search for low-energy conformations and selection of energetically stable shapes for each compound in the series under consideration. Molecular position is defined as a specific conformation associated with the activity exhibited by this compound. The molecular position is described by several descriptors.

The weighted holistic invariant (WHIM) molecular descriptors are constructed in such a way as to cover the information on the three-dimensional structure as fully as possible including the size, shape and symmetry of molecules and the distribution of atoms in them.[164–169] The calculation algorithm includes the analysis of the main components of the centred molecular coordinates using the weighted covariant matrix derived from various weighted schemes for atoms.

Different covariance matrices and principal axes are obtained depending on the type of the weighted scheme used. For example, when the atomic weights are used as the weighted scheme, the directions of the three principal axes coincide with the principal inertia axes.

For each weighted scheme, a set of statistical indices of atoms projected onto each principal component t_m ($m = 1, 2, 3$) is calculated. The invariance of the calculated parameters with respect to translation is attained by the procedure of centring the atomic coordinates, while the invariance with respect to rotation is ensured by using the principal components.

The directional WHIM descriptors are statistical indices calculated based on each individual principal component (1, 2, 3). The non-directional WHIM descriptors are found directly from directional descriptors. The latter type of descriptors characterise the overall shape of the molecule, any information related to the principal axes being lost in this case.

Eleven directional descriptors (λ_1, λ_2, λ_3, ϑ_1, ϑ_2, γ_1, γ_2, γ_3, η_1, η_2, η_3) characterising the molecular properties within each weighted scheme have been proposed.[164–169] Thus, the total number of directional WHIM descriptors is 66. In the case of planar structures, only eight directional (λ_1, λ_2, ϑ_1, ϑ_2, γ_1, γ_2, η_1, η_2) and

five non-directional descriptors (T, A, V, K, D) should be taken into account for each weighted scheme.

Analysis of the shape of electron cloud.[170—174] This method is based on the assumption that the shape of the electron cloud carries the whole body of information on the properties of compounds. This approach was used primarily to analyse molecular surfaces. The surface of a molecule is cut into areas according to the local curvature properties. Convex D_2, saddle-like D_1 and concave D_0 local areas are distinguished. The ratios of these descriptors determine groups of molecular shapes, which include two- one- and zero-dimensional truncated surfaces obtained from the initial surface by removing all areas of a given type (D_2, D_1 or D_0). The quantitative measure for the similarity of shapes of 3D molecules can be found by comparing the numerical codes of shapes. This procedure is performed automatically using a computer without subjective visual comparison of the shapes of objects.

Examples of using molecular shape descriptors in structure–biological activity correlations are given in the Appendix [Eqns (A.9) and (A.12)].

VII. Intermolecular interaction descriptors

Quantitative description of the intermolecular interactions in the elucidation of structure–biological activity correlations can be accomplished using the free energy difference (ΔG).[175, 176]

$$\Delta G = G_f - G_i, \tag{13}$$

where G_f and G_i are the free energies of the final and initial states, respectively.

A more detailed quantitative description of intermolecular interactions is provided by using the enthalpy (ΔH) and entropy (ΔS) differences, which are related to ΔG as follows:

$$\Delta G = \Delta H - T\Delta S, \tag{14}$$

where T is the reaction temperature. By measuring experimentally the equilibrium constant (K), one can determine ΔG

$$\Delta G = \Delta G^\circ - RT\ln K, \tag{15}$$

where ΔG° is the free energy change for a reaction under standard conditions.

The ΔG, K, ΔH and ΔS values for a reaction or interaction can be regarded as descriptors of intermolecular interactions. For example, dissociation constants of acids and bases (pK_a) are used fairly frequently in the modelling of structure–activity correlations.[177–180]

A large group of intermolecular interaction descriptors estimated from experimental data on reactivity comprises various constants describing the electronic influence of substituents on the reactivity of chemical compounds. This line of studies was initiated by Hammett who defined the parameter σ (Hammett constant) based on the relation [181]

$$\sigma_X = \log K_X - \log K_H, \tag{16}$$

where K_X is the corresponding constant for *meta*- (σ_m) or *para*-substituted (σ_p) benzoic acid, K_H is the ionisation constant for benzoic acid in water at 25 °C. By now, a great number of constants have been proposed to describe the electronic effects of substituents. For example, the inductive (σ_i) and resonance (σ_r, σ^+, σ^-) constants, the inductive Taft constant (σ^*) and the Swain and Lupton constants [182–184] are widely used as descriptors of intra- and inter-molecular interactions.

The reactivity descriptors, calculated by quantum-chemical methods, constitute an important group of intermolecular interaction descriptors. First of all, mention should be made of indices of electrophilic (S_e) and nucleophilic (S_n) superdelocalisabilities on a reaction centre (r), which are estimated from the equations

$$S_e(r) = 2\sum_k^{vac}\sum_{\sigma(r)}\frac{(C_{\sigma k})^2}{0.5(E_{HOMO}+E_{LUMO})-E_k}, \tag{17}$$

$$S_n(r) = 2\sum_i^{occ}\sum_{\sigma(r)}\frac{(C_{\sigma i})^2}{E_i-0.5(E_{HOMO}+E_{LUMO})}, \tag{18}$$

where the numerators contain coefficients for the atomic orbitals and denominators contain the energies of the corresponding molecular orbitals.

Some other descriptors such as self-atom polarisability $\pi(r)$,[185] reflecting the influence of the variation of electronegativity of atomic orbitals on the charge of the same atom

$$\pi(r) = 4\sum_i^{occ}\sum_k^{vac}\sum_{\sigma(r)}\frac{C_{\sigma i}^2 C_{\sigma k}^2}{E_k-E_i}, \tag{19}$$

and atom–atom polarisability $\pi(AB)$,[107] reflecting the effect of a perturbation at one atom (A) on the charge of another atom (B)

$$\pi(AB) = 4\sum_i\sum_a\sum_p\sum_r\frac{C_\pi^A C_{pa}^A C_{ri}^B C_a^B}{E_i-E_a}, \tag{20}$$

can be used as well to describe intermolecular interactions.

These descriptors have also found application in the description of structure.[107, 186–188]

Molecular refraction and polarisability. When a molecule is placed in an external electrostatic field, the charge distribution in it changes. The resultant dipole moment is determined from the relation

$$\mu = \alpha E, \tag{21}$$

where α is the molecular polarisability and E is the electric field intensity. The polarisability can be estimated experimentally based on the expression relating it to the molecular refraction (MR)

$$MR = \frac{4\pi N_A \alpha}{3}, \tag{22}$$

(N_A is the Avogadro number) and the relation of the molecular refraction to the refractive index (n)

$$MR = MV\frac{n^2-1}{n^2+2}, \tag{23}$$

where MV is the molecular volume. The MR and α values are estimated either based on the additivity of the experimentally determined contributions of separate atoms and fragments [189] or by quantum-chemical calculations.[191–194]

Lipophilicity. The group of intermolecular interaction descriptors used most commonly to analyse the structure–activity correlations is related to the notion of lipophilicity, which characterises the transport properties of compounds in biological objects. In the vast majority of cases lipophilicity is described quantitatively in terms of the partition coefficients (P) of compounds in the model octanol–water system. The $\log P$ value for the neutral form of the compound is used as the descriptor for this system. The corresponding descriptors for the cationic, anionic and zwitter-ion forms are designated by $\log P^+$, $\log P^-$ and $\log P^{+/-}$.[195] The partition of ionised compounds between octanol and water depends on pK_a of the compound itself and pH of the medium. The distribution of substances in biological media is characterised by the index $\log D$ with indication of the pH, instead of $\log P$. In the vast majority of publications, $\log D$ refers to a physiological medium with pH 7.4.

The relationship between the descriptors $\log D$, $\log P$, $\log P^+$, $\log P^-$ and pK_a is given by the following expressions:[196]
for the equilibrium $XH \rightleftharpoons X^- + H^+$,

$$\log D = \log P_{XH} - \log(1 + 10^{-pK_a+pH}); \tag{24}$$

for the equilibrium $X + H^+ \rightleftharpoons XH^+$,

$$\log D = \log P_X - \log(1 + 10^{-pK_a-pH}); \tag{25}$$

for the equilibria $X^- + H^+ \rightleftharpoons XH + H^+ \rightleftharpoons XH_2^+$,

$$\log D = \log (A_0 + A_1 + A_2), \tag{26}$$

where

$$A_0 = \frac{P_{X^-}}{1 + 10^{pK_{a2} - pH} + 10^{pK_{a1} + pK_{a2} - 2pH}}, \tag{27}$$

$$A_1 = \frac{P_{XH}}{1 + 10^{pH - pK_{a2}} + 10^{pK_{a1} - pH}}, \tag{28}$$

$$A_2 = \frac{P_{XH_2^+}}{1 + 10^{pH - pK_{a1}} + 10^{2pH - pK_{a2} - pK_{a1}}}. \tag{29}$$

The experimental methods used to estimate the lipophilicity parameters have been described in detail in reviews.[197–199]

Lipophilicity can be calculated using several schemes, for example, based on the additivity of the contributions of lipophilic constants of atoms [200] or fragments [202, 202] or using conformationally dependent approaches,[203, 204] molecular descriptors [205] and various solvatochromic parameters.[206] The octanol–water partition of compounds is mainly governed by two factors — the molecular volume and the capacity of molecules to form hydrogen bonds.[207, 208]

The contribution of the molecular volume to the lipophilicity has been described quantitatively [209, 210] using several descriptors such as molecular weight MW, molecular volume MV, total molecular surface area SA, the integral of the van der Waals interaction spectra VWI, molecular refraction MR and molecular polarisability α. To describe quantitatively the contribution of hydrogen bonds, free energy factors for hydrogen bond donors and acceptors C_d and C_a were employed, which are considered in greater detail below. Since the descriptors corresponding to the contribution of the molecular volume are correlated with one another, only one of them can be included in the regression equation. The best result was obtained by using the polarisability and the sum of the factors for the hydrogen bond acceptor capacity [see the Appendix, Eqn (A.14)]. This approach to the quantitative description of the lipophilicity makes it possible to estimate the contributions of polarisability and hydrogen bonding acting in the opposite directions and thus to modify the structure of compounds in order to optimise their properties. Table 1 presents the values for lipophilicity of some chemicals determined experimentally and calculated by Eqn (A.14) and the contributions to the lipophilicity of the two parameters mentioned above.

Analysis of numerous computer programmes meant for the calculation of log *P* (see [211–213]) demonstrated that they provide satisfactory accuracy in the calculation of the lipophilicities of simple neutral organic compounds but in the case of complex compounds, especially medicines, the accuracy is fairly low. This may be due to the fact that most of the calculations are based on the additivity of the logarithms of the contributions of atoms and fragments rather than on an accurate additive scheme for the contributions of the octanol (oct)–water partition coefficients of compounds. In addition, only a few experimentally determined lipophilicities of compounds containing several cationic and/or anionic groups have been published. Meanwhile it is these data that could provide the basis for precise calculations of the lipophilicities of complex ionised compounds.

Table 1. Experimental and calculated values for lipophilicity and polarisability and acceptor capacity contributions in the formation of hydrogen bonds.[210]

Compound	$\sum C_a$	α	$\sum C_d$	α	log*P*	
					calculation	experiment
Chloracetamide	3.80	8.00	−2.57	1.99	−0.59	−0.53
Acetaldehyde oxime	2.64	5.59	−1.79	1.39	−0.41	−0.12
Ethanolamine	4.16	6.43	−2.82	1.60	−1.23	−1.30
Nicotinamide	5.88	13.19	−3.98	3.28	−0.71	−0.37
Hydroquinoline	3.00	11.32	−2.04	2.82	0.77	0.59
Dimethyl sulfite	1.50	10.44	−1.02	2.60	1.57	1.77
Acyclovir	10.90	21.44	−7.39	5.34	−2.06	−1.52
Sephopyrazone	24.10	60.97	−16.34	15.18	−1.17	−0.84
Fluorouracil	5.78	10.32	−3.92	2.57	−1.36	−0.85
Hydroflumethiazide	9.31	25.16	−6.31	6.26	−0.06	0.04
Paracetamol	5.49	16.01	−3.72	3.99	0.26	0.64
Warfarin	7.41	32.83	−5.02	8.17	3.14	2.47

The application of Eqn (A.14) to the calculation of log P_{oct} values for various neutral organic compounds has resulted in the following expression

$$\log P_{oct}(\text{exp.}) = 1.00\,(\pm 0.02)\,\log P_{oct}(\text{calc.}), \tag{30}$$

$N = 2781$, $R = 0.971$, $S = 0.32$ (see ‡).

Comparison of the lipophilicities calculated from Eqn (A.14) with those determined experimentally for compounds that had been tested with 14 appropriate commercial programmes gave the following results: for 90 simple organic compounds, $R = 0.993$; for 48 medicines, $R = 0.935$ and for all the 138 chemicals tested, $R = 0.971$. These statistical criteria are commensurable with those attained in the best of the programmes tested. This outcome, together with the possibility of calculating the full pH-dependent profile for ionised compounds, attests to the value of the above-mentioned physico-chemical approach to the estimation of lipophilicity.

In recent years, in addition to the above 'classical' descriptors, another descriptor called molecular lipophilicity potential (MLP),[214–218] which quantitatively describes lipophilicity in three-dimensional space, has been widely used. The lipophilicity potential at a given point in space is a result of intermolecular interactions governed by the lipophilicity of all fragments of the molecule. Two components are needed to calculate the lipophilicity potential, namely, the fragment lipophilicities and a function of distance describing the variation of lipophilicity in space [214]

$$\text{MLP}_k = \sum_{i=1}^{N} f_i f_{ct}(d_{ik}), \tag{31}$$

where k is the label for a given point in space, i is the label of the fragment, N is the total number of fragments in the molecule, f_i is the lipophilicity constant for fragment i, f_{ct} is the distance function and d_{ik} is the distance between fragment i and point k in space.

A large number of examples in which the lipophilicity parameters have been used successfully as intermolecular interaction descriptors in the elucidation of structure–property correlations have been reported in a book [184] and in papers.[214–224] Examples of using log P_{oct} in structure–property (activity) regression equations are given in the Appendix [Eqns (A.14)–(A.16)].

It should be noted that the majority of procedures used to calculate both descriptors of intermolecular interactions and energies of reactions are based on the assumption of additivity of contributions of various types of intermolecular interactions [175]

$$\Delta G = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{motion}, \tag{32}$$

where $\Delta G_{solvent}$ is the solvent contribution, ΔG_{conf} is the change in the free energy caused by conformational changes in the reacting molecules, ΔG_{int} is the change in the free energy due to specific interactions and ΔG_{motion} is the change in the free energy of the molecule caused by changes in the translational, rotational and vibrational constituents.

‡ From here on, N, S and R are statistical criteria of equations: N is the number of observations, S is the standard deviation and R is the correlation coefficient; the Fischer criterion F is also used.

Specific interactions are calculated on the assumption that the contributions of various types of interaction obey the additivity principle. At present it is believed that the greatest contributions to the specific interactions are made by steric (st) and electrostatic (el) effects and hydrogen bonds (hb). Within the framework of this assumption, the free energy of specific interactions can be represented as follows:[225]

$$\Delta G_{\text{int}} = \Delta G_{\text{st}} + \Delta G_{\text{el}} + \Delta G_{\text{hb}} . \quad (33)$$

Below we consider the parametrisation of these three types of specific interactions, analyse the existing descriptors and describe computer programmes used to calculate the energies of these interactions.

1. Parametrisation of steric effects

The first steric parameter E_s was proposed by Taft.[226] When studying the reactivity of esters, he obtained the following relation:

$$E_s = \log \frac{k_R}{k_{Me}} , \quad (34)$$

where k_R and k_{Me} are the rate constants for hydrolysis of the compounds R^1COOR^2 and $MeCOOR^3$, respectively. It is assumed that the inductive and electronic effects have no influence on the system under consideration. Tables containing refined E_s values are presented in the study cited. Examples of successful use of this steric parameter can be found in the literature.[15, 184]

In the simplest case, steric effects can be characterised by van der Waals radii. Empirical steric parameters based on these radii have been developed in greater detail by Charton.[182, 227, 228] The one-parameter method, which is the simplest among those proposed by this researcher, uses steric parameter v, which depends on the van der Waals radii and expresses the steric effect of the first atom in the substituent. In the case of atomic groups having different substituents at the atoms through which the hindered internal rotation axis passes, it is proposed to describe steric effect either by sets of effective steric parameters for certain reactions or by branching equations, which are empirical in their character.

Note that the empirical approach developed by Charton involves many simplifications and requires introduction of additional empirical constants and coefficients when considering new sets of compounds.

The Lennard – Jones potential takes into account the steric constituent of the total energy of an interaction E_{1j}, which is represented as a combination of the attraction and repulsion forces[225]

$$E_{1j} = \frac{A}{d^{12}} - \frac{B}{d^6} , \quad (35)$$

where d is the distance between a pair of non-bonded atoms. When d is small, the term A/d^{12} accounts for the predominance of the repulsion forces and, as a consequence, positive E_{1j} values. When d is large, the term B/d^6 is greater. The minimum E_{1j} value corresponds to the optimum distance between a pair of non-bonded atoms. The Lennard – Jones potential can be used as a self-contained descriptor of intermolecular interactions and also as a measure of the contribution of steric interactions to the total interaction energy. In particular, it has been shown that for a series of substituted esters of acetic acid, a steric descriptor based on the Lennard – Jones potentials is well correlated with the Taft parameter E_s.

2. Description of electrostatic interactions

Electrostatic interactions are described using the Coulomb's law for point charges located on atoms and using standard molecular-mechanics approaches.[175]

When comparing the capacity of molecules for electrostatic interactions $V(R_i)$, molecular electrostatic potential (MEP) is used almost always; this parameter can be calculated at each point R of the space surrounding the molecule[230]

$$V(R_i) = -\int \frac{\delta(R)}{|R_i - R|} dr + \sum \frac{Z_I}{|R_i - R_I|} , \quad (36)$$

where $\delta(R)$ is the electron density, r is the distance, Z_I are the charges of the nuclei and R_I are the coordinates of the nuclei in the molecule. The results of the calculations of MEP can be represented graphically as maps. Comparison of the MEP maps for a number of compounds having the same type of activity (examples of these studies can be found in Refs 231 – 235) may prove useful for judging the character of specific electrostatic interactions of these compounds with a biological target exhibiting the given type of activity. However, it should be borne in mind that visual comparison is subjective.

In recent years, formal approaches based on comparison of quantum-chemical parameters related to electrostatic interactions have been vigorously developing. First of all, molecular quantum similarity measures (MQSM) should be mentioned.[236 – 239] According to this approach, quantum-chemical similarity of two molecules is identified by considering the electron density functional $Z_{AB}(\Omega)$

$$Z_{AB}(\Omega) = \int \rho_A(\mathbf{r}_1)\, \Omega(\mathbf{r}_1,\mathbf{r}_2) \rho_B(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2, \quad (37)$$

where $\rho_A(\mathbf{r}_1)$ and $\rho_B(\mathbf{r}_2)$ are the electron densities of molecules A and B, and Ω is a positively defined operator. Various operators can be chosen. It is convenient to present the operator in the form

$$\Omega(\mathbf{r}_1,\mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_2), \quad (38)$$

which transforms the integral in Eqn (37) into an integral of the product of density distributions in the given pair of molecules.

$$Z_{AB} = \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r})\, d\mathbf{r}. \quad (39)$$

If molecules A and B being compared are identical, we obtain a self-similarity measure

$$Z_{AA} = \int \rho_A^2(\mathbf{r})\, d\mathbf{r}. \quad (40)$$

The self-similarity measure is used to construct similarity indices, which can be regarded as normalised values of MQSM. The Carbo index was the first similarity index to be proposed[240]

$$C_{AB} = \frac{Z_{AB}}{(Z_{AA} Z_{BB})^{1/2}} . \quad (41)$$

This value ranges from zero (for comparison of dissimilar molecules) to unity (when the molecules are identical). Examples of using the Carbo index for elucidation of similarity of electronic structures have been reported.[241 – 244]

3. Quantitative description of the ability of compounds to form hydrogen bonds

Hydrogen bonding plays a crucial role in the formation of many molecular and ionic complexes in important chemical and biological processes including enzymic catalysis. The most glaring example demonstrating the significant role of hydrogen bonding in biological systems can be found in a monograph[245] in the description of DNA replication. This process is accompanied by a new binding of purine and pyrimidine bases *via* the formation of two (binding of adenine with thymine) or three (binding of guanidine with cytosine) hydrogen bonds. Even a rather rough calculation of the free energy difference for this complexation, carried out by Pauling (it was assumed that one hydrogen bond provides an energy benefit of 20 kJ mol^{-1}), demonstrated that this system of hydrogen bonds ensures a ratio of the numbers of 'correct' and 'false' recognitions equal to ten million to one!

Despite this important role of hydrogen binding in the formation of the properties of substances, until recently this intermolecular interaction has been taken into account mainly at the level of an indicator variable.[12, 13, 246–248] Therefore, in this review it is pertinent to discuss the quantitative approaches to the description of hydrogen bonding in detail.

According to Eqn (14), full thermodynamic description of a hydrogen bond requires a pair of parameters, namely, enthalpy and entropy or enthalpy and free energy. Four combinations of the changes in the enthalpy (ΔH) and the entropy (ΔS) are possible.[249–252] Note that in the majority of studies devoted to hydrogen bonding in complexes, the change in the entropy is either totally neglected or considered to be proportional to the enthalpy change. The roughness of this approach becomes clear from the examples given in a review.[251]

The development of the empirical scales of hydrogen bonding based on the experimental values of thermodynamic parameters started in the 1960s. Let us consider two approaches, the additive-multiplicative and multiplicative ones. In the additive-multiplicative method, the enthalpy change accompanying the formation of a hydrogen bond is described by six parameters [253–255]

$$-\Delta H_{hb} = e_A e_B + c_A c_B + t_A t_B, \quad (42)$$

where e is a parameter reflecting the electrostatic contribution, c is a parameter related to the covalent interaction, t is a parameter describing the charge transfer and indices A and B correspond to Lewis acids (hydrogen bond donors) and bases (hydrogen bond acceptors), respectively. Most of studies dealing with the structure–biological activity correlations present no exact data on the molecular structure of one partner of the interaction (*e.g.*, receptor, biological target). This fact hampers the use of Eqn (42). The method describing the thermodynamic parameters of hydrogen bonds based on the multiplicative approach is more widely distributed. The enthalpy of the hydrogen bond is described by the equation [256]

$$\Delta H_{ij} = \Delta H_{11} P_i E_j, \quad (43)$$

where ΔH_{11} is the enthalpy change upon the formation of a hydrogen bond between phenol and diethyl ether (standard hydrogen bond partners) in tetrachloromethane,§ P_i is a dimensionless factor for an ith hydrogen bond donor and P_j is a factor for a jth hydrogen bond acceptor.

The change in ΔH_{ij} can also be described using the donor (E_d) and acceptor (E_a) factors for the enthalpy of a hydrogen bond, which have opposite signs [92, 257]

$$\Delta H_{ij} = \Delta H_{11} E_d E_a. \quad (44)$$

The E_d and E_a values for various neutral and ionised compounds have been estimated.[258–270]

Similar equations have been proposed to describe the change in the free energy (ΔG) upon the formation of a hydrogen bond [271, 272]

$$\Delta G = a C_d C_a + a_0, \quad (45)$$

where a and a_0 are coefficients and C_d and C_a are the donor and acceptor factors for the free energy of the hydrogen bond; and to describe the complex formation constant for hydrogen bonding [273–276]

$$\log K = b\,\alpha\,\beta + b', \quad (46)$$

where b and b' are coefficients and α and β are the donor and the acceptor factors for the hydrogen bond formation constants. Taft *et al.*[277] have estimated a large number of β values for phosphoryl compounds based on the data of several publications.[261–263] With allowance for the relationship between the change in the free energy and the constants [see Eqn (15)], it is obvious that the difference between Eqns (45) and (46) reduces to the factor $2.302RT$. In the case where the same array of experimental data is considered and the same method is used to estimate the hydrogen bond parameters, there should be a strict correlation between C_d and α and between C_a and β. No general correlation has been found between E_d and C_d or between E_a and C_a; however, these correlations do exist for separate groups of compounds.[272]

Development and successful use of the empirical approaches described above are possible only when the set of experimental data on the thermodynamics of hydrogen bonds is being extended. A detailed description of an extensive set of data of this type has been reported.[210] The C_a and C_d values were calculated by selecting 414 hydrogen bond donors and 1298 hydrogen bond acceptors for which experimental thermodynamic parameters for interaction with several partners were available.[251] The matrix generated consisted of 414 columns and 1298 rows. This permitted rigorous calculations of the C_a and C_d values.

The factors for the free energy of hydrogen bonding for some compounds are presented below

Compound	C_d	Compound	C_a
CF_3COOH	−4.78	CCl_3CN	0.55
CCl_3COOH	−4.75	C_4H_9Cl	0.60
$CHCl_2COOH$	−4.28	$C_6H_5OCH_3$	1.07
3,5-$(NO_2)_2C_6H_3OH$	−4.09	CH_3OH	1.76
C_6F_5COOH	−3.82	$CH_3C(O)CH_3$	1.92
3-$NO_2C_6H_4OH$	−3.49	$(CH_3)_2SO_2$	2.01
$(CF_3)_2CHOH$	−3.31	C_5H_5N	2.45
$CF_3C(O)NHC(O)CF_3$	−2.98	$(C_2H_5)_3N$	2.57
C_6H_5OH	−2.50	$(CH_3)_2NC(O)N(CH_3)_2$	2.96
$CH(CN)_2Br$	−2.06	$(CH_3)_2SO$	3.13
4-$NO_2C_6H_4NH_2$	−1.82	$(C_6H_5)_3PO$	3.65
$CH_3C(O)SH$	−1.62	$[(CH_3)_2N]_3PO$	4.00
C_2H_5OH	−1.42	$(C_4H_9)_2NO$	4.25
$C_6H_5NH_2$	−1.04		
$CHCl_3$	−0.81		
2-$CH_3C_6H_5NH_2$	−0.50		
C_2H_5SH	−0.24		
$C_6H_5{\equiv}CH$	−0.12		

It should be emphasised that the limits of variation of the factors for the free energy of hydrogen bonding for functional groups overlap (Table 2). This indicates that substituents at a donor or acceptor atom play an exceptionally important role in the formation of the capacity of these atoms for hydrogen bonding.

Comparison of the experimental values for enthalpy and free energy with the values calculated using Eqns (44) and (45) proves

Table 2. Minimum and maximum free energy factors for hydrogen bonds for various functional groups.

Hydrogen bond donors	C_d^{min}	C_d^{max}	Hydrogen bond acceptors	C_a^{min}	C_a^{max}
$R^1R^2R^3C-H$	−2.06	−0.12	R−O−R	0.78	2.01
RS−H	−1.92	−0.24	R−S−R	0.57	1.51
R^1R^2N-H	−3.32	−0.50	$R^1R^2R^3N$	0.58	3.65
PhO−H	−4.24	−0.76	R≡N	0.55	2.77
AlkO−H	−3.50	−1.06	$R^1R^2R^3P=O$	2.40	4.02
RC(O)O−H	−4.78	−2.50	$R^1R^2S=O$	1.16	3.25

§ Most of the experiments on the thermodynamics of hydrogen bonding have been carried out in tetrachloromethane. This solvent is non-polar and therefore convenient for determining the values of the thermodynamic factors described below.

the adequacy of the multiplicative approach to the quantitative description of the thermodynamics of hydrogen bonding [251]

$$\Delta H(\text{calc.}) = -0.49\,(\pm 0.29) + 0.99\,(\pm 0.08)\,\Delta H(\text{exp.}), \quad (47)$$

$N = 2787$, $R = 0.970$, $S = 2.40$, $F = 44350$,

$$\Delta G(\text{calc.}) = -0.07\,(\pm 0.12) + 1.04\,(\pm 0.05)\,\Delta G(\text{exp.}), \quad (48)$$

$N = 3301$, $R = 0.991$, $S = 1.12$, $F = 175000$.

In the case where charge transfer makes a significant contribution to the hydrogen bond in a particular complex, the values calculated in this way may deviate from experimental results.[278]

Using the known C_d values for 414 hydrogen bond donors and C_a values for 1298 hydrogen bond acceptors, one can calculate ΔG_{hb} for 537 372 hydrogen-bonded complexes. When a similar approach was used to calculate the constants of formation of hydrogen bonds, the α values for 150 hydrogen bond donors and the β values for 500 hydrogen bond acceptors were determined; this permitted the calculation of K_{hb} for 75 000 complexes.[279]

A whole series of descriptors have been proposed for the multiplicative description of the thermodynamic parameters of a hydrogen bond, including the donor and acceptor factors for the hydrogen bond free energy and enthalpy for the strongest donor and acceptor sites in a molecule (C_a^{max}, C_d^{max}, E_a^{max}, E_d^{max}), the sum of the free energy and enthalpy factors over all the donor and acceptors sites in a molecule ($\sum C_a$, $\sum C_d$, $\sum E_a$, $\sum E_d$) and the four last-mentioned descriptors divided by the molecular weight ($\sum C_a$/MW, $\sum C_d$/MW, $\sum E_a$/MW, $\sum E_d$/MW).[280,281] These descriptors have been used to model the structure–property relationships in complex processes such as complex formation,[91,282–286] solubility in water,[210], lipophilicity,[210] inhibition of cholinesterase,[287,288] inhibition of dopamine back capture,[289] determination of the boiling points of hydrocarbons and acids,[290] insecticide activities of nitrogen and phosphorus organic compounds,[291] ionophoric activity of macrocycles,[292] fungicidal activity of phenols and β-adrenergic activity of alicyclic acids and phenols,[293] toxicity against *Golden orfe fish*[294] and permeability of biological membranes.[210,295] Descriptors based on α and β factors have been described in a review.[279] Similar indices for hydrogen bonding have been reported in one more study.[296]

VIII. Indicator descriptors

Indicator descriptors (sometimes called *de novo* constants) have found fairly wide use in modelling the structure–property (activity) relationships. It is difficult to classify these descriptors into any particular class because they can encode in an implicit form various structural features of compounds (properties of hydrogen bond donors and acceptors, the presence or absence of hydrogen bonds within the molecules, *ortho*-effects, *cis–trans* isomerism, stereoisomerism, various fragments, *etc.*).

The expediency of using indicator variables in some cases has been demonstrated in a review.[297] Thus for two sets of papain ligands, similar two-parameter relations were obtained [see the Appendix, Eqns (A.7) and (A.8)], having fairly close coefficients at MR and σ but substantially different constant terms. By combining these equations and by using the indicator variable I ($I = 1$ for mesylamides and $I = 0$ for benzamides), the researchers derived a common relation applicable to both series [see the Appendix, Eqn (A.24)].

Other examples of using indicator descriptors can also be found in the literature.[298,299]

Indicator descriptors are especially useful at initial stages of investigations. These descriptors can be used for tentative combinations of compounds to give various subsets of compounds with a particular property, which is needed to form a training set meant for the modelling of the structure–property relationship based on a more extensive structure description.

In analyses of the structure–biological activity relationships, various physico-chemical molecular properties or experimental parameters obtained by various physical methods are often used as structure descriptors.[15] The use of physicochemical properties as descriptors has been discussed in the preceding Sections. A more detailed discussion of these descriptors can be found in the literature.[300]

IX. Modern computer software for the calculation of descriptors

At present the market of computer software meant for chemistry and, in particular, for molecular design is filled by diversified programme products. Therefore, to be competitive, every new programme must have obvious novelty elements or obvious advantages over other programmes.

1. MOLSURF is a generator of chemical descriptors for QSAR.[193] This programme has three distinctive features.

Firstly, it is assumed that a researcher prefers describing molecules using chemical terms such as hydrogen bonds, nucleophiles, electrophiles, bases, acids, hydrophobicity, polarisability, π-interaction, *etc.* rather than using quantum-chemical terms (HOMO, LUMO, electrostatic potential, force fields, *etc.*). Therefore, the MOLSURF programme calculates the descriptors related the above chemical notions.

Secondly, the programme is based on correlations of the calculated values of descriptors with experimental data.

Thirdly, the programme prefers descriptors of substituents to molecular descriptors, because the former reflect more accurately the variation of properties upon structure modification and because their values can be calculated more precisely.

2. The calculations and the use of theoretical descriptors within the framework of the concept of linear correlations of free energy have been described in a number of publications.[301–306] In this fairly simple approach, a property is described by regression equations containing descriptors related to steric interactions, polarisability and hydrogen bonding as independent variables. The steric interactions are estimated quantitatively using the van der Waals volumes V_{mc}. A descriptor called polarisability index (π_I) is defined as polarisability divided by volume. To describe hydrogen bonds, the researchers distinguish two constituents, the acidic and basic ones. Each constituent is divided in turn into covalent and electrostatic components. The covalent contribution is assumed to be determined by the energy of the highest occupied molecular orbital (HOMO). The electrostatic contribution to the basicity is characterised by the formally most negative atomic charge (q_-) in the molecule (it is assumed that the most negative atom should interact most efficiently with the proton in a neighbouring molecule). The covalent component in the acidic constituent (ε_a) is characterised by the energy of the lowest unoccupied molecular orbital (the difference between the LUMO energy of the compound and the HOMO energy of water divided by hundred). A smaller ε_a value points to a higher tendency to form hydrogen bonds (the ε_b component) with water. The contribution of the electrostatic component to the acidity is determined by the most positive charge of a hydrogen atom in the molecule (q_+).

According to this approach a property (P) can be described by the following six-parameter equation

$$\log P = \log P_0 + \frac{AV_{mc}}{100} + B\pi_I + C\varepsilon_b + Dq_- + E\varepsilon_a + Fq_+. \quad (49)$$

where A, B, C, D, E and F are coefficients.

This method has been used to describe the properties and biological activity of chemical compounds and gave formally rather good results. Nevertheless, this approach seems to be based on a simplified physical model. In addition, when initialising sets are relatively small, multiparameter regression equations of this type are not stable and may lead to faulty results when used to predict the properties of compounds.

3. The CoMFA programme (Comparative Molecular Field Analysis) is currently the most widespread computer programme

used to calculate the energies of intermolecular interactions in terms of the corresponding descriptors establishing the structure–property relationships.[307]

The parameters calculated initially include the energies of steric and electrostatic interactions between the compound considered and a probe atom being placed into various points of a regular three-dimensional grid. The size of this grid should be sufficiently large to accommodate the largest molecule of the initialising set. The size of the cell in this grid is usually taken to be 2.0 Å and, in some cases, 1 Å (the smaller the grid dimensions, the larger the number of descriptors and the time it takes to calculate them). An sp^3 hybridised carbon atom with the charge +1.0 was used initially as the probe atom; later, various atoms and groups have been used (see, *e.g.* Ref. 308). It is clear that these calculations give an enormous number of descriptors, exceeding the number of compounds in the initialising set by a large factor. Most of the descriptors thus obtained are correlated with one another. Under these conditions, a stable model can be designed only by using a special procedure. The PLS (Partial Least Squares)[309] method permits selection of orthogonal components. During the design of stable models, the number of these components can be optimised by calculating the correlation coefficients for predictive models, formed by the cross-validation procedure.

During the last decade, the CoMFA programme has been constantly improved. For example, several studies have been devoted to the methodological aspects of the programme development[310,311] (description of non-linear CoMFA, the use of various semiempirical methods for optimisation of the calculations of the geometry and the properties of compounds). The programme has been used to study many types of biological activity of various compounds, for example, the relationship between the structure of 2-hetarylquinoline-4-amines and their biological activity against HIV,[312] the structures and steric and electronic effects of phosphorus ligands[313] and the structure–inhibitory activity correlations for various compounds,[314] steroid inhibitors,[315] melatonin receptor ligands[316] and calcium blocking agents.[317]

However, the CoMFA programme is not free from drawbacks. First of all, it should be noted that until recently it had been unable to calculate the energies of hydrogen bonds, which play a crucial role in many biological processes. In addition, based on the parameters that act as independent variables in the final regression equation, it is hardly possible to construct any useful physical model for a structure–property relationship because these variables are linear combinations of many thousands of descriptors. Yet another drawback of this programme is that it cannot be run on a personal computer.

4. The GRID programme,[225] having some common features with CoMFA, also uses atoms and functional groups of various natures as the probe atoms to estimate the intermolecular interactions with the macromolecules under study, which are put in a three-dimensional grid. It is significant that this programme, unlike CoMFA, takes into account not only steric and electrostatic interactions but also hydrogen bonds (E_{hb}). Initially this has been done using a direction-dependent 6–4 function with a broad energy minimum[225,318]

$$E_{hb} = \left(\frac{C}{d^6} - \frac{D}{d^4}\right)\cos^m Q. \tag{50}$$

where Q is the DHP angle (D is the donor atom of the protein, H is a hydrogen atom and P is the probe electron-withdrawing atom); m is usually taken to be 4.

If the interaction involves two identical atoms, the interatomic separation d_{min} is determined by tabulated values for C and D. If different atoms interact, the geometrical mean of their individual D values and the arithmetic mean of their separations (d_{min}) are used. Later, the 6–4 function has been replaced by a 8–6 function.[319–320]

$$E_r = \frac{C}{r^8} - \frac{D}{r^6}, \tag{51}$$

where r is the distance between the probe atom and some atom in the compound considered. The hydrogen bond energy is represented by the following expression:

$$E_{hb} = E_r E_\tau E_p, \tag{52}$$

where E_τ and E_p are angle functions taking account of the mutual arrangement in space of the hydrogen atom and the lone electron pair of the hydrogen bond acceptor.

The GRID programme has been used successfully to design models of various biological receptors.[321–327]

5. As an alternative to the CoMFA and GRID programmes, approaches reflecting a three-dimensional structure and the capacity for intermolecular interactions as a spectrum of interatomic distances or interatomic interactions can be considered.[251,270,328–340] In these cases, a molecular structure is represented as a radial scattering curve, which can be constructed for a molecule of any complexity provided that its Cartesian coordinates are known. A similar approach is employed in gas phase electron diffraction studies of molecular structures. The radial distribution function [$M(s)$] in the spherical approximation for atoms i and j and in the absence of intramolecular vibrations is a spectrum of interatomic distances, the intensity of the spectrum being determined by the natures of the atoms spaced at a given distance and by the dynamics of the variation of this distance during intramolecular vibrations

$$\mathrm{M}(s) = \sum_{i=1}^{N}\sum_{j>i}^{N} C_{ij}\frac{\sin sr_{ij}}{sr_{ij}}, \tag{53}$$

where C_{ij} is a coefficient characterising the scattering capacity of this pair of atoms, s is an angle parameter: $s = (4\pi/\lambda)\sin(Q/2)$ (λ is the electron wavelength and Q is the scattering angle) and r_{ij} is the interatomic distance in the molecule.

It can be said that the radial distribution function obtained in an electron diffraction experiment is the pattern 'seen' by an electron interacting with the molecule in a scattering event. This analogy has been extended to the case where a receptor interacts with molecules of a substance administered into an organism.[328] It is known that a receptor interacts with parts of the molecule that are complementary to the active sites of the receptor rather than with the whole molecule. The governing factors are the distances between the active sites and the susceptibility to intermolecular interactions of the atoms located at these distances from one another in the reagent molecules. A sort of scattering of the reagent molecules on the active sites of the receptor occurs, *i.e.* the receptor 'sees' the molecule as a radial distribution function. In this approach, the radial distribution function can be modified in such a way as to take into account quantitatively the ability of atoms of the reagent to undergo various intermolecular interactions. This approach was embodied in some studies[251,340] and in the MOLTRA (Molecular Transform Analysis) programme package, which can be used to calculate the spectra for the steric interactions of atoms; interactions of positively charged atoms with one another, negatively charged atoms with one another and positively charged atoms with negatively charged atoms; interactions of hydrogen bond donors with one another, hydrogen bond acceptors with one another and hydrogen bond donors with hydrogen bond acceptors. Each of these spectra is a superposition of bands corresponding to pairs of atoms in the molecule. The positions of the band maxima are matched by the equilibrium distances between these atoms and the areas below the band contours are proportional to the products of the given local properties for a given pair of atoms.

Figure 4 shows the spectra for the interactions of hydrogen bond acceptors in some macrocyclic polyethers. Each point in the spectra presented can be considered as a descriptor of intermolecular interactions describing the interaction of hydrogen bond

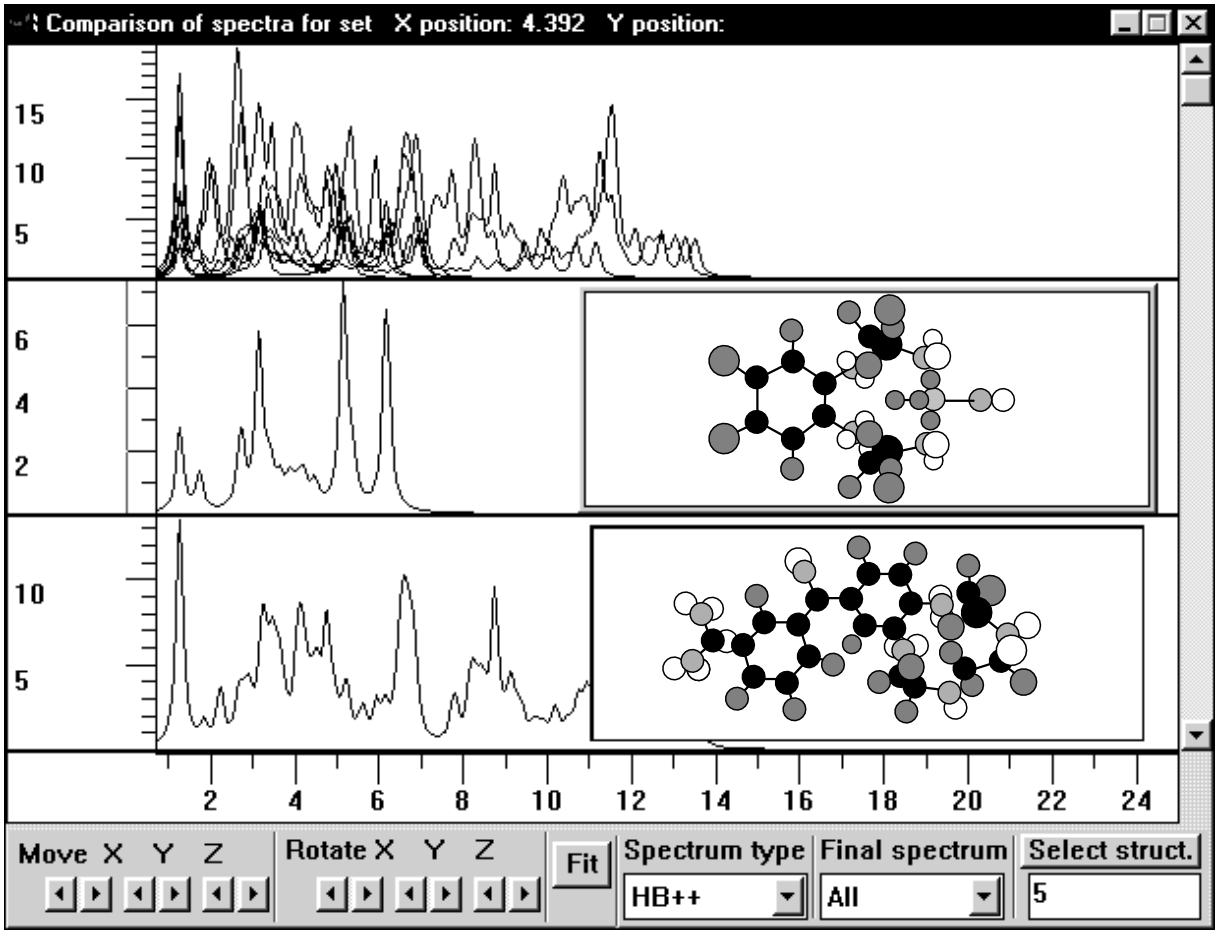


Figure 4. General view of the spectrum of interactions of hydrogen bond acceptors in some macrocyclic polyethers.

acceptors located at a given distance. For distances ranging from 1.1 to 20.0 Å and with a step of 0.1 Å, each spectral pattern provides 190 descriptors. Seven spectral patterns permit the description of the possible steric and Coulomb interactions and hydrogen bonds in terms of 1340 descriptors.

X. The necessary properties of descriptors

A great number of descriptors have been reported by now. It is clear that the extension of the views on molecular structure stimulates the design of new structure models and new descriptors reflecting these views. The attempts of developing new descriptors which, having been introduced in computer programmes, would preserve and utilise the information on the molecular structure better than the earlier descriptors are also justified.

Nevertheless, the introduction of new descriptors is expedient only in two cases — either when none of the existing descriptors or their combination provides the elucidation of stable structure–property correlations for the training set used or when the new descriptor ensures a substantial improvement of the statistical criteria of the model.[9]

New descriptors should be readily calculable, convenient to use and versatile; they should ensure comprehensive description of a given structural feature and be interpretable.[341]

Unfortunately, the above requirements are by far not always taken into account.

The efficiency of descriptors of various classes used in establishing the structure–biological activity correlations becomes clear upon comparison of the statistical criteria of the models designed using these descriptors. For example, Raevsky *et al.*[290, 293] compared the models for classification of agonists and antagonists among substituted phenylethylamines based on a number of topological and physico-chemical descriptors using linear discriminant analysis, a discrete regression model[290] and significant cluster analysis.[342] In all of these types of classification, the best results were attained by using hydrogen bond descriptors. Other examples can also be found in the literature.[14–16, 337, 343, 344]

In this review, the descriptors were introduced in the order of increasing amount of information on the molecular structure contained in them. However, this does not mean that the use of a more informative descriptor guarantees better results in elucidation of the structure–property correlations. Indeed, there are many examples of successful employment of topological descriptors for the calculation of the melting and boiling points, evaporation enthalpies and other 'simple' physical properties within narrow series of related compounds, which differ, for example, in the length of the chain in the alkyl substituent. In these cases, the statistical criteria of regression equations containing more informative descriptors as independent variables prove sometimes to be inferior to those for equations based on simpler descriptors, because the magnitudes of more informative descriptors are estimated with lesser accuracy and because they carry 'excessive' information for the set considered. The foregoing leads to the following conclusions: it is expedient to use combinations of known descriptors with different information contents; it is useful to develop new descriptors and perfect the procedures to attain more precise calculations of all descriptors; well-founded selection of descriptors for a particular set of compounds is necessary.¶

Note that comprehensive description of a molecular structure cannot be attained in terms of any single descriptor. One descriptor can predominantly take into account one feature of

the electronic or steric structure or one type of intermolecular interactions. For this reason, description of a molecular structure using several descriptors and construction of structure–property models based on several independent variables appears expedient.

In the construction of structure–biological activity models, the modern views on the processes occurring in an organism with participation of chemical compounds play an important role in the selection of appropriate descriptors. For example, it is currently believed that the complex mechanism of action of a substance on an organism consists of three stages — pharmaceutic, pharmacokinetic and pharmacodynamic stages.[345] The pharmaceutic stage is associated with the administration of a substance into an organism and includes disintegration of the preparation. The pharmacokinetic stage is determined by the transport of the substance from the administration site to the particular biological target and includes the processes of absorption, distribution, metabolism and excretion. At this stage, the reactivity of the initial compounds and their capacity for complex formation are important. The pharmacodynamic stage is characterised by the interaction of the initial substance or the products with biological targets responsible for the particular type of biological activity. At this stage, the three-dimensional structure of the interacting partners is the most important factor.

The exceptional popularity of the Hansch–Fujita three-parameter equation is due most of all to the fact that it uses descriptors of the electronic structure (Hammett constants), steric interactions (Taft steric constants), and transport properties (octanol–water partition coefficients). At present, similar models and the corresponding equations are also used successfully in the molecular design of biologically active substances.[346–349] For example, for a series of bicyclic compounds, an equation describing their ability to bind the muscarinic receptor based on three descriptors has been derived [210] [see the Appendix, Eqn (A.23)]. One descriptor characterises the transporting properties of the compounds, the second descriptor is related to the solubility of the compounds in water and the ability to form hydrogen bonds with the donor sites of the receptor and the third one points to the important role of the distance between two particular hydrogen bond acceptors in the formation of the substrate–receptor complex.

The development of models requires the observance of formal criteria. From the statistical viewpoint, the number of compounds considered should exceed the number of descriptors by a large factor. The information contents of the descriptors must be sufficiently large and, simultaneously, they should not correlate with one another. This can be attained by several procedures including experimental design, [350] the SIMCA/PLS method [351, 352] and the principal component and factor analyses.[353]

XI. Conclusion

At present, molecular structure descriptors constitute an integral part of the majority of studies dealing with structure–property (including biological activity) correlations; they are widely used to optimise characteristics of compounds in the molecular design. The Appendix to this review presents examples of successful application of various types of descriptors for quantitative description of diverse properties and types of biological action.

It should be emphasised that the molecular structure descriptors used should finally provide the design of structure–property models, which satisfy certain requirements. These models should be consistent with the laws of Nature, be relatively simple and have a clear physical meaning; in addition, they must be stable and hence possess predictive capacity and provide the possibility of identifying the relationships that were not specified in an explicit form when the model was developed. All these points broach one more important section of the computer-aided molecular design, namely, modelling of the structure–property relationships, and require separate discussion.

¶ An expert can rely not only on the generally recognised views on the mechanism of formation of the property being considered but also on the knowledge not yet formalised and even on intuition.

XII. Appendix

Examples of successful application of molecular structure descriptors in structure–property (biological activity) correlatons are presented below.

1. Element-level descriptors

Solubility of gases in water, $\log L^{w}$ (see [354]).

$$\log L^{w} = 42.37(\pm1.11)\mathrm{HDCA}(2) + 0.65(\pm0.02)[2n(\mathrm{N}) + m(\mathrm{O})] - 0.16(\pm0.02)\Delta E + 0.12(\pm0.01)\mathrm{PCWT}^{E} + 0.82(\pm0.01)N_{ring} + 2.65(\pm0.22), \quad (A.1)$$

where HDCA(2) is a hydrogen bond descriptor, $[2n(\mathrm{N}) + m(\mathrm{O})]$ is a descriptor characterising the number of nitrogen and oxygen atoms in the molecules of compounds studied, $\Delta E = E_{LUMO} - E_{HOMO}$, PCWT^{E} is an electronic topological index, N_{ring} is the number of aromatic rings; $N = 406$, $R = 0.971$, $S = 0.52$, $F = 1300$.

2. Topological descriptors

Lipophilicity of organic compounds (based on topological and information topological indices), $\log P$ (see [355])

$$\log P = 0.19P_{10} - 1.46\,\mathrm{IC}_0 + 1.09\,\mathrm{CIC}_2 - 0.77\,\mathrm{CIC}_3 - 1.36\,{}^{6}X_b + 5.34\,{}^{0}X^{V} - 3.41\,{}^{1}X^{V} + 0.55\,{}^{4}X^{V} - 0.41\,{}^{3}X_C^{V} + 1.10X_w - 0.17W - 5.60, \quad (A.2)$$

$N = 219$, $R = 0.955$, $S = 0.35$, $F = 194$.

β-*Blocking activity of medicines*, pLD_{50} (see [356]).

$$\mathrm{pL}D_{50} = 1.79\,{}^{1}\chi - 1.81\,{}^{1}\chi^{V} + 0.36\,{}^{2}\chi + 2.37\,{}^{4}\chi_p^{V} - 2.76\,{}^{4}\chi_p + 1.56, \quad (A.3)$$

where χ are Kier and Hall connectivity indices; $N = 17$, $R = 0.905$, $S = 0.208$, $F = 10.1$.

3. Descriptors of the electronic structure of molecules

Binding of an estrogen receptor, pEC_{50} (see [357]).

$$\mathrm{pEC}_{50} = 3.38(\pm0.12)\,D^2 + 3.22(\pm0.20)\,\mathrm{Gap} - 44.3(\pm1.8) \quad (A.4)$$

where D is the Balaban topological index, Gap is the $E_{LUMO} - E_{HOMO}$ difference; $N = 33$, $R = 0.99$, $S = 0.30$, $F = 595.0$.

Toxicity against Daphnia magna, EC_{50} (see [358]).

$$\mathrm{EC}_{50} = -3.63\frac{V_{mc}}{100} - 45.8\pi + 3.71q_{-} - 2.0q_{+} + 11.4, \quad (A.5)$$

where V_{mc} is the molecular volume, π is the polarisability index, q_{-} is the electrostatic basicity, q_{+} is the electrostatic acidity; $N = 38$, $R = 0.977$, $S = 0.37$, $F = 176$.

Fungicide activity of phosphonates, pEC_{50} (see [359]).

$$\mathrm{pEC}_{50} = 0.296(\pm0.124)\sum\pi - 0.047(\pm0.012)\mathrm{MR}(\mathrm{R}^2) + 0.543(\pm0.331)\sum F' + 3.556(\pm0.215), \quad (A.6)$$

where π is the hydrophobicity constant, $\mathrm{MR}(\mathrm{R}^2)$ is the molar refraction of the second substituent, F' is the Swain–Lupton electronic parameter; $N = 57$, $R = 0.851$, $S = 0.290$, $F = 46.32$.

Binding of papain ligands by phenyl N-mesylglycinates (see [297]).

$$\log\frac{1}{K_m} = 0.529(\pm0.230)\mathrm{MR} + 0.379(\pm0.200)\,\sigma + \quad (A.7)$$

$$+ 1.877(\pm 0.130),$$

where σ is an electronic constant; $N = 13$, $R = 0.935$, $S = 0.105$, $F = 34.51$.

Binding of papain ligands by phenyl N-benzoylglycinates (see [297]).

$$\log\frac{1}{K_m} = 0.771(\pm 0.670)\mathrm{MR} + 0.728(\pm 0.370)\,\sigma + \quad \text{(A.8)}$$
$$+ 3.623(\pm 0.340),$$

$N = 7$, $R = 0.971$, $S = 0.148$, $F = 32.85$.

4. Molecular shape descriptors

Inhibition of ligand binding, $\mathrm{pIC_{50}}$ (see [360]).

$$\mathrm{pIC_{50}} = 10.25(\pm 0.97)\,V + 1.43(\pm 0.56), \quad \text{(A.9)}$$

where V is the shape index; $N = 26$, $R = 0.912$, $S = 0.462$, $F = 111.1$.

Sweetness of aniline derivatives, logRS (see [361]).

$$\mathrm{logRS} = 0.52\,\mathrm{L} - 1.37\,\mathrm{W1} + 3.71, \quad \text{(A.10)}$$

where L and W1 are STERIMOL parameters; $N = 20$, $R = 0.90$, $S = 0.32$.

Psychomimetic activity of phenyl-2-aminopropane derivatives, logMU (see [362]).

$$\mathrm{logMU} = 1.99(\pm 0.13) - 1.78(\pm 0.17)\,\mathrm{MTD} + \quad \text{(A.11)}$$
$$+ 0.93(\pm 0.10)\,S_3 - 0.77(\pm 0.12)\,q_4,$$

where MTD is the minimum topological difference, S_3 is the electron topological index for the atom with number 3 in the benzene ring, q_4 is the charge on atom 4 of the benzene ring; $N = 49$, $R^2 = 0.885$, $S = 0.241$, $F = 84.7$.

Inhibition of hepatic receptors, $\mathrm{pEC_{50}}$ (see [169]).

$$\mathrm{pEC_{50}} = 0.447\,\mathrm{Tm} + 0.242\,\mathrm{Ts} - 0.158\,\mathrm{Vu} + 2.361, \quad \text{(A.12)}$$

where Tm, Ts and Vu are WHIM descriptors; $N = 71$, $R = 0.914$, $S = 0.64$.

Inhibition of photosystem II, $\mathrm{pIC_{50}}$ (see [363]).

$$\mathrm{pIC_{50}} = 1.081\log P - 0.57\sigma(\mathrm{Y,Al}) + 0.29L(\mathrm{Y,Al}) - \quad \text{(A.13)}$$
$$- 0.95T(\mathrm{Y,Al}) - 2.21T(\mathrm{Y,Ar}) - 0.32W(\mathrm{X}) +$$
$$+ 1.02\,I_u + 8.16,$$

where Y and X are substituents, Al is alkyl, Ar is aryl, L, W and T are shape descriptors, I_u is an indicator variable (1 for urea derivatives and 0 for all other compounds); $N = 69$, $R = 0.921$.

5. Intermolecular interaction descriptors

Lipophilicity of various compounds, log$P$ (see [209]).

$$\log P = -0.01(\pm 0.08) + 0.249(\pm 0.005)\,\alpha - \quad \text{(A.14)}$$
$$- 0.68(\pm 0.02)\textstyle\sum C_a,$$

where α is the molecular polarisability, $\sum C_a$ is the sum of the factors for the hydrogen bond acceptor capacity; $N = 234$, $R = 0.961$, $S = 0.34$, $F = 1383$.

Effect of protein binding, log HSAI (see [364]).

$$\log \mathrm{HSAI} = 0.59(\pm 0.03)\log P + 0.27(\pm 0.08), \quad \text{(A.15)}$$

$N = 57$, $R = 0.92$, $S = 0.28$, $F = 283$.

Toxicity against Daphnia, log(1/$\mathrm{IC_{50}}$) (see [365]).

$$\log\frac{1}{\mathrm{IC_{50}}} = 0.809\log P + 1.64, \quad \text{(A.16)}$$

$N = 23$, $R^2 = 0.84$, $S = 0.30$, $F = 118$.

Bactericidal activity, pMIC (see[366]).

$$\mathrm{pMIC} = -0.35\,\mathrm{p}K_a + 0.68, \quad \text{(A.17)}$$

$N = 8$, $R = 0.973$, $S = 0.063$.

Ligand–receptor binding, ΔG (see [367]).

$$\Delta G = 0.06(\pm 0.01)\,E_L - 0.05(\pm 0.02)\,\Delta E_{solv} + 7.74(\pm 0.57), \quad \text{(A.18)}$$

where E_L is the change of the intramolecular energy of the ligand upon binding, ΔE_{solv} is the change of the solvation upon binding; $N = 10$, $R = 0.92$, $F = 20.61$.

Inhibitory activity of aryl-substituted piperazine and piperylene, log(1/$\mathrm{EC_{50}}$) (see [368]).

$$\log\frac{1}{\mathrm{EC_{50}}} = 0.62(\pm 0.04)\log P - 0.39(\pm 0.05)\textstyle\sum C_a - \quad \text{(A.19)}$$
$$- 2.16(\pm 0.18),$$

$N = 17$, $R^2 = 0.90$.

Toxicity of carboxylic acids against Xenopus embryos, $\mathrm{pLC_{50}}$ (see [358]).

$$\mathrm{pLC_{50}} = 0.696\log P - 0.358\,\mathrm{p}K_a - 0.541, \quad \text{(A.20)}$$

$N = 44$, $R = 0.947$.

Membrane permeability (*human red cell basal*), logBP (see [369]).

$$\mathrm{logBP} = -0.70(\pm 0.64) + 1.08(\pm 0.16)\textstyle\sum C_d, \quad \text{(A.21)}$$

$N = 10$, $R = 0.983$, $S = 0.433$.

Membrane permeability (cells of the alga *Chara ceratophylla*), log Per (see [369]).

$$\log \mathrm{Per} = 0.83(\pm 0.57) + 0.59(\pm 0.12)\textstyle\sum C_d, \quad \text{(A.22)}$$

$N = 27$, $R = 0.903$, $S = 0.49$.

Binding of the muscarinic receptor by bicyclic ligands, log K (see [210]).

$$\log K = 1.28(\pm 0.15)\log P - 0.09(\pm 0.06)\textstyle\sum C_a + \quad \text{(A.23)}$$
$$+ 0.27(\pm 0.08)\,\mathrm{HBA5.3A} - 4.58(\pm 0.48),$$

where HBA5.3A is a descriptor characterising the interaction between two hydrogen bond acceptors spaced 5.3 Å apart; $N = 27$, $R = 0.918$, $S = 0.38$, $F = 41.1$

6. Indicator descriptors

Binding of papain ligands by phenyl N-mesylglycinate and N-benzoylglycinate, log(1/K_m) (see [297]). (The combination of Eqns (A.7) and (A.8) with the aid of an indicator variable I ($I = 1$ for mesylamides and $I = 0$ for benzamides)].

$$\log\frac{1}{K_m} = 0.57(\pm 0.26)\mathrm{MR} + 0.56(\pm 0.19)\sigma - \quad \text{(A.24)}$$
$$- 1.92(\pm 0.15)I + 3.74(\pm 0.17),$$

$N = 20$, $R = 0.990$, $S = 0.148$, $F = 272.04$.

Inhibition of HIV-RT, $\mathrm{pIC_{50}}$ (see [370]).

$$\mathrm{pIC_{50}} = 1.47\,I + 0.41\,\mu + 2.69\log \mathrm{MW} - 4.36, \quad \text{(A.25)}$$

where I is the indicator variable, MW is the molecular weight, $N = 14$, $R = 0.94$, $S = 0.42$, $F = 27.4$.

Cytotoxic activity, $\mathrm{pIC_{50}}$ (see [371]).

$$\mathrm{pIC_{50}} = 0.73(\pm 0.46)\,I_{2'} + 1.14(\pm 0.38)\,I_{5'} + \quad \text{(A.26)}$$
$$+ 1.12(\pm 0.43)\,I_{xy} + 3.15,$$

where $I_{2'}$, $I_{5'}$, and I_{xy} are indicator variables; $N = 22$, $R = 0.915$, $S = 0.32$, $F = 30.78$.

References

1. C Hansch *Drug Metabolism Rev.,* **15** 1279 (1984-85)
2. C Hansch, T Fujita *J. Am. Chem. Soc.* **86** 1616 (1964)
3. R Lipnick, V A Filov *Trends Pharm. Sci.* **13** 56 (1992)
4. N V Lazarev *Neelektrolity* (Nonelectrolytes) (Leningrad: VMMA, 1944)
5. O A Raevsky, A M Sapegin *Usp. Khim.* **57** 1565 (1988) [*Russ. Chem. Rev.* **57** 897 (1988)]
6. S Nikolic, N Trinajstic *Croat. Chem. Acta* **70** 777 (1997)
7. O A Raevsky *Ros. Khim. Zh.* **39** 109 (1995) [a]
8. J Tomasi *J. Mol. Struct. (Theochem)* **179** 273 (1988)
9. M Charton *Newsl. Int. QSAR Soc.* **1** 3 (1991)
10. C Hansch *Acc. Chem. Res.* **26** 147 (1993)
11. C Hansch, D Hoekman, H Gao *Chem. Rev.* **96** 1045 (1996)
12. E J Lien *SAR. Side Effects and Drug Design* (New York: Marcel Dekker, 1987)
13. E J Lien *Chin. Pharm. J.* **42** 93 (1990)
14. P C Jurs, S L Dixon, L Egolf, in *Chemometric Methods in Molecular Design* Vol. 2 (Ed. H Waterbeemd) (Dusseldorf: VCH, 1995) p. 15
15. B Testa, L B Kier *Med. Res. Rev.* **11** 35(1991)
16. A J Stuper, W E Brugger, P Jurs *Computer Assisted Studies of Chemical Structure and Biological Function* (New York: Wiley, 1979)
17. N S Zefirov, S I Kuchanov (Ed.) *Primenenie Teorii Grafov v Khimii* (Application Graph Theory in Chemistry) (Novosibirsk: Nauka, 1988)
18. M I Stankevich, I V Stankevich, N S Zefirov *Usp. Khim.* **57** 337 (1988) [*Russ. Chem. Rev.* **57** 191 (1988)]
19. Yu G Papulov, V R Rozenfel'd, T G Kemenova *Molekulyarnye Grafy* (Molecular Graphs) (Tver': Tver' State University, 1990)
20. R B King (Ed.) *Chemical Applications of Topology and Graph Theory* (Amsterdam: Elsevier, 1983)
21. S S Tratch, M I Stankevitch, M S Zefirov *J. Comput. Chem.* **11** 899 (1990)
22. L B Kier, L H Hall *Molecular Connectivity in Structure – Activity Analysis* (New York: Wiley, 1986)
23. M Randic, X Guo, T Oxley, H Krishnaprijan *J. Chem. Inf. Comput. Sci.* **33** 709 (1993)
24. M Randic *J. Am. Chem. Soc.* **97** 6609 (1975)
25. L B Kier, L H Hall *J. Pharm. Sci.* **64** 1971 (1975)
26. L B Kier, L H Hall *J. Pharm. Sci.* **65** 1806 (1976)
27. L B Kier, L H Hall *Pharm. Res.* **8** 801 (1990)
28. L H Hall, B Mohney, L B Kier *Quantum Struct. – Act. Relat.* **10** 43 (1991)
29. L H Hall, L B Kier *J. Chem. Inf. Comput. Sci.* **35** 1039 (1995)
30. L H Hall, L B Kier, B B Brown *J. Chem. Inf. Comput. Sci.* **35** 1074 (1995)
31. A K Saxena *Quantum Struct. – Act. Relat.* **14** 31 (1995)
32. A K Saxena *Quantum Struct. – Act. Relat.* **14** 142 (1995)
33. D L Dowdy, T E McKone *Environ. Toxicol. Chem.* **16** 2448 (1997)
34. L H Hall, C T Story *J. Chem. Inf. Comput. Sci.* **36** 1004 (1996)
35. M T D Cronin *Toxicol. Vitro* **10** 103 (1996)
36. P-Z Lang, X-F Ma, G-H Lu, Y Wang, Y Bian *Chemosphere* **32** 1547 (1996)
37. M Soskis, D Plavsic, N Trinajstic *J. Chem. Inf. Comput. Sci.* **36** 829 (1996)
38. R Garcia-Domenech, F J Garcia-March, R M Soler, J Galvez, G M Anton-Fos, J V de Julian-Ortiz *Quantum Struct. – Act. Relat.* **15** 201 (1996)
39. R R A Abon-Shaaban, H A Al-Khamees, H S Abou-Auda, A P Simonelli *Pharm. Res.* **13** 129 (1996)
40. H Kubinyi *Quantum Struct. – Act. Relat.* **14** 149 (1995)
41. S Nikolic, N Trinajstic, Z Mihalic *Croat. Chem. Acta* **68** 105 (1995)
42. I Lukovits *J. Chem. Inf. Comput. Sci.* **34** 1079 (1994)
43. A A Dobrynin, A A Kochetova *J. Chem. Inf. Comput. Sci.* **34** 1082 (1994)
44. M Randic *Stud. Phys. Theor. Chim.* **54** 101 (1988)
45. M Randic *Int. J. Quantum Chem. Quantum Biol. Symp.* **15** 201 (1988)
46. M Randic, A F Kleiner, L M DeAlba *J. Chem. Inf. Comput. Sci.* **34** 277 (1994)
47. M Randic, M Razinger *J. Chem. Inf. Comput. Sci.* **35** 140 (1995)
48. O G Mekenyan, D Peitchev, D Bonchev, N Trinajstic, I Bangov *Drug. Design* **36** 176 (1986)
49. M V Diudea, D Horvath, A Graovac *J. Chem. Inf. Comput. Sci.* **35** 129 (1995)
50. L B Kier *Quantum Struct. – Act. Relat.* **4** 109 (1985)
51. L B Kier *Quantum Struct. – Act. Relat.* **5** 1 (1986)
52. L B Kier *Quantum Struct. – Act. Relat.* **5** 7 (1986)
53. L B Kier *Quantum Struct. – Act. Relat.* **6** 8(1987)
54. L B Kier *Quantum Struct. – Act. Relat.* **6** 117 (1987)
55. Y-Y Yao, Lu Xu, Y-Q Yang, X-S Yuan *J. Chem. Inf. Comput. Sci.* **33** 590 (1993)
56. G Rucker, C Rucker *J. Chem. Inf. Comput. Sci.* **33** 683 (1993)
57. H P Schultz, T P Schultz *J. Chem. Inf. Comput. Sci.* **33** 240 (1993)
58. H P Schultz, E B Schultz, T P Schultz *J. Chem. Inf. Comput. Sci.* **33** 863 (1993)
59. H P Schultz, E B Schultz, T P Schultz *J. Chem. Inf. Comput. Sci.* **34** 1154 (1994)
60. M V Diudea, M Topan, A Graovac *J. Chem. Inf. Comput. Sci.* **34** 1072 (1994)
61. T Pisanski, A Zithik, A Graovac, A Baumgartner *J. Chem. Inf. Comput. Sci.* **34** 1090 (1994)
62. J Galvez, R Garcia, M T Salabert, R M Soler *J. Chem. Inf. Comput. Sci.* **34** 520 (1994)
63. E Estrada *J. Chem. Inf. Comput. Sci.* **35** 701 (1995)
64. E Estrada *J. Chem. Inf. Comput. Sci.* **35** 708 (1995)
65. B Llorente, N Rivero, R Carrasco, R M Martinez *Quantum Struct. – Act. Relat.* **13** 419 (1994)
66. I Lukovits *Quantum Struct. – Act. Relat.* **9** 227 (1990)
67. G G Gash, J J Breen *J. Chem. Inf. Comput. Sci.* **33** 275 (1993)
68. A R Katritzky, E V Gordeeva *J. Chem. Inf. Comput. Sci.* **33** 835 (1993)
69. J Galvez, R Garcia-Domenech, J V de Julian-Ortiz, R M Soler *J. Chem. Inf. Comput. Sci.* **35** 272 (1995)
70. B Lucic, S Nikolic, N Trinajstic, D Juretic, A Juric *Croat. Chem. Acta* **68** 435 (1995)
71. J Niederfellner, D Lenoir, G Matuschek, F Rehfeldt, H Utschick, R Bruggemann *Quantum Struct. – Act. Relat.* **16** 38 (1997)
72. J V de Julian-Ortiz, R Garcia-Domenech, J Galvez Alvarez, R Soler Roca, F J Garcia-March, G M Anton-Fos *J. Chromatogr., A* 719 37 (1996)
73. R B King (Ed.) *Chemical Applications of Topology and Graph Theory* (Amsterdam: Elsevier, 1983)
74. G Bazylak *J. Planar. Chromatogr.* **7** 202 (1994)
75. S C Basak, S Bertelsen, G D Grunwald *J. Chem. Inf. Comput. Sci.* **34** 27 (1994)
76. W J Boecklen, G J Niemi *SAR QSAR Environ. Res.* **2** 79 (1994)
77. A T Balaban, S C Basak, T Colburn, G D Grunwald *J. Chem. Inf. Comput. Sci.* **34** 1118 (1994)
78. S C Basak, G D Grunwald *J. Chem. Inf. Comput. Sci.* **35** 366 (1995)
79. T Wieland *Arzheim.-Forsch. Drug. Res.* **46** 273 (1996)
80. A T Balaban, T S Balaban *J. Chim. Phys. Phys.-Chim. Biol.* **89** 1735 (1992)
81. W J Wisswesser *The Wisswesser Line-Formula Chemical Notation* (New York: McGraw-Hill, 1968)
82. V V Avidon *Khim.-Farm. Zh.* **8** 22 (1974) [b]
83. V V Avidon, V S Aralovich *NTI, Ser. 2, Inform. Prots. Sist.* (5) 26 (1975)
84. D Weininger *J. Chem. Inf. Comput. Sci.* **28** 31 (1988)
85. D Weininger *J. Chem. Inf. Comput. Sci.* **29** 97 (1989)
86. D Weininger *J. Chem. Inf. Comput. Sci.* **30** 237 (1990)
87. A Drefahl, M Reinhard *J. Chem. Inf. Comput. Sci.* **33** 886 (1993)
88. K K Agarwal, H L Gelernter *J. Chem. Inf. Comput. Sci.* **34** 463 (1994)
89. V V Avidon, V E Golender, A B Rozenblit, in *Metody Predstavleniya i Obrabotki Strukturnoi Informatsii dlya Analiza Svyazi Struktura – Aktivnost'* (The Methods of Presentation and Processing of Structural Information for Analysis of Structure – Activity

Relationship) (Riga: Institute of Organic Compounds, Academy of Science, Latv. SSR, 1981) p. 8
90. A B Rozenblit, V E Golender *Logiko-Kombinatornye Metody v Konstruirovanii Lekarstv* (Logical Combinatorial Methods in Drug Design) (Riga: Zinatne, 1983)
91. O A Raevsky, A M Sapegin, V V Chistyakov, V P Solov'ev, N S Zefirov *Koord. Khim.* **16** 1175 (1990) [c]
92. O A Raevsky, V V Avidon, V P Novikov *Khim.-Farm. Zh.* **16** 968 (1982) [b]
93. V P Avidon, O A Raevsky, V S Arolovich, R I Zhdanov, in *Bioactive Spin Labels* (Ed. R I Zhdanov) (Berlin: Springer, 1992) p. 549
94. G M Barenboim, A G Malenkov *Biologicheski Aktivnye Veshchestva* (Bioactive Compounds) (Moscow: Nauka, 1986) p. 90
95. I Paeva, E Golovinsky *Quantum Struct. – Act. Relat.* **9** 16 (1990)
96. A Ormerod, P Willett, D Bawden *Quantum Struct. – Act. Relat.* **8** 115 (1989)
97. A Ormerod, P Willett, D Bawden *Quantum Struct. – Act. Relat.* **9** 302 (1990)
98. W J Streich, R Franke *Quantum Struct. – Act. Relat.* **4** 13 (1985)
99. R Franke, W J Streich *Quantum Struct. – Act. Relat.* **4** 51 (1985)
100. R Franke, W J Streich *Quantum Struct. – Act. Relat.* **4** 63 (1985)
101. V K Gombar, E P Jaeger, P C Jurs *Quantum Struct. – Act. Relat.* **7** 225 (1988)
102. R A Lewis, A C Good, S D Pickett, in *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry*. (Eds H van de Waterbeemd, B Testa, G Folkers) (Weinheim: VCH, 1997) p. 137
103. J E Dubois, M Loukianoff, C Mercier *J. Chim. Phys.* **89** 1493 (1992)
104. C Mercier, G Trouiller, J-E Dubois *Quantum Struct. – Act. Relat.* **7** 88 (1990)
105. E J Martin, J M Blaney, M A Siani, D C Spellmeyer, A K Wong, W H Moos *J. Med. Chem.* **38** 1431 (1995)
106. H-J Bohm, in *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry*. (Eds H van de Waterbeemd, B Testa, G Folkers) (Weinheim: VCH, 1997) p. 125
107. M Karelson, V S Lobanov, A R Katritzsky *Chem. Rev.* **96** 1027 (1996)
108. O Kikuchi *Quantum Struct. – Act. Relat.* **6** 179 (1987)
109. H Bauknecht, A Zell, H Bayer, P Levi, M Wagener, J Sadowski, J Gasteiger *J. Chem. Inf. Comput. Sci.* **36** 1205 (1996)
110. S Karabunarliev, O G Mekenyan, W Karcher, C L Russom, S P Bradbury *Quantum Struct. – Act. Relat.* **15** 302 (1996)
111. M A Ordocica, M L Velazquez, J G Ordocica, J L Escobar, P A Lehmann *Quantum Struct. – Act. Relat.* **12** 246 (1993)
112. F T Hatch, M E Colvin, E T Seidl *Environ. Mol. Mutagen.* **27** 314 (1996)
113. A Inoue, T Kawai, M Iimura, H Sugimoto, H Kawakami *J. Med. Chem.* **39** 4460 (1996)
114. M T D Cronin, T W Schultz *Chemosphere* **32** 1453 (1996)
115. M Cocchi, M C Menziani, P G De Benrdrtti, G Cruciani *Chemometr. Intell. Lab. Syst.* **14** 209 (1992)
116. K Tuppurainen *J. Mol. Struct. (Theochem)* **306** 49 (1994)
117. B Balasubramanian *SAR QSAR Environ. Res.* **2** 59 (1994)
118. R G Parr, R Pearson *J. Am. Chem. Soc.* **105** 7512 (1983)
119. Z Medven, H Gusten, A Subljic *J. Chemometr.* **10** 135 (1996)
120. C Graham, R Gealy, O T Macina, M H Karol, H S Rosenkranz *Quantum Struct. – Act. Relat.* **15** 224 (1996)
121. W H J Vaes, C Hamwijk, E U Ramos, H J Mverhaar, J L M Hermens *Anal. Chem.* **68** 4458 (1996)
122. D F W Lewis, B G Lake *Toxicology* **125** 31 (1998)
123. R C Young, C Graham, M L Roantree, in *QSAR in Drug Design and Toxicology* (Eds D Hadzy, B Jerman-Blazic) (Amsterdam: Elsevier, 1987) p. 91
124. T I Oprea, L Rurunczi, E E Moret, in *Trends in QSAR and Molecular Modelling 92* (Ed. C G Wermuth) (Leiden: ESCOM, 1993) p. 398
125. S Gibson, R McGuire, D C Rees *J. Med. Chem.* **39** 4065 (1996)
126. D Chapman *J. Comput.-Aid. Mol. Design* **10** 501 (1996)
127. J H van Drie *J. Chem. Inf. Comput. Sci.* **37** 38 (1997)
128. A Verloop, W Hoogenstraaten, J Tipker, in *Drug Design* (Ed. J Ariens) (New York: Academic Press, 1976) p. 165
129. J Jurgens, H Schwedletzky, P Heising, J K Seydel, B Wiedemann, U Holzgrabe *Arch. Pharm. Med. Chem.* **329** 179 (1996)
130. A Tanaka, H Fujiwara *J. Med. Chem.* **39** 5017 (1996)
131. W Draber *Z. Naturforsch., C Biosci.* **51** 1 (1996)
132. M M Davila-Jimenez, M P Elizalde-Gonzales, S Garcia-Diaz *J. Liq. Chrom. Rel. Technol.* **21** 1629 (1998)
133. J E Amoore *Ann. N. Y. Acad. Sci.* **116** 457 (1964)
134. R H Rohrbaugh, P C Jurs *Anal. Chim. Acta* **199** 99 (1987)
135. G M Crippen *J. Med. Chem.* **22** 988 (1979)
136. G M Crippen *J. Med. Chem.* **23** 599 (1980)
137. G M Crippen *J. Med. Chem.* **24** 198 (1981)
138. A K Ghose, G M Crippen *J. Med. Chem.* **25** 892 (1982)
139. A K Ghose, G M Crippen *J. Med. Chem.* **27** 901 (1984)
140. A K Ghose, G M Crippen *J. Med. Chem.* **28** 333 (1985)
141. A K Ghose, G M Crippen, G R Revankar, P A McKernan, D F Smee, R K Robins *J. Med. Chem.* **32** 746 (1989)
142. G M Crippen *J. Comput. Chem.* **8** 943 (1987)
143. G M Crippen *J. Comput. Chem.* **10** 673 (1989)
144. M P Bradley, G M Crippen *J. Med. Chem.* **36** 4171 (1993)
145. L G Bouli, G M Crippen, H A Barton, H Kwon, M A Marletta *J. Med. Chem.* **33** 771 (1990)
146. A J Hopfinger *J. Am. Chem. Soc.* **102** 7196 (1980)
147. M Mabilia, R A Pearlstein, A J Hopfinger *Eur. J. Med. Chem.* **20** 1133 (1985)
148. M G Koehler, K Rowberg-Schaefer, A J Hopfinger *Arch. Biohem. Biophys.* **266** 152 (1988)
149. B J Burke, A J Hopfinger *J. Med. Chem.* **33** 274 (1990)
150. A J Hopfinger, B J Burke, W J Dunn III *J. Med. Chem.* **37** 3768 (1994)
151. B J Burke, W J Dunn III, A J Hopfinger *J. Med. Chem.* **37** 3775 (1994)
152. P I Nagy, J S Tokarski, A J Hopfinger *J. Chem. Inf. Comput. Sci.* **34** 1190 (1994)
153. J S Tokarski, A J Hopfinger *J. Med. Chem.* **37** 3639 (1994)
154. K-B Rhyu, H C Patel, A J Hopfinger *J. Chem. Inf. Comput. Sci.* **35** 771 (1995)
155. J W Dunn, A J Hopfinger, C Catana, C Duraiswami *J. Med. Chem.* **39** 4825 (1996)
156. A M Doweyko *J. Med. Chem.* **31** 1396 (1988)
157. A M Doweyko *J. Math. Chem.* **7** 273 (1991)
158. A M Doweyko, W B Mattes *Biochemistry* **31** 9388 (1992)
159. A M Doweyko *J. Med. Chem.* **37** 1769 (1994)
160. J J Kaminsky, A M Doweyko *J. Med. Chem.* **40** 427 (1997)
161. A N Jain, T G Dietterich, R H Lathrop, D Chapman, R E Critchlow, B E Bauer, T A Webster, T Lozanj-Perez *J. Comput.-Aid. Mol. Design* **8** 635 (1994)
162. A N Jain, K Koile, D Chapman *J. Med. Chem.* **37** 2315 (1994)
163. A N Jain, N L Harris, J Y Park *J. Med. Chem.* **38** 1295 (1995)
164. R Todeschini, M Lasagni, E Marengo *J. Chemometr.* **8** 263 (1994)
165. R Todeschini, P Gramatica, R Provenzani *Chemometr. Intell. Lab. Syst.* **27** 221 (1995)
166. R Todeschini, M Vigni, R Provenzani, A Finizio, P Gramatica *Chemosphere* **32** 1527 (1996)
167. R Todeschini, C Bettiol, G Giurin, P Gramatica, P Miana, E Argese *Chemosphere* **33** 71 (1996)
168. R Todeschini, P Gramatica *Quantum Struct. – Act. Relat.* **16** 113 (1997)
169. R Todeschini, P Gramatica *Quantum Struct. – Act. Relat.* **16** 120 (1997)
170. P G Mezey *J. Chem. Inf. Comput. Sci.* **32** 650 (1992)
171. P G Mezey *Shape in Chemistry. Introduction to Molecular Shape and Topology* (New York: VCH, 1992)
172. P D Walker, P G Mezey *J. Am. Chem. Soc.* **115** 12 423 (1993)
173. P D Walker, P G Mezey *J. Am. Chem. Soc.* **116** 12 022 (1994)
174. P G Mezey, Z Zimpel, P Warburton, P D Walker, D G Irvine, S L Dixon, B Greenberg *J. Chem. Inf. Comput. Sci.* **36** 602 (1996)
175. A Murcko, M A Murcko *J. Med. Chem.* **38** 4953 (1995)
176. M Charton, in *Lipophilicity in Drug Action and Toxicology* (Eds V Pliska, B Testa, H Waterbeemd) (Weinheim: VCH, 1996) p. 387
177. C Dauwe, B Sellergren *J. Chromatogr., A* **753** 191 (1996)

178. P G Tratnyek, in *Perspectives in Environmental Chemistry* (Ed. D L Macalady) (New York: Oxford University Press, 1998) p. 167
179. L Quadri, A Cerri, P Ferrari, E Folpini, M Mabilia, P Melloni *J. Med. Chem.* **39** 3385 (1996)
180. Z Nikolovska-Coleska, L Suturkova, K Dorevski, A Krbavcic, T Solmajer *Quantum Struct. – Act. Relat.* **17** 7 (1998)
181. L P Hammet *J. Am. Chem. Soc.* **59** 66 (1937)
182. M Charton *Prog. Phys. Org. Chem.* **13** 119 (1981)
183. C Hanch, A Leo, R W Taft *Chem. Rev.* **91** 165 (1991)
184. C Hansch, A Leo, D Hoekman *Exploring QSAR. Hydrophobic, Electronic and Steric Constants* (Washington, DC: American Chemical Society, 1995)
185. G Schuurman *Quantum Struct. – Act. Relat.* **9** 326 (1990)
186. O G Mekenyan, G D Veith *SAR QSAR Environ. Res.* **2** 129 (1994)
187. S Karabunarliev, O G Mekenyan, W Karcher, C L Russom, S P Bradbury *Quantum Struct. – Act. Relat.* **15** 311 (1996)
188. B W Clare, C T Supuran *J. Pharm. Sci.* **83** 768 (1994)
189. K J Miller *J. Am. Chem. Soc.* **112** 8533 (1990)
190. R M Smith, P S Pearlman, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giraldo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p. 222
191. J Jaerv, K Sak, M Eller, P Ek, A Engstroem, I Engstroem *Bioorg. Chem.* **24** 159 (1996)
192. J D Hirst *J. Med. Chem.* **39** 3526 (1996)
193. P Sjoberg, in *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry* (Eds H van de Waterbeemd, B Testa, G.Folker) (Weinheim: VCH, 1997) p. 83
194. S -S So, M Karplus *J. Med. Chem.* **39** 5246 (1996)
195. V Pliska, B Testa, H Waterbeemd, in *Lipophilicity in Drug Action and Toxicology* (Eds V Pliska, B Testa, H Waterbeemd) (Weinheim: VCH, 1996) p. 1
196. A Avdeef, in *Lipophilicity in Drug Action and Toxicology* (Eds V Pliska, B Testa, H Waterbeemd) (Weinheim: VCH, 1996) p. 109
197. J C Dearden, G M Bresnen *Quantum Struct. – Act. Relat.* **7** 133 (1988)
198. L-G Danielsson, Y-H Zhang *Trends Anal. Chem.* **15** 188 (1996)
199. M L C Montanari, C A Montanari, D P Veloso, A E Beezer, J C Mitchell, P L O Volpe, *Quantum Struct. – Act. Relat.* **17** 102 (1998)
200. V N Viswanadhan, A K Ghose, G R Revankar, R K Robins *J. Chem. Inf. Comput. Sci.* **29** 163 (1989)
201. R F Rekker, R Mannhold *Calculation of Drug Lipophilicity* (Weinheim: VCH, 1992)
202. G Klopman, J-K Wang, M J Dimayuga *J. Chem. Inf. Comput. Sci.* **34** 752 (1994)
203. E Kellogg, D Abraham *J. Comput.-Aid. Mol. Design* **5** 545 (1991)
204. H Waterbeemd, H Karajiannis, H Kansy, D Obrecht, K Muller, C Lehmann, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giraldo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p. 78
205. N Bodor, M J Huang *J. Pharm. Sci.* **81** 272 (1992)
206. M H Abraham, H S Chadra, in *Lipophilicity in Drug Action and Toxicology* (Eds V Pliska, B Testa, H Waterbeemd) (Weinheim: VCH, 1996) p. 311
207. H Waterbeemd, B Testa *Adv. Drug. Res.* **16** 87 (1987)
208. N El Tayar, B Testa, in *Trends in QSAR and Molecular Modelling 92* (Ed. C G Wermuth) (Leiden: ESCOM, 1993) p. 101
209. O A Raevsky, K-J Shaper, J K Seydel *Quantum Struct. – Act. Relat.* **14** 433 (1995)
210. O A Raevsky, in *Computer-Aided Lead Finding and Optimization. Current Tools for Medicinal Chemistry* (Eds H Waterbeemd, B Testa, G Folkers) (Wienheim: VCH, 1997) p. 367
211. H Waterbeemd, R Mannhold, in *Lipophilicity in Drug Action and Toxicology* (Eds V Pliska, B Testa, H Waterbeemd) (Weinheim: VCH, 1996) p. 401
212. R Mannhold, K Dross *Quantum Struct. – Act. Relat.* **15** 403 (1996)
213. H Waterbeemd, R Mannhold *Quantum Struct. – Act. Relat.* **15** 410 (1996)
214. P A Carrupt, P Gaillard, F Billois, P Weber, B Testa, C Meyer, S Perez, in *Lipophilicity in Drug Action and Toxicology* (Eds V Pliska, B Testa, H Waterbeemd) (Weinheim: VCH, 1996) p. 195
215. C Altomare, A Carotti, V Carta, S Kneubihler, P A Carrupt, B Testa, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giraldo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p. 463
216. M Laguerre, M Saux, A Carpy, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giraldo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p. 588
217. I Rozas, M Martin *J. Chem. Inf. Comput. Sci.* **36** 872 (1996)
218. S Immel, F W Lichtenhaler *Leibigs. Ann. Chem.* 39 (1996)
219. T Fujita *Quantum Struct. – Act. Relat.,* **16** 107 (1997)
220. B Testa, P- A Carrupt, P Gaillard, F Billois, P Weber *Pharm. Res.* **13** 335 (1996)
221. C Hansch, D Hoekman, A Leo, L Zhang, P Li *Toxicol. Lett.* **79** 45 (1995)
222. M T D Cronin, J C Dearden *Quantum Struct. – Act. Relat.* **14** 1 (1995)
223. M T D Cronin, J C Dearden *Quantum Struct. – Act. Relat.* **14** 117 (1995)
224. M T D Cronin, J C Dearden *Quantum Struct. – Act. Relat.* **14** 329 (1995)
225. P J Goodford *J. Med. Chem.* **28** 849 (1985)
226. R W Taft, in *Steric Effects in Organic Chemistry* (Ed. M S Newman) (New York: Wiley, 1956) p. 556
227. M Charton *Environ. Health Perspect.* **61** 229 (1985)
228. M Charton, in *Similarity Models in Organic Chemistry, Biochemistry and Related Fields* (Eds R I Zalewski, T M Krygowski, J Shorter) (Amsterdam: Elsevier, 1991) p. 629
229. K H Kim, Y C Martin, in *QSAR, Rational Approaches to the Design of Bioactive Compounds* (Eds C Silipo, A Vittoria) (Amsterdam: Elsevier, 1991) p. 151
230. F Sanz, M Martin, F Lapena, F Manaut *Quantum Struct. – Act. Relat.* **5** 54 (1986)
231. I Alkorta, H O Villar *J. Med. Chem.* **37** 210 (1994)
232. D B Kireev, O A Raevsky, V I Fetisov, in *Trends in QSAR and Molecular Modelling 92* (Ed. C G Wermuth) (Leiden: ESCOM, 1993) p. 331
233. L Bonati, E Fraschini, M Lasagni, E P Madoni, D Pitea *J. Mol. Struct. (Theochem)* **340** 83 (1995)
234. R Bureau, J C Lancelot, H Prunier, S Rault *Quantum Struct. – Act. Relat.* **15** 373 (1996)
235. A K Bhattacharjee, J M Karle *J. Med. Chem.* **39** 4622 (1996)
236. R Carbo, B Calabuig *Int. J. Quantum Chem.* **42** 1681 (1992)
237. R Carbo, B Calabuig *J. Chem. Inf. Comput. Sci.* **32** 600 (1992)
238. E Besalu, L Amat, X Fradera, R Carbo, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giraldo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p. 396
239. X Fradera, L Amat, E Besalu, R Carbo-Dorca *Quantum Struct. – Act. Relat.,* **16** 25 (1997)
240. R Carbo, L Leyde, M Arnau *Int. J. Quantum Chem.* **17** 1185 (1980)
241. A C Good, S-S So, W G Richard *J. Med. Chem.* **36** 433 (1993)
242. R Carbo, B Calabuig *J. Mol. Struct. (Teochem)* **254** 517 (1992)
243. R Carbo, E Besalu, B Calabuig, L Vera *Adv. Quant. Chem.* **25** 253 (1994)
244. D J Wild, P Willett *J. Chem. Inf. Comput. Sci.* **36** 159 (1996)
245. L Pauling, P Pauling *Chemistry* (San Francisco: W H Freeman and Co., 1975)
246. P Magee *Sci. Total Environ.* **109**/**110** 155 (1991)
247. N Froloff, A Faurion, P M Leod *Chem. Sens.* **21** 425 (1996)
248. E Forgas, T Cserhati *J. Liq. Chromatogr., Relat. Technol.* **19** 1849 (1996)
249. O A Raevsky *Teor. Eksp. Khim.* **22** 450 (1986) [d]
250. O A Raevsky *Usp. Khim.* **59** 375 (1990) [*Russ. Chem. Rev.* **59** 219 (1990)]
251. O A Raevsky *J. Phys. Org. Chem.* **10** 405 (1997)
252. O A Raevsky, V Yu Grigor'ev, V P Solov'ev, I V Martynov *Dokl. Akad. Nauk SSSR* **299** 1433 (1988) [e]
253. R S Drago, B B Wayland *J. Am. Chem. Soc.* **87** 3571 (1965)
254. R S Drago *Structure and Bonding* (Heidelberg: Springer, 1973)
255. M K Kroeger, R S Drago *J. Am. Chem. Soc.* **103** 3250 (1981)

256. A V Iogansen *Teor. Eksp. Khim.* **7** 302 (1971) [d]
257. O A Raevsky, V P Novikov *Khim.-Farm. Zh.* **16** 583 (1982) [b]
258. I V Martynov, O A Raevsky *Dokl. Akad. Nauk SSSR* **265** 664 (1982)
259. I V Martynov, O A Raevsky *Vestn. Akad. Nauk SSSR* (7) 93 (1983)
260. I V Martynov, O A Raevsky *Zh. Vses. Khim. O-va im D I Mendeleeva* **28** (1983) [a]
261. O A Raevsky, M M Gilyazov, Ya A Levin *Zh. Obshch. Khim.* **53** 563 (1983) [f]
262. O A Raevsky, M M Gilyazov, Ya A Levin *Zh. Obshch. Khim.* **53** 1720 (1983) [f]
263. O A Raevsky, M M Gilyazov, Ya A Levin *Zh. Obshch. Khim.* **53** 1724 (1983) [f]
264. O A Raevsky, V Yu Grigor'ev, V P Solov'ev *Khim.-Farm. Zh.* **18** 578 (1984) [b]
265. T G Valeeva, Ya A Levin, O A Raevsky *Zh. Obshch. Khim.* **55** 1590 (1985) [f]
266. O A Raevsky, V Yu Grigor'ev, V P Solov'ev, A I Ivanov, V B Sokolov, I V Martynov *Zh. Obshch. Khim.* **57** 2073 (1987) [f]
267. O A Raevsky, V Yu Grigor'ev, V P Solov'ev, I V Martynov *Dokl. Akad. Nauk SSSR* **298** 1166 (1988) [e]
268. O A Raevsky, A F Solotnov, V P Solov'ev *Zh. Obshch. Khim.* **57** 1240 (1987) [f]
269. A M Sapegin, O A Raevsky, V V Chistyakov, I V Martynov *Khim.-Farm. Zh.* **21** 578 (1987) [b]
270. O A Raevsky, in *QSAR in Drug Design and Toxicology* (Eds D Hadzy, B Jerman-Blazic) (Amsterdam: Elsevier, 1987) p. 31
271. O A Raevsky, V Yu Grigor'ev, V P Solov'ev *Khim.-Farm. Zh.* **23** 1294 (1989) [b]
272. O A Raevsky, V Ju Grigor'ev, D B Kireev, N S Zefirov *Quantum Struct.–Act. Relat.* **11** 49 (1992)
273. R W Taft, D Gurka, L Loris, P R Schleyer, J W Rakshys *J. Am. Chem. Soc.* **91** 4801 (1969)
274. J-L M Abboud, L Bellon *Ann. Chim.* **5** 63 (1970)
275. M H Abraham, P E Grellier, D V Prior, R W Taft, J J Moris, P J Taylor, C Lourence, M Bertholot, R M Doherty, M J Kamlet, J-L M Abboud, K Sraidi, G Guiheneuf *J. Am. Chem. Soc.* **110** 8534 (1988)
276. M H Abraham *Chem. Soc. Rev.* **22** 73 (1993)
277. R W Taft, W J Shuely, R H Doherty, M J Kamlet *J. Org. Chem.* **53** 1737 (1988)
278. P-C Maria, J-F Gal, J Franceschi, E Fargin *J. Am. Chem. Soc.* **109** 483 (1987)
279. M H Abraham, in *Quantitative Treatments of Solute/Solvent Interactions. Theoretical and Computational Chemistry* Vol. 1 (Eds P Polizer, J S Murray) (Amsterdam: Elsevier, 1994) p. 83
280. O A Raevsky, V Ju Grigor'ev, V P Solov'ev, in *QSAR Rational Approaches to the Design of Bioactive Compounds* (Eds C Silipo, A Vittoria) (Amsterdam: Elsevier, 1991) p. 135
281. O A Raevsky, A M Sapegegin *Khim.-Farm. Zh.* **24** 43 (1990) [b]
282. O A Raevsky, A M Sapegin, V V Chistyakov, A N Razdol'skii, N S Zefirov *Dokl. Akad. Nauk SSSR* **309** 623 (1989) [e]
283. H-J Schneider, V Rmdiger, O Raevsky *J. Org. Chem.* **58** 3648 (1993)
284. H-J Schneider, T Blatter, A Eliseev, V Rmdiger, O A Raevsky *Pure Appl. Chem.* **65**. 2329 (1993)
285. O A Raevsky, I G Shnaider *Zh. Org. Khim.* **31** 1793 (1995) [g]
286. V P Solov'ev, N N Strakhova, O A Raevsky, V Rudiger, H-J Schneider *J. Org. Chem.* **61** 5221 (1996)
287. O A Raevsky, V V Chistyakov, R S Agabekyan, A M Sapegin, N S Zefirov *Bioorg. Khim.* **16** 1509 (1990) [h]
288. I I Kitova, O A Raevsky, V G Blinova, N S Zefirov, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giralo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p.160
289. S O Bachurin, N V Lukoyanov, L N Petrova, L S Solyakov, S E Tkachenko, O A Raevsky *Dokl. Akad. Nauk* **346** 549 (1996) [e]
290. O Raevsky, A Sapegin, N Zefirov *Quantum Struct.–Act. Relat.* **13** 412 (1994)
291. A N Razdol'skii, A M Sapegin, O A Raevsky, E V Shaveleva, G I Smirnova, N G Rozhkova *Khim.-Farm. Zh.* **26** (9–10) 92 (1992)
292. U Z Mirkhodzhaev, N V Lukoyanov, I I Kitova, Kh F Abdulaev, O A Raevsky, A K Tashmukhamedova *Uzb. Biol. Zh.* (1) 8 (1991)
293. O Raevsky, V Ju Grigor'ev, D B Kireev, N S Zefirov *J. Chim. Phys. Phys.-Chim. Biol.* **89** 1747(1992)
294. O A Raevsky, V Ju Grigor'ev, E Mednikova, in *Trends in QSAR and Molecular Modelling 92* (Ed. C G Wermuth) (Leiden: ESCOM, 1993) p. 116
295. H Waterbeemd, G Gamenisch, G Folkers, O A Raevsky *Quant. Struct.-Act. Relat.* **15** 480 (1996)
296. J Li, Y Zhang, A J Dallas, P W Carr *J. Chromatogr.* **550** 101 (1991)
297. H Kubinyi, in *Burger's Medicinal Chemistry and Drug Discovery* (Ed. M E Wolff) (New York: Willey, 1995) p. 497
298. J J Hostynek, P S Magee *Quantum Struct.–Act. Relat.* **16** 473 (1997)
299. J Halova, O Strouf, P Z A Sochozova, N Uchida, T Yuzyri, K Sakakibara, M Hirota *Quantum Struct.–Act. Relat.* **17** 37 (1998)
300. W H Streng *Drug Design Today* **2** 415 (1997)
301. L Y Wilson, G R Famini *J. Med. Chem.* **34** 1668 (1991)
302. G R Famini, V P Loumbev, E K Frykman, L Y Wilson *Quantum Struct.–Act. Relat.* **17** 558 (1998)
303. G R Famini, W P Ashman, A P Miciewicz, L Y Wilson *Quantum Struct.–Act. Relat.* **11** 162 (1992)
304. G R Famini, L Y Wilson, C A Penski *J. Phys. Org. Chem.* **5** 395 (1992)
305. A H Lowrey, C J Cramer, J J Urban, G R Famimi *Comput. Chem.* **19** 209 (1995)
306. N A Chester, G R Famini, M V Haley, C W Kurnas, P A Sterling, L Y Wilson *ACS Symp. Ser.* **643** 191 (1996)
307. R D Cramer III, D E Patterson, J D Bunce *J. Am. Chem. Soc.* **110** 5959 (1988)
308. R J Vaz *Quantum Struct.–Act. Relat.* **16** 303 (1997)
309. W Lindberg, J A Persson, S Wold *Anal. Chem.* **55** 643 (1983)
310. K Hasegava, T Kimura, K Funatsu *Quantum Struct.–Act. Relat.* **16** 219 (1997)
311. C Navajas, A Poso, K Tuppurainen, J Gynther *Quantum Struct.–Act. Relat.* **15** 189 (1996)
312. D B Kireev, J R Chretien, O A Raevsky *Eur. J. Med. Chem.* **30** 395 (1995)
313. W E Steinmetz *Quantum Struct.–Act. Relat.* **15** 1 (1996)
314. M Akamatsu, Y Ozoe, T Ueno, T Fujita, K Mochida, T Nakamura, F Matsumura *Pest. Sci.* **49** 319 (1997)
315. T I Oprea, A E Garcia *J. Comput.-Aid. Mol. Design* **10** 186 (1996)
316. S Sicsic, I Serraz, J Andrieux, B Bremont, M Mathe-Allanmat, A Poncet, S Shen, M Langlois *J. Med. Chem.* **40** 739 (1997)
317. F Corelli, F Manetti, A Tafi, G Campiani, V Nacci, M Botta *J. Med. Chem.* **40** 125 (1997)
318. D N A Boobbyer, P J Goodford, P M McWhinne, R C Wade *J. Med. Chem.* **32** 1083 (1989)
319. R C Wade, K J Clark, P J Goodford *J. Med. Chem.* **36** 140 (1993)
320. R C Wade, P J Goodford *J. Med. Chem.* **36** 148 (1993)
321. G Cruciani, P J Goodford *J. Mol. Graph.* **12** 116 (1994)
322. A M Davis, N P Gensmantel, E Johansson, D P Marriot *J. Med. Chem.* **37** 963 (1994)
323. E E Polymeropoulos, N Hofgen *Quantum Struct.–Act. Relat.* **16** 231 (1997)
324. T Langer *Quantum Struct.–Act. Relat.* **15** 469 (1996)
325. M Pastor, G Cruciani *J. Med. Chem.* **38** 4637 (1995)
326. M Itzstein, J C Dyason, S W Oliver, H F White, W-Y Wu, G B Kok, M S Pegg *J. Med. Chem.* **39** 388 (1996)
327. S Clementi, G Cruciani, P Fifi, D Riganelli, R Valigi *Quantum Struct.–Act. Relat.* **15** 108 (1996)
328. V P Novikov, O A Raevsky *Khim.-Farm. Zh.* **16** 574 (1982) [b]
329. I Csorvassy, L Tozser, in *QSAR. Rational Approaches to the Design of Bioactive Compounds* (Eds C Silipo, A Vittoria) (Amsterdam: Elsevier, 1991) p. 193
330. J W King, R Kassel, B B King *Int. J. Quantum Chem. Quantum Biol. Symp.* **17** 27 (1990)
331. J W King, R J Kassel *Int. J. Quantum Chem. Quantum Biol. Symp.* **18** 289 (1991)
332. J W King, R J Kassel *Int. J. Quantum Chem. Quantum Biol. Symp.* **19** 179 (1992)

333. J W King *Int. J. Quantum Chem. Quantum Biol. Symp.* **20** 139 (1993)
334. G R Famini, R J Kassel, J W King, L Y Wilson *Quantum Struct. – Act. Relat.* **10** 344 (1991)
335. J W King *Int. J. Quantum Chem. Quantum Biol. Symp.* **21** 209 (1994)
336. B I Kuchkaev, B A Knyazev, V K Kurochkin, V A Petrunin *Khim.-Farm. Zh.* **26** (6) 60 (1992) [b]
337. B I Kuchkaev, B A Knyazev, V K Kurochkin, V A Petrunin *Khim.-Farm. Zh.* **31** (8) 30 (1997)
338. J Schuur, P Selzer, J Gasteiger *J. Chem. Inf. Comput. Sci.* **36** 334 (1996)
339. J Gasteiger, J Sadowski, J Schuur, P Selser, L Steinhauer, V Steinhauer *J. Chem. Inf. Comput. Sci.,* **36** 1030 (1996)
340. O Raevsky, L Dolmatova, V Grigor'ev, I Lisyansky, S Bondarev, in *QSAR and Molecular Modelling. Concepts, Computational Tools and Biological Applications* (Eds F Sanz, J Giraldo, F Manaut) (Barcelona: J R Prous Science Publishers, 1995) p. 241
341. R Franke *Theoretical Drug Design Method* (Berlin: Springer, 1984)
342. J McFarland, D Gan *Drug Inform. J.* **24** 705 (1990)
343. J McFarland, D Gan *Quantum Struct. – Act. Relat.* **13** 11 (1994)
344. H Matter *J. Med. Chem.* **40** 1219 (1997)
345. E J Ariens, A M Simonis *Top. Curr. Chem.* **52** 1 (1974)
346. K-t Chung, L Kirkovsky, A Kirkovsky, W P Purcell *Mutat. Res.* **387** 1 (1997)
347. C A Montanari, M S Tute *Quantum Struct. – Act. Relat.* **16** 480 (1997)
348. Y Nezu, N Wada, F Yoshida, M Miyazawa, T Shimizu, T Fujita *Pestic. Sci.* **52** 343 (1998)
349. H Kubinyi *Drug Design Today* **2** 457 (1997)
350. V Austel, in *Chemometric Methods in Molecular Design* Vol. 2 (Ed. H Waterbeemd) (Dusseldorf: VCH, 1995) p. 49
351. W Dunn, S Wold, in *Chemometric Methods in Molecular Design* Vol. 2 (Ed. H Waterbeemd) (Dusseldorf: VCH, 1995) p. 179
352. S Wold, in *Chemometric Methods in Molecular Design* Vol. 2 (Ed. H Waterbeemd) (Dusseldorf: VCH, 1995) p. 195
353. R Franke, A Gruska, in *Chemometric Methods in Molecular Design* Vol. 2 (Ed. H Waterbeemd) (Dusseldorf: VCH, 1995) p. 115
354. A R Katritzky, L Mu, M Karelson *J. Chem. Inf. Comput. Sci.* **36** 1162 (1996)
355. S C Basak, B D Gute, G D Grundwald *J. Chem. Inf. Comput. Sci.* **36** 1054 (1996)
356. R Carcia-Domenich, C de Grigoria Alapont, J V de Julian-Ortiz, J Galvez, L Popa *Bioorg. Med. Chem. Lett.* **7** 567 (1997)
357. S P Bradbury, O G Mekenyan, G T Ankley *Environ. Toxicol. Chem.* **15** 1945 (1996)
358. D A Dawson, T W Schultz, R S Hunter *Teratogen Carcinogen Mutagen* **16** 109 (1996)
359. N K Roy, E S J Nidiry, K Vasu, S Bedi, B Lalljee, B Singh *J. Agric. Food Chem.* **44** 3971(1996)
360. A Cappelli, M Anzini, S Vomero, P G De Benedetti, M C Menziani, G Giorgi, C Manzoni *J. Med. Chem.* **40** 2910 (1997)
361. J Polanski, J Gasteiger, M Wagener, J Sadowski *Quantum Struct. – Act. Relat.* **17** 27 (1998)
362. M Mracec, S Muresan, M Mracec, Z Simon, G Naray-Szabo *Quantum Struct. – Act. Relat.* **16** 459 (1997)
363. R Vaz *Weed. Sci.* **44** 718(1996)
364. M Rowley, J J Kulagowski, A P Watt, D Rathbone, G I Stevenson, R W Carling, R Baker, G R Marshall, J A Kemp, A C Foster, S Grimwood, R Hargreaves, C Hurley, K L Saywell, M D Tricklebank, P D Leeson *J. Med. Chem.* **40** 4053 (1997)
365. Y H Zhao, M T D Cronin, J C Dearden *Quantum Struct. – Act. Relat.* **16** 131 (1997)
366. M J B Mengelers, P E Hougee, L H M Janssen, A S J P A M van Miert *J. Vet. Pharmacol. Ther.* **20** 276 (1997)
367. J S Tokarski, A J Hopfinger *J. Chem. Inf. Comput. Sci.* **37** 792 (1997)
368. P Chiba, M Hitzler, E Richter, M Huber, C Tmej, E Giovagnoni, G Ecker *Quantum Struct. – Act. Relat.* **16** 361 (1997)
369. O A Raevsky, K-J Schaper *Eur. J. Med. Chem.* **33** 799 (1998)
370. K C S C Liu, S-S Lee, M-T Lin, C-W Chang, C-L Liu, J-Y Hsu, S Ren, E J Lien *Med. Chem. Res.* **7** 168 (1997)
371. R Garg, S P Gupta *J. Enzyme Inhib.* **11** 171 (1997)

[a] — *Mendeleev Chem. J. (Engl.Transl.)*
[b] — *Pharm. Chem. J. (Engl. Transl.)*
[c] — *Russ. J. Coord. Chem. (Engl. Transl.)*
[d] — *Theor. Exp. Chem. (Engl. Transl.)*
[e] — *Dokl. Chem. Technol., Dokl. Chem. (Engl. Transl.)*
[f] — *Russ. J. Gen. Chem. (Engl. Transl.)*
[g] — *Russ. J. Org. Chem. (Engl. Transl.)*
[h] — *Russ. J. Bioorg. Chem. (Engl. Transl.)*