Russian Chemical Reviews **65** (2) 117 – 130 (1996)
© 1996 Russian Academy of Sciences and Turpion Ltd

UDC 541.12 + 681.142

Computer-aided studies of reaction mechanisms

A V Zeigarnik, L G Bruk, O N Temkin, V A Likholobov, L I Maier

Contents

I. A rational strategy for the mechanistic studies of catalytic reactions 117
II. The formulation of hypotheses on the mechanisms of catalytic reactions 119
III. Programs for the formulation of hypotheses on the mechanisms of complex reactions 120
IV. Conclusion 129

Abstract. The principal stages in a rational strategy for the investigation of reaction mechanisms are described, in which the key stage is the formulation of hypotheses at the initial stage of the study. Computer programs are examined in which hypotheses are formulated about the mechanisms of complex reactions with participation of the user. Empirical and formal-logical algorithms and programs as well as various aspects of their application are discussed. The programs are compared with those for computer-aided organic synthesis design, which are similar in their structures and algorithms. The bibliography includes 119 references.

I. A rational strategy for the mechanistic studies of catalytic reactions

Until recently, the dominant approach to the study of the mechanisms of catalytic reactions involved a method which might be called 'inductive'. Its essential feature is that it implies the following stages of the study. Initially an experiment is carried out (mainly a kinetic experiment). A mathematical description is found for the results of this experiment. The reaction mechanism is deduced from the mathematical description. A chemical content is then attributed to this scheme, abstract symbols (such as A, B, C, X, and Y) being replaced by the formulae of specific substances. The reaction mechanism is thus obtained. This approach is so deeply ingrained in the minds of many researchers that, despite the regularly held conferences devoted to the methodological problems of kinetic studies in catalysis, it has been scarcely criticised. On the other hand, the cognitive potentials of this method are extremely limited. Evidently an arbitrary number of models to describe the same experimental data may be proposed, and it is impossible to unambiguously deduce a reaction mechanism from such a description (especially in complex cases).

In practice, investigators studying reaction mechanisms in accordance with the scheme are guided initially by some hypothesis about the mechanism (there is usually only one hypothesis), justifying and elaborating it when necessary if the experimental data do not fit within the framework of the existing ideas. Doing so, they understand that the interpretation of the kinetic curves does point out how the existing hypothesis should be modified.

Analysis of the literature on philosophy [1–5] and the applications of computational methods and mathematical statistics in chemical kinetics [6–10] has clearly shown that the inductive method for the investigation of reaction mechanisms is methodologically unsound. However, the crisis of the inductive method in chemical kinetics did not arise solely as a result of studies by philosophers and mathematicians, who are more skilled in logic than other researchers. The knowledge about reaction mechanisms, including catalytic reaction mechanisms, expanded significantly over several years. As this knowledge accumulated, ideas about the complexity of catalytic processes also changed. Nowadays investigators put forward much more complex mechanisms for the description of catalytic reactions than before. These mechanisms involve intermediates, which have more complex structures, and new types of elementary steps. They include a greater number of intermediates and steps, so that the kinetic models (mathematical equations) corresponding to such mechanisms have become appreciably more complex. It became clear that the path 'from the rate law to a mechanism' is no longer useful for mechanistic studies and kinetic modeling. In general this method is fruitful only in relatively simple cases.

The inductive method was replaced by the deductive method, which can be represented by the following logical sequence:

formulation of hypotheses ⟶ experiment design
⟶ experiment ⟶ discrimination between hypotheses
⟶ mechanisms not in conflict with experiment.

This method is quite natural and has found extensive applications in many branches of science. Interestingly, the American Chemical Society has recommended that papers be written according to a plan that corresponds to the stages of the investigation within the framework of the method.[10] Its undoubted advantage is that the hypotheses are formulated explicitly, making it possible to estimate the width of the 'search space' and to draw conclusions about the value of the 'winning' hypothesis.

Figuratively speaking, the investigator actually addresses questions to nature. According to Polya's apt expression, 'Nature may answer Yes or No, but it whispers one answer and thunders the others; its Yes is provisional and its No is definitive'.[11] Thus the difference between the two methods described here is clear: an in the case of 'inductive' approach, the an investigator is in a search for an arbitrary 'yes', whereas, in the case of the 'deductive' method, the hypotheses for which an unambiguous 'no' is obtained are to be rejected. It is then important to know not only the hypotheses remaining after the

A V Zeigarnik, L G Bruk, O N Temkin Laboratory for Chemical Kinetics and Catalysis, Lomonosov Academy of Fine Chemical Technology, prosp. Vernadskogo 86, 117571 Moscow, Russian Federation. Fax (7-095) 430 79 83
V A Likholobov (Laboratory for Catalysis by Metal Complexes), **L I Maier** (Laboratory for Information Systems and Computer Programs), Boreskov Institute of Catalysis, Siberian Division of the Russian Academy of Sciences, prosp. akad. Lavrent'eva 5, 630090 Novosibirsk, Russian Federation. Fax (7-383) 235 57 76

Received 10 May 1995
Uspekhi Khimii **65** (2) 125 – 139 (1996); translated by A K Grzybowski

experimental test, but also the entire set of the initial hypotheses. The method works better if the set of hypotheses is more complete.

Several papers have been devoted to the popularisation of the deductive method.[12-14] Comparison of the two strategies allowed us to conclude that the concept of a 'reaction mechanism' combines two types of information, topological and physicochemical, which are interrelated and interdependent.[14]

The topological information characterises the structure of the chemical reaction mechanism (the interrelation of the intermediates, reaction paths, and their coupling). The identification of the mechanism structure is based on kinetic and physicochemical methods. The structure of the mechanism can be well represented by graphs, which are widely used in chemistry in general [15-20] and in particular the theory of chemical reaction mechanisms.[21-23]

The physicochemical information reflects the specific chemical content of the mechanism — the composition and structure of intermediates, their reactivities (the values of rate constants, the structures of transition states). The topological and physicochemical types of information complement one another and cannot be obtained independently; i.e., 'in general' one cannot first establish the kinetic scheme and then attribute chemical content to it.[14]

An important argument in favour of the idea that hypotheses must be formulated first is the multiplicity of mechanisms in catalysis. By multiplicity of mechanisms, we understand the possibility for the reaction to occur via several different mechanisms (and different intermediates) and the possibility of several such mechanisms dominating over others depending on the type of catalyst (in the case of metal-complex catalysis, on the nature of metal, its oxidation state, and even on the ligands coordinated to this metal).[24]

For example, the reaction

$$C_2H_2 + CO + ROH \longrightarrow CH_2{=}CHCOOR$$

may occur via three mechanisms:[24,25]
(1) the hydride mechanism:

$$HMX \xrightarrow{C_2H_2} CH_2{=}CHMX \xrightarrow{CO}$$

$$\longrightarrow CH_2{=}CHCOMX \xrightarrow{ROH} CH_2{=}CHCOOR + HMX,$$

(2) the alkoxide mechanism:

$$M^{II} \xrightarrow[-H^+]{ROH} M^{II}OR \xrightarrow{CO} M^{II}COOR \xrightarrow{C_2H_2}$$

$$\longrightarrow M^{II}CH{=}CHCOOR \xrightarrow{H^+} CH_2{=}CHCOOR + M^{II},$$

(3) the metallocyclic mechanism:

$$(M)_n \xrightarrow{C_2H_2} (M)_n\langle \overset{CH}{\underset{CH}{\|}} \rangle \xrightarrow{CO} \overset{(M)_n-CO}{\underset{HC=CH}{|\quad|}} \xrightarrow{ROH}$$

$$\longrightarrow H-(M)_n-CH{=}CHCOOR \longrightarrow (M)_n + CH_2{=}CHCOOR.$$

The hydride mechanism is more probable for Pd(0) complexes; the alkoxide mechanism is more probable for Pd(II) complexes; the metallocyclic mechanism is more probable in solutions of Pd(I) clusters.

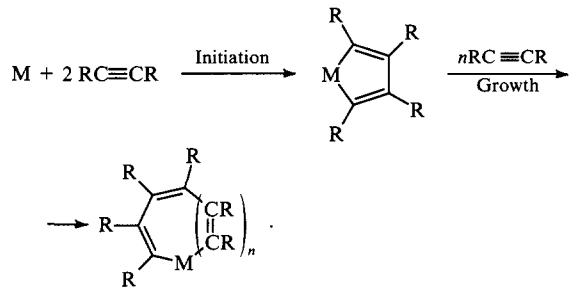
The multiplicity of mechanisms and the increase in their complexity are best manifested in the polymerisation reactions of alkynes.[24] The simplest mechanism includes two steps — chain initiation and chain propagation.

$$M{-}X + R^1C{\equiv}CR^2 \xrightarrow{\text{Initiation}}$$

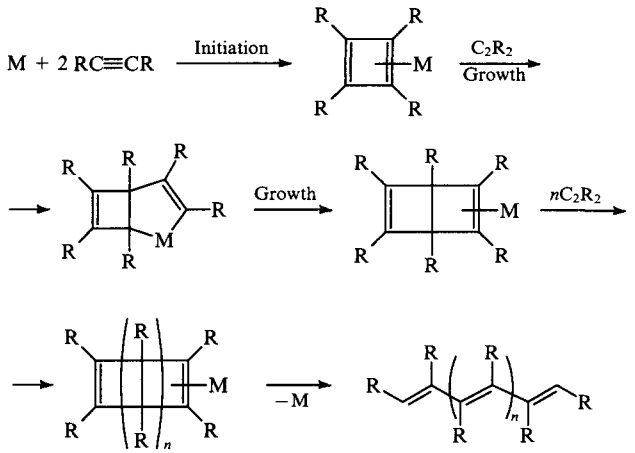
$$\longrightarrow M{-}CR^1{=}CR^2{-}X \xrightarrow[\text{Growth}]{nR^1C{\equiv}CR^2}$$

$$\longrightarrow M{-}(CR^1{=}CR^2)_n{-}CR^1{=}CR^2X$$

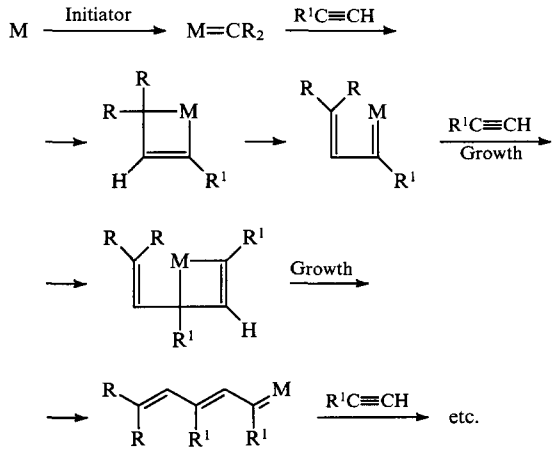
Other alkyne polymerisation mechanisms are the metallocyclic mechanism



and the [2 + 2]-cycloaddition mechanism.



A mechanism involving several chain propagation and initiation steps is also possible:



As can be seen from the above examples, the same reactions can occur via different mechanisms. The researcher usually knows from the literature some plausible reaction pathways and uses them explicitly or implicitly as hypotheses.

II. The formulation of hypotheses on the mechanisms of catalytic reactions

The formulation of hypotheses is a creative task and cannot be fully formalised either now or in the future. However, certain elements of this procedure can be formalised. Furthermore, it is the investigator's intuition combined with the ability of a computer to execute complex algorithmic procedures that constitutes a good basis for the formulation of mechanistic hypotheses.

Three groups of hypotheses were discussed:[14] (1) hypotheses about the nature of a catalyst, (2) hypotheses about elementary steps, and (3) hypotheses about the conjugation nodes. In addition, a fourth group, which combines the elements of the above three groups, was considered: hypotheses about the reaction mechanism as a whole. Each of these groups of hypotheses can serve as a basis for the design of discriminating experiments (hereafter, we shall use the term 'experiment design' in a broader sense than that adopted by the specialists in statistics).

In the case of homogeneous metal-complex catalysis, the group of hypotheses about the nature of the catalyst includes the following problems: the condition of a catalyst in solution, the existence of complexation equilibria, the specific role of solvent, the degree of nonideality of solution, the influence of the solution components on the solubility of substrates (when the reaction is a heterophase one), and the effect of the reaction medium on metal complexes. In heterogeneous catalysis, the following factors should be considered: the composition and nature of the surface, the phase composition of the catalyst, the degree and the nature of the surface nonuniformity, the adsorption of reactants and products, and the transformations of the active sites under the influence of the reaction medium.

The group of hypotheses about elementary steps of the reaction mechanism permits the discrimination between the mechanisms at the level corresponding to the discrimination between the individual steps. Here it is also essential to have a set of hypotheses which is as complete as possible.

In the complex cases of stoichiometrically ambiguous reactions, the third group of hypotheses proved very useful — the hypotheses about the conjugation nodes in the kinetic scheme of a reaction.[13] The conjugation node is a fragment of the structure of the reaction mechanism incorporating the intermediate that is transformed by two or more steps. The form of the expression for the ratios between the rates of product formation via different routes having a conjugation node determines the structure of this node. When the reaction is far from equilibrium and the routes via which the products are formed include irreversible stages, the ratios between the product formation rates are far simpler functions of the reactant concentrations than the rates themselves or even the expressions for the selectivity. It is worthwhile analysing the conjugation nodes bearing in mind some hypothesis about the possible reaction mechanism (i.e., at the stage of designing the discriminating experiment). However, even if the general reaction mechanism is unknown, such analysis is all the same useful.[13] Examples of analysis of conjugation nodes were discussed in several papers.[13,26-28]

When a priori information about the process is available, it is useful to formulate, before putting forward the hypotheses, the requirements that should be satisfied for all the hypothetical mechanisms. Naturally, the greater the amount of a priori information, the smaller the number of hypotheses which should be tested.

In this review, we restrict the consideration to the formal aspect of the problem of the formulation of hypotheses and discuss the methods of computer-aided hypothesis generation.

Numerous publications have been devoted to the problems of computer-assisted synthesis design (CASD). Investigators — users of the CASD program — usually solve the following problems: 'which reactants must be used and which reactions must be carried out to prepare a given substance in the simplest way?' and 'what can be obtained from the specified set of substances?'.[29-33] By the most conservative estimate about 500 original papers and several tens of reviews have been devoted to the CASD problem. Here we shall refer only to those reviews that contain a fairly complete list of references.

The CASD methods are in many respects the same as the methods for the computer generation of hypotheses about reaction mechanisms. There is also the resemblance between the algorithms of the programs. Furthermore, problems associated with the complexity of the algorithms for combinatorial search inevitably arise in both cases. In order to avoid a combinatorial explosion, particular attention is devoted to the development of effective selection criteria, which make it possible to shorten the working time of a program. However, computer generation of reaction mechanisms also has its specific features.

Firstly, in contrast to the majority of CASD algorithms, the programs designed to formulate the mechanistic hypotheses deal with elementary reactions rather than synthetic ones. At any rate, the developers of the programs for mechanism generation devote much more attention to the problem of the reactions being elementary. Secondly, the central problem which then arises is not that of the way in which the multistage synthesis can be carried out but concerns the alternative variants of the behaviour which may be expected from the chemical system when it is 'left alone' (after specifying particular conditions). Thirdly, in the computer-assisted synthesis design, use is also made of the evaluation criteria, which are inapplicable to the reaction mechanisms. These include criteria associated with the choice of a 'synthesis strategy'[31,34-36] and with the estimation of the cost of the synthetic steps,[37] the criteria characterising the similarity between intermediates and target products,[29,38-54] the criteria employing the complexity and similarity information-theoretic indices for the description of the complexity of the synthesis,[55-57] etc. In certain programs, some of these criteria have been erroneously extended to the study of the mechanisms. Criteria such as the mechanistic complexity indices may be used, in principle, in the study of reaction mechanisms; however, they may not be used as evaluation criteria underlying the program and must be employed for discrimination between hypotheses if it is necessary to choose the simplest mechanism for the description of experimental observations. Otherwise the correct but more complex hypothesis may be neglected whereas sets of simpler but incorrect hypotheses would be subjected to experimental tests.

In the study of a reaction mechanism, the physicochemical criteria come to the fore. These are associated with the estimation of the chemical reactivity and are based, for example, on the calculations of electronegativity, polarisability,[58] bond energies,[59-64] thermodynamic functions,[65,66] and other physicochemical quantities. In addition, use is made of criteria based on heuristic rules such as Tolman's rules,[67-69] the restrictions the molecularity of the stages, and others. The choice of the evaluation criteria is a key factor in the development of programs for the study of reaction mechanisms, since they determine in many respects the plausibility of the results and the program execution time.

Yet another difference between the CASD programs and those for the generation of mechanisms is that the completeness of the combinatorially feasible hypotheses concerning the synthetic pathways is less important than the completeness of the hypotheses about the reaction mechanisms.

The computer programs which are being developed for the generation of reaction networks and for the study of the possible reaction routes can manipulate chemical information at a more or less general level. In some cases, the reaction is formulated in a compact form (only the transformation of the molecular fragment is described or it is simply indicated which bonds — and between which atoms — are formed or broken), whilst in others the reaction is specified in greater detail (not only the reacting fragment, but also all the reacting molecules or even the reaction conditions are specified). Depending on how detailed is the description of the reaction, one speaks of a greater or lesser degree of its generality.

Scheme 1

	H^1	H^2	H^3	C^4	O^5	H^6	C^7	O^8
H^1	0	0	0	1	0	0	0	0
H^2	0	0	0	1	0	0	0	0
H^3	0	0	0	1	0	0	0	0
C^4	1	1	1	0	1	0	0	0
O^5	0	0	0	1	4	1	0	0
H^6	0	0	0	0	1	0	0	0
C^7	0	0	0	0	0	0	2	2
O^8	0	0	0	0	0	0	2	4

B

+

	H^1	H^2	H^3	C^4	O^5	H^6	C^7	O^8
H^1	0	0	0	0	0	0	0	0
H^2	0	0	0	0	0	0	0	0
H^3	0	0	0	0	0	0	0	0
C^4	0	0	0	0	−1	0	1	0
O^5	0	0	0	−1	0	0	1	0
H^6	0	0	0	0	0	0	0	0
C^7	0	0	0	1	1	0	−2	0
O^8	0	0	0	0	0	0	0	0

R

=

	H^1	H^2	H^3	C^4	O^5	H^6	C^7	O^8
H^1	0	0	0	1	0	0	0	0
H^2	0	0	0	1	0	0	0	0
H^3	0	0	0	1	0	0	0	0
C^4	1	1	1	0	0	0	1	0
O^5	0	0	0	0	4	1	1	0
H^6	0	0	0	0	1	0	0	0
C^7	0	0	0	1	1	0	0	2
O^8	0	0	0	0	0	0	2	4

E

The developers of programs have encountered several typical problems. The generation of hypotheses always involves combinatorial searching. The algorithm frequently contains several modules designed for the solution of different problems associated with backtracking. One such module is usually designed for the solution of the problem of the generation of intermediates and another, for testing the intermediates for isomorphism (i.e., the program checks whether a given intermediate was obtained previously). In order to reduce the computation time, specific procedures, which depend on the type of the specific program, are employed.

In principle, there are two types of program: empirical programs and programs tending to be of the formal-logical type.[31,32] In the former, use is made of libraries of generalised reactions † (the degree of generality can differ greatly, but specific reactions are usually not considered, since with their aid one can obtain only trivial hypotheses which have already been described in the literature).

Generalised reactions play the role of operators, which by acting on a molecular fragment, give rise to a new fragment. Since the initial molecule is known, the entire reaction also becomes known. Thus, the program algorithm reduces to the exhaustive search for molecular fragments to which specified generalised reactions may be applied. Apart from all the possible limitations, which are specified by the user, the degree of generality also plays the role of an evaluation factor.

In programs of the second (formal-logical) type, the reaction is formulated on the basis of some logical rules (these may be stoichiometric relations, the search for all possible bond redistributions, etc.). Obviously, the programs of the second type are highly inter-active and require stronger constraints and evaluation criteria.

III. Programs for the formulation of hypotheses on the mechanisms of complex reactions

1. Programs of the formal-logical type

One of the most common methods for the development of programs for the study of chemical reaction mechanisms is that based on the Dugunji–Ugi mathematical model.[38] This method was used in a program package of the Munich project.[29]

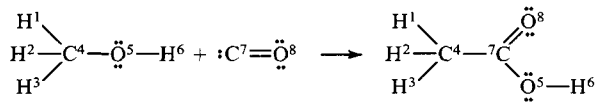
† By a generalised reaction, we understand a description of the chemical reaction lacking the particular details distinguishing each specific reaction from similar ones. Such details may be atoms or whole fragments which are not involved directly in the redistribution of electrons or in the formal redistribution of bonds. In generalised reactions, the designations of atoms may be replaced by symbols. The representation of a generalised reaction may look nothing like a reaction, which is unusual for chemists: matrices, graphs, special-format files, or their individual components may be used.

The Dugunji–Ugi model, also known as the *be*-matrix model, uses the matrix representations of chemical structures. Suppose a set of atoms (not necessarily different) $\mathscr{A} = \langle A_1, A_2, ..., A_n \rangle$ is specified. The set of the possible sets of molecules made up of these atoms may be regarded as isomeric ensembles of molecules EM. Each atom A_i, $A_i \in \mathscr{A}$, is included in each isomeric ensemble EM once and only once. The complete set of k isomeric ensembles of molecules forms a family of isomeric ensembles $\mathrm{FIEM}(\mathscr{A}) = \{\mathrm{EM}_1, \mathrm{EM}_2, ..., \mathrm{EM}_k\}$. Two isomeric ensembles of the same family are different (nonisomorphic), if they contain different sets of molecules or other species. The initial species and final products may be represented by structural *be*-matrices of the two isomeric ensembles (for the initial species and the products). The non-diagonal entries of the *be*-matrix reflect the connectivity of the atoms from the set $\mathscr{A}$ in the molecules belonging to the corresponding isomeric ensemble. The diagonal entries are equal to the number of free valence electrons of the corresponding atom. Suppose that the *be*-matrix ***B*** describes the initial species and the *be*-matrix ***E*** describes the products of some reaction: $\mathrm{EM}(\boldsymbol{B}) \longrightarrow \mathrm{EM}(\boldsymbol{E})$.‡ The bond redistribution can then be described by the fundamental equation of the Dugunji–Ugi model:

$$\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E},$$

where ***R*** is the matrix transforming ***B*** into ***E*** (the *r*-matrix).[38]

Consider as an example the methanol carbonylation reaction:



In this case, the set of atoms comprising the isomeric ensembles of FIEM may be described as $\langle H^1, H^2, H^3, C^4, O^5, H^6, C^7, O^8 \rangle$, while the equation of the Dugunji–Ugi model assumes the form presented in Scheme 1.

The matrix ***R*** can be represented as the sum of the *r*-matrices of the elementary steps of the mechanism:

$$\boldsymbol{R} = \boldsymbol{R}_1 + \boldsymbol{R}_2 + \ ... \ + \boldsymbol{R}_S$$

(S is the total number of elementary steps), while the entire mechanism can be represented as a sequence of transformations of the isomeric ensembles:

$$\mathrm{EM}(\boldsymbol{B}) \equiv \mathrm{EM}_0 \longrightarrow \mathrm{EM}_1 \longrightarrow \mathrm{EM}_2 \longrightarrow \ ... \ \longrightarrow \mathrm{EM}_S \equiv \mathrm{EM}(\boldsymbol{E})$$

‡ Hereafter we use the abbreviations and designations adopted in the English language literature: EM = ensemble of molecules; FIEM is the family of isomeric ensembles of molecules. The Latin symbols *be* and *r*, appearing in the designations of the matrices, arise from the words *bond-electron* matrix and *reaction* matrix. The symbols ***B*** and ***E*** arise from the words *beginning* and *end*.

For catalytic reactions, the set $\mathscr{A}$ must be supplemented by the atoms entering into the composition of the catalyst. For example, the reaction described above takes place in the presence of the Rh(I)–HI catalytic system.[70–72] The isomeric ensemble $EM_0 \equiv EM(\boldsymbol{B})$ must then include the HI and $[Rh(CO)_2I_2]^-$ species. A simplified mechanism of the methanol carbonylation reaction can be written as follows:[72]

$$\{CH_3OH + HI + CO + [Rh(CO)_2I_2]^-\} \xrightarrow{1}$$
$$EM_0$$
$$\longrightarrow \{CH_3I + H_2O + CO + [Rh(CO)_2I_2]^-\} \xrightarrow{2}$$
$$EM_1$$
$$\longrightarrow \{H_2O + CO + [(CH_3)Rh(CO)_2I_3]^-\} \xrightarrow{3}$$
$$EM_2$$
$$\longrightarrow \{H_2O + [CH_3C(O)Rh(CO)_2I_3]^-\} \xrightarrow{4}$$
$$EM_3$$
$$\longrightarrow \{H_2O + CH_3C(O)I + [Rh(CO)_2I_2]^-\} \xrightarrow{5}$$
$$EM_4$$
$$\longrightarrow \{HI + CH_3C(O)OH + [Rh(CO)_2I_2]^-\}.$$
$$EM_5$$

In this example, each of the five stages of the mechanism can be represented in a matrix form:

$$\boldsymbol{B}_i + \boldsymbol{R}_i = \boldsymbol{E}_i, \quad i = 1, 2, ..., 5,$$

where $\boldsymbol{B}_i = \boldsymbol{E}_{i-1}$ and the overall transformation can be expressed as follows

$$\boldsymbol{B}(EM_0) + \boldsymbol{R} = \boldsymbol{E}(EM_5).$$

Here $\boldsymbol{R} = \sum_i \boldsymbol{R}_i$.

The computer program RAIN was developed for the investigation of reaction mechanisms within the framework of the Munich project.[29,30,73–78] A distinctive feature of this program is the so called 'bilateral' search. The RAIN program inherited from CASD the idea of retrosynthesis,[79,80] in which the possiblity of searching for possible starting reactants by inspecting the products (retrosearch) is employed. In RAIN, the search is carried out simultaneously in the forward and reverse directions (bilateral search). Reaction generators, which make it possible to find both the reactions and retroreactions, are used in the RAIN program.[77] The reaction generators constitute, in fact, a compilation of logical rules controlling the process of formulation of a specific reaction (or retroreaction). These rules are based on the combinatorial enumeration of all possible bond redistributions. The enumeration is constrained by the user's settings. Transition tables should apparently be regarded as the principal constraints. These tables, compiled for atoms of each type, restrict the number of transformations (and retrotrans-formations) of the valence states of atoms in the course of the generation of elementary steps. No generalised reactions of any kind are thereby used. Therefore, the RAIN program can be categorised as program of the formal-logical type.

The idea of the bilateral search is applied as follows. Isomeric ensembles of the initial species [EM($\boldsymbol{B}$)] and final products [EM($\boldsymbol{E}$)] are specified. The reactions and retroreactions are sought for the ensembles EM($\boldsymbol{B}$) and EM($\boldsymbol{E}$), respectively. At some step, two (or more) intermediate isomorphic isomeric ensembles are obtained as a result of the search in different directions. In other words, the generation of reaction pathways in the direction of the products and in the direction of the initial species should lead to isomorphic isomeric ensembles. Methods have been developed for the optimisation of the search for isomeric ensembles of molecules connecting the search tree to the retrosearch tree[30] on the basis of the principle of the minimum chemical distance.[29,30,38–43] A chemical distance is regarded as the number of electrons which must be redistributed in order to convert one ensemble of molecules into another.

The action of this principle as an optimising factor is based on the following hypothesis. Let an ensemble of molecules be subjected to various transformations. The processes having smaller chemical distances between the reactants and products are more probable than those having greater distances. According to Ugi et al.[77] the permissible deviations from this rule are within two units of the chemical distance (one or two electrons); i.e., the reaction can sometimes choose a pathway not corresponding to the minimum structural changes, but these deviations from the general rule are not very large.

The combinatorial search in the RAIN program is limited by the user-defined constraints. In particular, apart from the transition tables, other formal constraints can be introduced: the maximum number of atoms entering into the reacting fragment of the molecule; the maximum change in bond order during the elementary reaction; the maximum number of bonds the order of which changes; the minimum size of the rings in the carbon skeletons of the molecules that contain triple or cumulated double bonds; the maximum number of charged atoms; the maximum number of covalent bonds between heteroatoms; the maximum number of rings; the maximum number of individual molecules in the intermediate ensembles of molecules, etc. (18 rules total).

The bilateral search has several drawbacks. In contrast to CASD, in the study of reaction mechanisms it is necessary to conduct the search in one (preferably forward) direction, since the intermediates formed during the reaction may interact with each other and may participate in reactions with the initial substances. It is then necessary to expand the set $\mathscr{A}$ at each step of the generation of the possible ensemble of molecules (and hence to increase the size of the ensemble). This is impossible within the bilateral search, which, within the framework of the Dugunji – Ugi model, requires the set $\mathscr{A}$ to be fixed and not to change during the execution of the program. According to Barone and Chanon,[32] this was first mentioned by Yoneda, but the paper containing a description of his GRACE program[81] does not deal with the problem of the possible expansion of the set $\mathscr{A}$ in the course of its execution.

Like RAIN, the GRACE program belongs to the formal-logical type, but it employs a somewhat different principle of reaction generation. Yoneda put forward the idea of reaction group matrices that describe the transformations of functional groups of the molecules. Upon imposing the constraints, the GRACE program identifies the functional groups within the molecules and generates the reaction group matrices by a combinatorial-logical procedure. Then, it detects all possible combinations of the reaction group matrices from which the elementary reaction matrices may be obtained (the r-matrices of steps). Thus, as in the RAIN program, the GRACE program employs neither the library of functional group transformations nor the library of elementary reactions, which are the combinations of the functional group transformations.

In the GRACE program, the evaluation criteria are formulated by the user according to the standard scheme. The following constraints are possible:

(1) the number of atoms (reaction centres) the valence state of which changes during the elementary reactions must not exceed 4;

(2) the number of broken bonds (or bonds the order of which is decreased) and formed bonds (or bonds the order of which is increased) should not be greater than 3;

(3) the molecularity of each reaction should not exceed 3;

(4) the number of unpaired electrons on one atom should not be greater than 1;

(5) the number of atoms with unpaired electrons in one molecule should not be greater than 4;

(6) the numbers of valence electrons at the carbon and hydrogen atoms should be 4 and 1 respectively;

(7) at least one active centre on the catalyst surface must participate in each step;

(8) the hydrogen atoms must be adsorbed on the catalyst surface and cannot exist as free radicals.

The program allows the user to change any particular constraints or not use some of them at all. Apart from these constraints, an upper limit to the complexity of the elementary transformation is predefined. The quantity calculated from the *r*-matrix and equal to half the chemical distance was adopted as a measure of complexity. A network of elementary reactions (the reaction network) is generated in the forward direction only, which makes it possible to expand the set $\mathscr{A}$ in the course of its growth when necessary. The execution of the program may be illustrated by the reaction network generated for the heterogeneous catalytic hydrogenation of ethene:[81]

$$CH_2=CH_2+2Z \longrightarrow ZCH_2-CH_2Z$$

$$H_2+2Z \longrightarrow 2HZ$$

$$ZCH_2-CH_2Z+HZ \longrightarrow ZCH_2-CH_3+2Z$$

$$ZCH_2-CH_3+HZ \longrightarrow CH_3-CH_3+2Z$$

$$CH_2=CH_2+2Z \longrightarrow ZCH=CH_2+HZ$$

$$ZCH=CH_2+2Z \longrightarrow ZCH=CHZ+HZ$$

$$2HZ \longrightarrow H_2+2Z$$

Z is an active site of a catalyst.

It is of interest to note that the same set of reactions, which are formulated as transformations of ensembles of molecules, is somewhat larger by virtue of the identity of certain transformations. For example, the following two transformations constitute essentially the same reaction.

$$\{CH_2=CH_2+4Z+H_2\} \longrightarrow \{ZCH=CH_2+HZ+2Z+H_2\}$$

$$\{CH_2=CH_2+2HZ+2Z\} \longrightarrow \{ZCH=CH_2+3HZ\}$$

In this sense, the Dugunji–Ugi model is somewhat redundant. Indeed, if the same reaction had not been duplicated, less time would be needed for the program run. When the number of elementary steps is large, this factor plays a significant role. There are seven reactions in the example presented above. If the reactants are represented by an ensemble of molecules, then 11 transformations of such ensembles are required (i.e., 1.5 times more). The time required for the interpretation of the results proportionally increases.

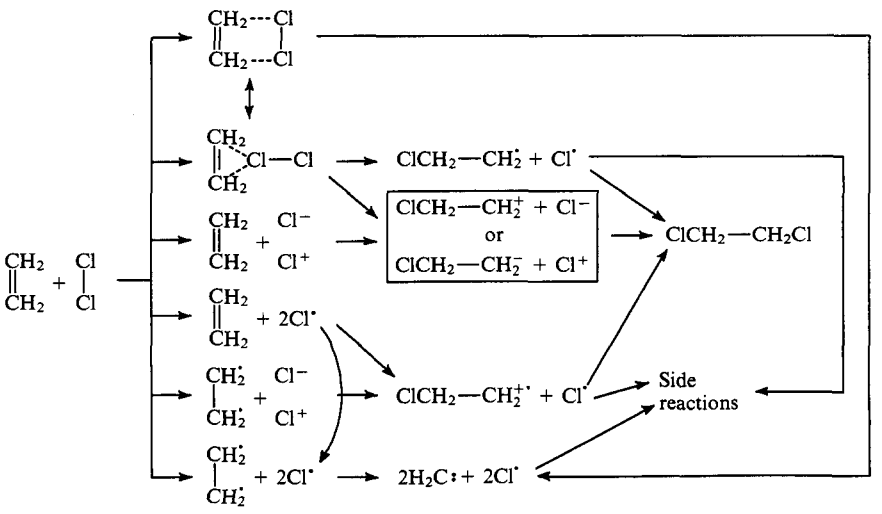
The Dugunji–Ugi model was also used in Dozmorov's program.[55] Dozmorov proposed a new similarity measure, which made it possible to diminish the number of steps and intermediates generated. This similarity measure is a topological information-theoretic index which involves information components of different nature.[82] In our opinion, the use of this characteristic as an evaluation criterion has no sufficient physical justification. Dozmorov does not report the details of the algorithm, but it appears that, as in the GRACE program, only the forward search is employed in his program, while the elementary reactions are generated by combinatorial-logical methods (i.e., it also belongs to the formal-logical type). The reaction network of the 'informationally' allowed steps of the ethene chlorination reaction, obtained with the aid of this program, is presented in Scheme 2.

Thus in all the programs of the formal-logical type, based on the Dugunji–Ugi model, similarity characteristics are used to reduce the number of 'false' solutions. In the RAIN program, the criterion is the chemical distance (of all possible transformations of the ensemble of molecules, the program selects only those corresponding to the minimum chemical distance to within an error of 2 units of the chemical distance.) In the GRACE program, the criterion is the complexity of the elementary reaction equal to half the chemical distance (the user specifies the upper limit to this quantity). In Dozmorov's program, the criterion is the information topological index, with the aid of which the program decides whether an elementary reaction is allowed or prohibited from the standpoint of imformation.

The MECHEM program developed by Valdés-Peréz[83–92] has appeared comparatively recently. Its algorithm is based on the stoichiometry principles. The initial version of the program operated not with the structural formulae of species, but with their compositions. Currently, the program also makes it possible to recognise the structures of substances. The MECHEM program belongs to the type of formal-logical programs because it does not employ any libraries of generalised reactions. The mechanisms are generated by the constrained combinatorial search for possible transformations of hypothetical intermediates. There are no reaction generators as such in the program. The 'skeleton' of the reaction mechanism is formed with further specification of the elementary steps and species. The model for the formulation of hypotheses is based on the following principles:[85]

(1) The program searches for the simplest mechanisms (hypotheses) by a progressive increase in the number of steps and species. If the mechanisms with a smaller number of steps and species do not satisfy all the user-defined constraints or the condition of elemental balance, the number of species is increased or whilst the number of species is held constant the number of

Scheme 2



steps is increased. Thus the complexity of the hypotheses is characterised by the number of species and steps.

(2) The set of reaction mechanisms in which the unknown intermediates are designated by wild cards X, Y, Z, etc., is initially generated. These wild cards are then replaced by the formulae of specific substances. Their empirical formulae are evaluated from the elemental balance (in the latest version of MECHEM, an algorithm for the determination of the structures of species from their formulae is employed;[90] it is then possible to impose constraints on the species structures[93]).

(3) The molecularity and the number of products of an elementary reaction do not exceed 2; in the same step, the same species cannot play the role of both the starting material and the product.

(4) The mechanisms are generated in such a way that each step cannot have starting materials that were not generated in the preceding steps or specified by the user as starting materials. (This and several other principles constitute the basis of the canonical representation of mechanisms[91]).

The user inputs the starting materials and products, the initial form of the catalyst (if it exists), and the intermediates observed. The user also inputs the constraints on the valence of each atom, the order of the appearance of the intermediates in the steps, the possibility (or impossibility) of the initial species being among the products of particular steps, the number of bonds the order of which changes, etc. A large number of heuristics, which make the algorithm more effective, are used in the program.[89]

Possible mechanisms of ethane hydrogenolysis were generated with the MECHEM program.[88,89] The following constraints were used: (1) the overall stoichiometry is $C_2H_6+H_2 \longrightarrow 2CH_4$; (2) one catalyst active site M can form only one bond (not necessarily a single bond); (3) intermediates with a bridged structure, MCH_2-CH_2M and $MCH-CHM$, are present in the reaction; (4) hydrogen is not formed in any of the steps; (5) all intermediates contain at least one catalyst site M; (6) none of the intermediates contain three carbon or metal atoms; (7) no more than three bonds are formed and broken (regardless of their order) in one step; (8) the MCH_2-CH_2M intermediate must appear in the sequence of steps before the $MCH-CHM$ intermediate. Apart from these constraints, account was also taken of the need to maintain the elemental balance in each step and the condition that the valences of hydrogen and carbon atoms do not exceed 1 and 4 respectively. Among other mechanisms of the same complexity (i.e., with the same number of steps and species), the following mechanism of ethane hydrogenolysis was generated: §

1. $H_2+2M \longrightarrow 2U$
2. $2M+C_2H_6 \longrightarrow U+V$
3. $2M+V \longrightarrow MCH_2-CH_2M+U$
4. $M+MCH_2-CH_2M \longrightarrow U+W$
5. $2W \longrightarrow MCH_2-CH_2M+MCH-CHM$
6. $MCH-CHM \longrightarrow 2X$
7. $U+X \longrightarrow M+Y$
8. $2Y \longrightarrow X+Z$
9. $U+Z \longrightarrow 2M+CH_4$

Here U, V, W, X, Y, and Z are intermediates which were interpreted by the program as HM, CH_3-CH_2M, CH_2-CHM, MCH, CH_2M, and CH_3M respectively. The stoichiometric numbers of steps 1–9 are 1, 1, 1, 2, 1, 1, 4, 2 and 2, respectively, with which all the intermediates vanish in the overall equation.

§ Since the program generates only unimolecular and bimolecular reactions (this is one of the principles inherent in the algorithm for the generation of a step), it is seemingly implied that 2M represents two linked active sites of the heterogeneous catalyst (M_2).

The intermediates MCH and CH_2M may be interpreted as carbyne and carbene complexes (at the time when the above investigation was carried out, the algorithm for the identification of bond orders was introduced into the program).

Generally, the MECHEM program is designed to search for hypotheses invoked to account for the results of observations (i.e., the hypotheses follow the observations). As mentioned above, another formulation of the problem is methodologically more correct: What variants of observations are to be expected? We may note that the weakest feature of this program is the concept of the selection of simplest mechanisms. It is reasonable to employ simplicity criteria at the step involving the discrimination between hypotheses and also at the instant when the experimental tests of the hypotheses have already been completed and the remaining hypotheses explain equally well the observations and agree with the background knowledge about the processes of a given type. In this case, when the future design of experiments is unclear or if it is impossible to experimentally discriminate between the hypotheses by means of the available methods, it is useful to apply criteria characterising the simplicity of the mechanism. These may be criteria based on a simple count of the number of steps and species (as in MECHEM) or more complex criteria.[94–97]

Apart from the formal-logical methods of generating reaction networks discussed here, in many studies the reaction mechanisms have been generated by the joint application of combinatorial methods and stoichiometry,[98–100] as well as graph theory.[101] Computer programs were not created on the basis of these methods, nor they have been adequately documented in the literature. Methods have also been developed for 'cutting off' the mechanisms from the reaction networks that match a given overall stoichiometry.[9,98,102–112] In these methods, it is postulated that all starting materials, intermediates, and final products and sometimes the equations of the steps are known. For this reason, despite the usefulness of such algorithms and programs for the study of mechanisms, their employment for the generation of hypotheses about the mechanisms of chemical reactions is extremely limited.[14]

2. Empirical programs

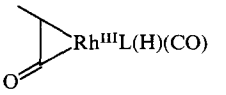
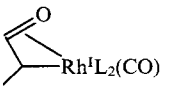
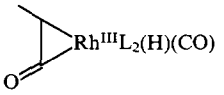
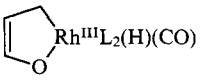
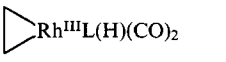
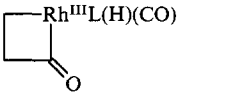
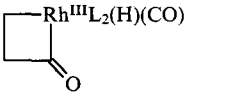
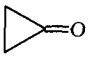
As stated above, all empirical programs are characterized by a general property: generalised reactions (transforms) are employed in them. The method used to describe the generalised reactions can be arbitrary. The general algorithm involves the recognition of molecular fragments to which particular generalised reactions (reaction templates) can be applied with subsequent generation of specific reactions and structures of substances.

One of the first empirical programs is the CAMEO program.[113] At the time of writing the present review, 25 communications have already been published. Here we refer only to the first publication[113] and to a review.[114]

The CAMEO program is designed to simulate the mechanisms of organic reactions. It is characteristic of the program that the search for each successive fragment to which a particular generalised reaction may be applied is performed in this program solely in the forward direction (side paths are then considered and account is taken of the interaction between the intermediates).

The starting materials and the reaction products, the reaction conditions, and also the list of modules which must be activated are initially specified. The program contains several modules, each of which is designed to describe a particular class of organic reactions. The molecules are represented in the form of special tables in which atoms and bonds are specified.[108] These tables contain the following information: the ordinal number of each atom, the type of each atom (C, H, ...), its coordinates on the graphical display, and the code of the charge (numbers 1, 2, and 3 stand for a cation, anion, or a radical, respectively; if the atom is neutral, then the code 0 is used); it is indicated between which atoms there are bonds and what are their orders. Each bond has a unique number. The location of the bond relative to the conventional plane of the figure is indicated. Furthermore, account is

Table 1. List of substances proposed by the TAMREAC program for the generation of the ethene hydroformylation reaction network.[68]

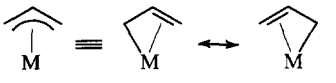
No. of species	Chemical formula	No. of species	Chemical formula
1	$Rh^{I}L_2(H)(CO)_2$		
2	$(\eta^2\text{-}C_2H_4)Rh^{I}L_2(H)(CO)_2$	21	
3	$Rh^{I}L(H)(CO)_2$		
4	$C_2H_5Rh^{I}L_2(CO)_2$	22	$[\eta^2\text{-}C(O)=CHCH_3]Rh^{I}L(H)(CO)$
5	$O=C(C_2H_5)-Rh^{I}L_2(CO)$	23	$O=CH-CH(CH_3)-Rh^{III}L_2H_2CO$
6	$O=C(C_2H_5)-Rh^{III}L_2H_2(CO)$		
7	$Rh^{I}L_2(H)(CO)$	24	
8	$CH_3CH=Rh^{I}L_2(H)(CO)_2$		
9		25	$CH_3CH=CH-O-Rh^{I}L_2(CO)$
		26	
10	$[\eta^2\text{-}C(O)=CHCH_3]-Rh^{I}L_2(H)(CO)$		
11	$O=CH-CH(CH_3)-Rh^{I}L_2CO$	27	H_2
		28	$HO-CH=CHCH_2-Rh^{I}L_2(CO)$
12		29	$HO-CH=CHCH_2-Rh^{III}L_2H_2(CO)$
		30	$CH_3CH=CH-O-Rh^{III}L_2H_2(CO)$
13		31	
		32	$[\eta^2\text{-}CH_2=CH-C(O)]Rh^{I}L_2(H)(CO)$
14		33	$CH_3CH=CH-OH$
		34	$O=CHCH_2CH_2-Rh^{I}L_2(CO)$
		35	$O=CHCH_2CH_2-Rh^{III}L_2H_2(CO)$
15	L		
16	$(\eta^2\text{-}C_2H_4)Rh^{I}L(H)(CO)_2$	36	
17	$C_2H_5Rh^{I}L(CO)_2$	37	C_2H_4
18	$C_2H_5Rh^{III}L(H)_2(CO)_2$	38	C_2H_6
19	CO	39	$CH_2=CH-CH=O$
20	$CH_3CH=Rh^{I}L(H)(CO)_2$	40	C_2H_5CHO

Note. The initial species were $Rh^{I}L_2(H)(CO)_2$ (L = PPh_3), C_2H_2, CO, and H_2.

taken of the evaluation rules, including the rules based on physicochemical principles (estimates of pK_a, ΔH_f, the HOMO–LUMO energies, heats of reaction, etc.).

The following modules were designed: reactions catalysed by bases and acids (reactions involving nucleophilic and electrophilic species) pericyclic reactions, redox reactions, free-radical reactions and chain processes, reactions of carbenoid species, reactions involving reduction by hydrides, and 'ene' and 'retro-ene' syntheses. With the aid of the CAMEO program, it is possible to determine the most probable sites of attack by active species and to estimate the comparative reactivities.

The TAMREAC program[67–69] was designed for investigators in the field of organometallic catalysis. Unfortunately, the authors[67–69] did not publish the details of the algorithm and it is therefore difficult to analyse the real possibilities of the program. Connectivity tables in which each atom is characterised by a number and a symbol (carbon, hydrogen, metal, arbitrary atom, arbitrary electrophilic atom, etc.) are used for the computer representation of the molecules within the framework of this program. In the tables, it is also indicated with which atoms and by what bonds the given element is linked and what is its overall charge (oxidation state). Furthermore, tables of η^2-bonds have been introduced for the description of the η^2-bonds with metals, in which it is indicated between which atoms there is a multiple bond and to which atom it is coordinated. The η^3-coordination is described with the aid of two isomeric structures:



The TAMREAC program provides for the possibility of generating elementary steps in each of which there is not more than one metal atom. The developers of the TAMREAC program systematically analysed a large number of elementary concerted reactions and constructed a well-organised library of generalised reactions. It includes the following reactions: (1) oxidation of the metal, (2) reduction of the metal, (3) reactions affording occupation of a coordination site, (4) reactions creating a vacant coordination site, (5) reactions without change in the oxidation state or the coordination number of the metal, and (6) radical reactions and one-electron transfers. The library consists of files each containing several generalised reactions, described in the form of a structural transformation of the reaction fragment.

The TAMREAC program contains an algorithm for the recognition of ligands, the number of electrons they can donate

Table 2. The reaction network for the hydroformylation of ethene generated by the TAMREAC program.[68]

No.	Reaction[a]	No.	Reaction[a]	No.	Reaction[a]	No.	Reaction[a]
1	1 → 3+15	23	7+39 → 32	45	13 → 12	67	23 → 7+40
2	1 → 7+19	24	7+40 → 6	46	13+15 → 14	65	23 → 11+27
3	1+37 → 2	25	7+40 → 23	47	14 → 7+36	69	24 → 11
4	2 → 1+37	26	8 → 4	48	14 → 13+15	70	24 → 25
5	2 → 4	27	8 → 9	49	15+16 → 2	71	25+27 → 30
6	2 → 16+15	28	8 → 20+15	50	16 → 3+37	72	25 → 24
7	3+15 → 1	29	9 → 10	51	16 → 9	73	25 → 26
8	3+37 → 16	30	9 → 11	52	16 → 12	74	26 → 7+31
9	4 → 2	31	9 → 16	53	16 → 20	75	26 → 25
10	4 → 5	32	9 → 21+15	54	17+15 → 4	76	26 → 28
11	4 → 8	33	10 → 5	55	17 → 20	77	28 → 26
12	4 → 17+15	34	10 → 9	56	17+27 → 18	78	28+27 → 29
13	5 → 4	35	10 → 11	57	18 → 3+38	79	29 → 7+33
14	5 → 10	36	10 → 22+15	58	18 → 17+27	80	29 → 28+27
15	5+27 → 6	37	11 → 9	59	20+15 → 8	81	30 → 7+33
16	6 → 5+27	38	11 → 10	60	20 → 16	82	30 → 25+27
17	6 → 7+40	39	11 → 24	61	20 → 17	83	32 → 7+39
18	7+19 → 1	40	11 → 32	62	20 → 21	84	32 → 11
19	7+31 → 26	41	11+27 → 23	63	21+15 → 9	85	32 → 34
20	7+33 → 30	42	12 → 9	64	21 → 22	86	34+27 → 35
21	7+33 → 29	43	12 → 13	65	22+15 → 10	87	34 → 32
22	7+36 → 14	44	12 → 16	66	22 → 21	88	35 → 7+40
						89	35 → 34+27

[a] The numbers of species are indicated in Table 1.

and the number of sites they can occupy. The number of valence electrons of the metal atom and its oxidation states are calculated from these data. The information obtained is used in the evaluation criteria. For example, it is ensured automatically that the number of valence electrons is within the limits specified by the user. The program also contains an algorithm for the recognition of the geometry of each complex. The developers of TAMREAC created a system of evaluation rules on the basis of which an expert assessment of the reaction plausibility is made using the data on the type of reaction (reductive elimination, migratory insertion, metallocycle formation, etc.), the coordination number of the metal, the number of valence electrons of the metal atom, and the geometry of the complex. This information, formalised in the form of evaluation matrices, is usually very useful to the investigator. However, one should understand that the data on the basis of which a conclusion is made are not sufficient. Therefore, such an estimate can be taken into account, but the possibilities of its employment within the framework of any specific study are extremely limited. Account must also be taken of the fact that the evaluation rules indicated above are usually developed exclusively on the basis of experimental observations which often lack a reliable theoretical justification. These rules are continuously revised and, generally speaking, they lack 'the strength of a law'. By studying a specific system, each investigator has ultimately the right to solve independently the problem of the need to take into account the expert recommendations. For this reason, the most acceptable variant is where the user (the investigator) is himself able to formulate the evaluation rules in accordance with various schemes proposed by the program.

The TAMREAC program was tested on a large number of specific reactions, and in a number of instances extremely interesting results were obtained. For example, this program was used to generate the reaction network for the ethene hydroformylation in the system containing $HRh(CO)_2L_2$ (L = PPh_3).[68] The starting materials were ethene, H_2, CO, and $HRh(CO)_2L_2$. It is seen from the results presented in Tables 1 and 2 that the program 'synthesised' fairly interesting by products, the search for which in a real catalytic system will help to confirm or reject some of the proposed pathways.

The KOMSIKAT program package for the investigation of the mechanisms of catalytic reactions has been developed at the Institute of Catalysis of the Siberian Division of the Russian Academy of Sciences on the basis of the Dugunji–Ugi model and the group matrix method (similar to that employed in the GRACE program[81]).[59–64] In general, the ideas underlying the KOMSIKAT program are similar to those implemented in the GRACE program, but it does not use all the combinatorially possible group matrices, only those specified by the user. Therefore, this program should be categorised as an empirical program. The authors of the KOMSIKAT project extracted from literature data the characteristic structural types of elementary transformations involving the catalyst. In particular, the authors postulated that the processes occurring in the CO + H_2 system with participation of a metal M may be described by five basic types of elementary transformations. Table 3 presents the reaction fragments and the corresponding blocks of *r*-matrices (index group matrices) employed. The diagonal elements of the index group matrices characterise the changes in the number of valence electrons.

In the KOMSIKAT program, the energy criterion was used in the evaluation of reaction pathways. The reaction network was generated as follows. In the first stage, all possible ensembles of

Table 3. The reaction fragments and the index group matrices used in the KOMSIKAT program.[64]

Bond redistribution scheme	Index group matrix
M+X−Y ⟶ X−M−Y	

	M	X	Y
M	−2	1	1
X	1	0	−1
Y	1	−1	0

X−M−Y ⟶ M+X−Y

	M	X	Y
M	2	−1	−1
X	−1	0	1
Y	−1	1	0

M−X−Y ⟶ Y−M−X

	M	X	Y
M	−1	0	1
X	0	1	−1
Y	1	−1	0

Y−M−X ⟶ M−X−Y

	M	X	Y
M	−1	0	1
X	0	1	−1
Y	1	−1	0

M−Z+X−Y ⟶ M−X+Z−Y

	M	X	Y	Z
M	0	−1	1	0
X	−1	0	0	1
Y	1	0	0	−1
Z	0	1	−1	0

molecules which may be obtained on the basis of the possible elementary transformations of the initial ensemble of molecules were generated. From the ensemble of molecules obtained in the first stage, only those were selected which corresponded to the maximum atomisation energy (i.e., to the most stable ensembles of molecules)¶ to within some accuracy, while the remaining ones were rejected. For each intermediate ensemble of molecules, remaining after the selection, a new set of ensembles of molecules, obtained by means of allowed elementary trans-formations, was found. Among them the most stable were also selected. The generation of a 'branch' of the reaction network was terminated in accordance with formal characteristics — after the isolation of the metal-catalyst in a free form (when the ensemble of molecules contained a free metal atom). As a result, the reaction network consisted of intermediate ensembles of molecules, which corresponded to the pathways leading to the transformation of their precursors into the most stable compounds. The atomisation energy was calculated by the method of interacting bonds[60] using semiempirical parameters of the bond energies of the atoms comprising the molecules in the ensemble. For bonds involving C, H, and O atoms, the parameters obtained from the experimental thermochemical data on the heats of formation of the corresponding compounds were used. Thus, the parameter $E_{C=O}$ was found from the heats of formation of aldehydes and ketones, the parameter E_{C-H} was obtained from the heats of formation of hydrocarbons and alcohols, the parameter E_{O-H} was found from the heats of formation of alcohols, etc.

At the next stage, the optimum catalyst was selected. For this purpose, the parameters E_{M-X} (E_{M-C}, E_{M-H}, E_{M-O}) characterising the metal were varied for each intermediate ensemble and each pathway in the reaction network (in each linear sequence of ensembles of molecules from the starting EM to the final one). Because the variation of the E_{M-X} parameter was organised as step-by-step procedure, the atomisation energy of each intermediate ensemble of each pathway was split into a number of values. The atomisation energy corresponding to the condition of a uniform distribution of energy over the elementary steps of the catalytic process was chosen for each ensemble. In other words, the most smoothed broken line, which corresponds to the atomisation energies of the intermediate ensembles of a the pathway, was chosen (Fig. 1). The semiempirical parameters of the bond energies calculated for this broken line were compared to the tabulated values for specific metals. The metal was regarded as optimal when the tabulated parameters of the bond energies for this metal was closest to the calculated values (i.e., values obtained for the most smoothed broken line).

Thus, it follows from a comparison of the calculated parameters of the formation of ethylene glycol in the reaction between CO and H_2 in the presence of a metal catalyst and the tabulated values that the best catalysts for this reaction are Rh, Co, Ir, Ru, and V. The ranges of the optimum E_{M-X} parameters for the formation of each product of the hydrogenation of CO (methane, methanol, ethane, ethanol, acetic acid, methyl formate, and ethylene glycol) were found and the corresponding metals were determined. The optimum parameters E_{M-X} and metals (Mn, Fe, Co, Ni, Ru, and Rh) for the formation of all the C_1 and C_2 products were found from the overlap of these ranges of parameters.

Baltanas and Froment[115] developed a program for the generation of reaction networks for the solution of specific

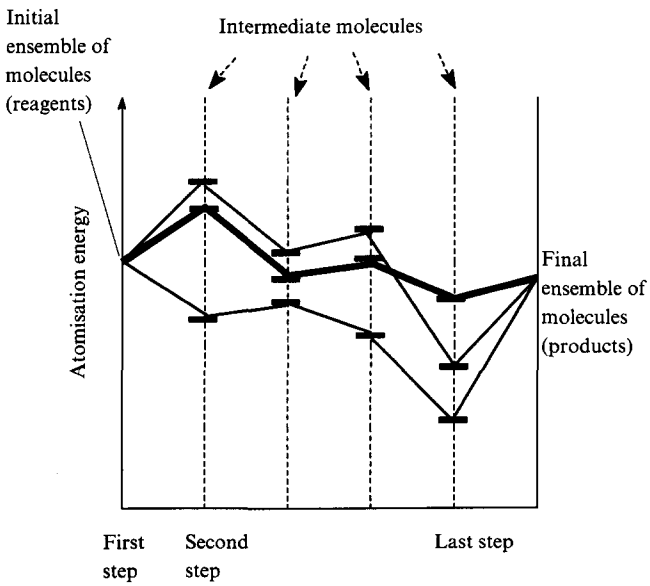


Figure 1. The splitting of the levels upon the variation of the E_{M-X} parameters. The optimum parameters E_{M-X} are those which correspond to the most 'smoothed' broken line, shown in the figure by a heavy line. The split levels of the atomisation energies of ensemble of molecules are shown in the dashed line corresponding to elementary steps.

¶ The ensemble of molecules for which the sum of the atomisation energies of the species was greatest was regarded as the most stable. Several most stable ensembles of molecules are possible because the metal atom is not known *a priori* and only the range of the possible atomisation energies, obtained by varying the semiempirical parameters of the bond energies, is known.[60]

problems, such as the study of catalytic hydroisomerisation and hydrocracking. Five generalised reactions, which are to be applied in succession to various species, are used in this program. These generalised reactions are designed solely for the investigation of the above catalytic processes. Species and reactions are described in terms of Boolean matrices which do not take into account the bond orders. The kinetic equations for the reaction network obtained were derived by the program. The principal disadvantage of this program is the narrow range of its applications.

Temkin's group (Lomonosov State Academy of Fine Chemical Technology, Moscow) are developing the ChemNet program for the formulation of hypotheses about the mechanisms of catalytic reactions. This program has been discussed in only a few publications, some of which are not readily available.[14,116] We shall, therefore, discuss it in more detail.

The first version of the program was written in the mid 1980s in FORTRAN IV, a language which has now become hopelessly obsolete.[117] A new version, designed for modern personal computers of the IBM PC type and using the principal ideas of the old program, has been developed recently.

The list of species participating in the reaction as the reactants and system components and the list of the elementary transformations (generalised reactions) are input data. The species are specified by the user in the form of structural formulae and are transformed into records having a special format. Part of the record contains an upper triangular adjacency matrix of the atoms, while another part comprises different characteristics of atoms (charges, oxidation states, etc.), bond orders, as well as multicentre bonds. The elementary transformations are analogues of chemical reactions and describe the characteristic features of the changes occurring. The extent to which the details of the reaction are specified is decided by the user.

The more detailed the description of the elementary transformation, the shorter the list of substances to which this transformation is applicable. For example, the two elementary transformations

$$M-C+C\equiv O \longrightarrow M-C(O)-C,$$

$$M-C\equiv C+C\equiv O \longrightarrow M-C(O)-C\equiv C,$$

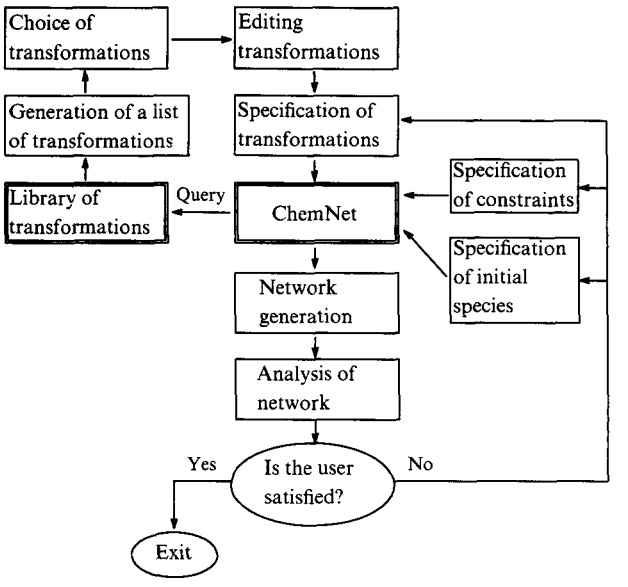


Figure 2. The bookkeeping scheme for the interactive execution of the ChemNet program.

describe the same reaction with specification of details to different extents. The difference consists only in that, in the first case, the program 'inserts' CO into all species containing the M–C bond, whilst, in the second case, it inserts it into metal alkynyl complexes. Note that, in the first case, the program will 'insert' CO also into M–C(O) fragments until the constraints imposed on the number of atoms in the species will begin to work. The elementary transformations can be inputted both in a symbolic form and in the form when standard chemical symbols of atoms are used. If the elementary transformation is formulated in a

Table 4. List of substances proposed by the ChemNet program for the generation of the reaction network using ten elementary transformations and five initial substances ($PdCl_2$, CuCl, MeOH, HC≡CMe, CO).

No. of species	Chemical formula	No. of species	Chemical formula	No. of species	Chemical formula
1[a]	$PdCl_2$	17	MeC≡CPdCOOMe	33	MeOPdCH=C(Me)COOMe
2[a]	CuCl	18	$Pd(C\equiv CMe)_2$	34	MeOPdC(Me)=CHCOOMe
3[a]	MeOH	19	ClPdCOC≡CMe	35	ClPdCH=C(Me)COC≡CMe
4[a]	HC≡CMe	20	MeOPdC≡CMe	36	ClPdC(Me)=CHCOC≡CMe
5[a]	CO	21	MeC≡CC(O)PdCOOMe	37	ClPdH
6	MeOPdCl	22	MeC≡CPdCOC≡CMe	38	MeOPdH
7	HCl	23	$Pd(COC\equiv CMe)_2$	39	MeOC(O)OMe
8	MeOCu	24	MeOC(O)–C(O)OMe	40	HPdCOOMe
9	$(MeO)_2Pd$	25	Pd^0	41	HPdC≡CMe
10	ClPdCOOMe	26	MeC≡CCOOMe	42	HPdC(O)C≡CMe
11	CuCOOMe	27	MeC≡C–C≡CMe	43	PdH_2
12	MeOPdCOOMe	28	MeOC(O)C(O)C≡CMe	44	HPdCH=C(Me)COOMe
13	$Pd(COOMe)_2$	29	MeC≡CC(O)C≡CMe	45	HPdC(Me)=CHCOOMe
14	CuC≡CMe	30	MeC≡CC(O)C(O)C≡CMe	46	HPdCH=C(Me)COC≡CMe
15	ClPdC≡CMe	31	ClPdCH=C(Me)COOMe	47	HPdC(Me)=CHCOC≡CMe
16	MeOPdC≡CMe	32	ClPdC(Me)=C(H)COOMe		

[a] Initial species.

Table 5. The reaction network generated by the ChemNet program on the basis of specified ten elementary transformations (ET) and five initial substances ($PdCl_2$, CuCl, MeOH, HC≡CMe, CO).

ET	Reaction[a]	ET	Reaction[a]	ET	Reaction[a]	ET	Reaction[a]
1	3+1 → 6+7	5	22+5 → 23	9	10+3 → 37+39	10	10+3 → 7+12
1	3+2 → 7+8	5	13 → 24+25	9	12+3 → 38+39	1	15+3 → 7+16
1	6+3 → 7+9	6	17 → 25+26	9	13+3 → 39+40	1	19+3 → 7+20
1	6+5 → 10	6	18 → 25+27	9	17+3 → 39+41	1	31+3 → 7+33
2	8+5 → 11	6	21 → 25+28	9	19+3 → 26+37	1	32+3 → 7+34
2	9+5 → 12	6	22 → 25+29	9	20+3 → 26+38	1	37+3 → 7+38
2	12+5 → 13	6	23 → 25+30	9	21+3 → 26+40	1	16+5 → 17
2	4+2 → 7+14	6	10+4 → 31	9	21+3 → 39+42	2	20+5 → 21
3	14+1 → 2+15	7	10+4 → 32	9	22+3 → 26+41	2	38+5 → 40
4	14+6 → 2+16	7	12+4 → 33	9	23+3 → 26+42	2	37+14 → 2+41
4	14+10 → 2+17	7	12+4 → 34	9	40+3 → 39+43	4	41+5 → 42
4	15+14 → 2+18	7	19+4 → 35	9	42+3 → 26+43	5	40+4 → 44
4	15+5 → 19	7	19+4 → 36	9	11+10 → 2+13	7	40+4 → 45
5	16+5 → 20	7	32 → 26+37	10	15+11 → 2+17	7	42+4 → 46
5	17+5 → 21	8	34 → 26+38	10	19+11 → 2+21	7	42+4 → 47
5	18+5 → 22	8	36 → 29+37	10	19+14 → 2+22	7	45 → 26+43
						8	47 → 29+43

[a] The numbers of the species are indicated in Table 1.

symbolic form, it is necessary to determine also the possible substituent of each symbol. For example, the generalised reaction

$$M+H-X \longrightarrow H-M-X,$$

is formulated in a symbolic form. The symbols M and X must be interpreted, indicating which atoms of the metal M and which atoms X can occur in the reaction.

Upon carrying out a 'rough' generation of the reaction network, the user can alter the degree of specification of detail in the records of steps, thereby excluding the directions of attack by particular species known a priori to be implausible.

A special unique format has been developed for the description of species and elementary transformations.[118] The substances and reactions are stored in the computer memory in the form of special codes, the description of which is outside the scope of the present review.

Several evaluation rules are used in the ChemNet program, which are recommended to the user.

1. In an elementary step, the change in the number of valence electrons and in the oxidation state of the metal atom must not exceed 2 in absolute magnitude, which agrees with Tolman's rules.[119]

2. The molecularity of the elementary reactions must not exceed 2.

3. The user specifies the maximum number of valence electrons, the maximum coordination number, and the range of the possible oxidation states for each metal atom (the quantities stored in a special file are used as default settings).

4. The user specifies the maximum number of atoms and/or the maximum numbers of atoms of a particular type (C, H, O, ...) in the products formed.

The generation of the reaction network is also constrained by the transforms which the user specifies himself or selects from the library. Furthermore, the generation of the reaction network is implicitly constrained by the degree of specification of details in the records of the transform.

After the specification of the initial species and elementary transformations (transforms), the program searchs for the initial species which are regarded as candidates for participation in each of the elementary transformations. After finding for each elementary transformation all the combinations of species to which the transformation is applicable and the new species have been stored in the computer memory, the program adds all the new substances to the list of the initial substances and repeats the entire procedure. The combinations of species and steps, tested at the preceding steps, are not examined. The process completes when each species (the starting material or the species obtained in the course of the network generation) was examined with respect to the possibility of its involvement in each elementary transformation, while new substances are not produced owing to the constraints imposed. The overall algorithm for the execution of the program can be represented by the bookkeeping scheme shown in Fig. 2.

One of the tests for the program was the synthesis of an alkynylcarboxylic acid ester in the $CuCl_2-CuCl-PdCl_2-R'OH$ system.

$$RC{\equiv}CH + CO + R'OH + 2CuCl_2 \xrightarrow{PdCl_2} RC{\equiv}CCOOR' + 2CuCl + 2HCl$$

The following elementary transformations were specified:

1. $MCl+ROH \longrightarrow MOR+HCl$ (M = Pd, Cu; R = C)
2. $MOR+CO \longrightarrow MCOOR$ (M = Pd, Cu; R = C)
3. $MCl+RC{\equiv}CH \longrightarrow MC{\equiv}CR+HCl$ (M = Cu; R = C)
4. $CuC{\equiv}CR+Pd-Cl \longrightarrow CuCl+PdC{\equiv}CR$ (R = C)
5. $PdC{\equiv}CR+CO \longrightarrow Pd-C(O)-C{\equiv}CR$ (R = C)
6. $C-M-C \longrightarrow M+C-C$ (M = Pd)
7. $M-C(O)-+C{\equiv}C \longrightarrow M-C{=}C-C(O)-$ (M = Pd)
8. $M-C{=}CH \longrightarrow C{\equiv}C+M-H$ (M = Pd)
9. $Pd-C(O)-+ROH \longrightarrow PdH+RO-C(O)-$ (R = C)
10. $CuC+C-Pd-X \longrightarrow CuX+C-Pd-C$ (X = Cl)

The following constraints were imposed: the maximum number of atoms in the conjectured species must be less than 20, the maximum oxidation state of the metal in the conjectured species is +1 for copper and +2 for palladium, and the maximum coordination numbers of copper and palladium are, respectively, 1 and 2.

The results of the generation of the reaction network are presented in Tables 4 and 5. These results may be used by the experimentalist in the discriminating experimental design.

IV. Conclusion

In conclusion, we may note that there are many computer programs, which make it possible to formulate hypotheses about the mechanisms of complex reactions and to generate reaction networks. Analysis of the programs considered makes it possible to indicate several typical disadvantages of these programs.

1. The impossibility of obtaining complete data on the transformations of the intermediates, side reactions, and by products because of the necessity for a rigid specification of the number of atoms entering into the isomeric ensemble and because of the bilateral nature of the search (RAIN, Dozmorov's program, GRACE).

2. The use of the simplicity criterion as a selection factor (MECHEM).

3. The narrowness of the range of applications (the program of Froment and Baltanas).

A feature common to the majority of the programs considered is that the key factor in the development of methods for the generation of hypotheses is the application of heuristic rules, which are used to different extents and at different stages of the execution of the programs.

1. The use of libraries of transformations (reaction generators) obtained as a result of the generalisation of the available literature data on the reactions (TAMREAC, KOMSIKAT, ChemNet, etc.).

2. The use of the more or less formalised rules for the assessment of the reaction plausibility (TAMREAC, CAMEO).

3. The use of heuristic rules for the optimisation, acceleration, and restriction of the search (RAIN, MECHEM).

4. The estimation of the thermodynamic and other characteristics, obtained by semiempirical or fully empirical methods, which in turn make it possible to assess the plausibility of individual steps or even entire reaction pathways (KOMSIKAT, the program of Froment and Baltanas, CAMEO).

In all cases, it is important for the user to know which heuristic rules are inherent in the program, what are the possibilities for their modification, and whether or not any of them can be abandoned. It is desirable to have such a possibility as regards the constraints as well.

Different levels of generalisation of the elementary reactions are employed in the empirical programs. In some programs, for example, in the ChemNet program, it is possible to apply simultaneously several levels of generalisation (i.e., reactions in which only the transformation of the fragments is characterised and reactions formulated in more detail).

Since different programs are based on different methods and heuristic rules, it is sometimes useful to combine them within the framework of a single program package. In particular, the present authors are working on the creation of a program package combining the ChemNet and KOMSIKAT programs. In our view, such a combination of programs may ensure an additional flexibility in the selection of substances for the formulation of hypotheses.

References

1. K R Popper *The Logic of Scientific Discovery* (London: Hutchinson, 1959)
2. K R Popper *Objective Knowledge* (Oxford: Oxford University Press, 1972)
3. K R Popper *Conjectures and Refutations: The Growth of Scientific Knowledge* (New York: Harper and Row, 1968)
4. L B Bazhenov, L Kh Samorodnitskii *Vopr. Filosofii* **6** 93 (1971)
5. D Hodson *Educ. Chem.* **19** 112 (1982)
6. V V Nalimov *Zavod. Lab.* **44** 325 (1978)
7. L S Polak, V Ya Gol'denberg, A A Levitskii *Vychislitel'nye Metody v Khimicheskoi Kinetike* (Computational Methods in Chemical Kinetics) (Moscow: Nauka, 1988)
8. V N Pisarenko, A G Pogorelov *Planirovanie Kineticheskikh Issledovanii* (Design of Kinetic Research) (Moscow: Nauka, 1969)
9. V G Gorskii *Planirovanie Kineticheskikh Eksperimentov* (Kinetic Experiment Design) (Moscow: Nauka, 1984)
10. *The ACS Style Guide: A Manual for Authors and Editors* (Ed. V S Dodd) (Washington, DC: American Chemical Society, 1986)
11. G Pólya *Mathematics and Plausible Reasoning* (Princeton: Princeton University Press, 1974)
12. O N Temkin, S M Brailovskii, L G Bruk, in *Khimicheskaya Kinetika v Katalize: Kineticheskie Modeli Zhidkofaznykh Reaktsii* (Chemical Kinetics in Catalysis: Kinetic Models of Liquid-Phase Reactions) (Ed. S L Kiperman) (Chernogolovka: Inst. Chem. Phys. Acad. Sci. of the USSR, Inst. Org. Chem. Acad. Sci. of the USSR, 1985) p. 59
13. O N Temkin, L G Bruk, D Bonchev *Teor. Eksp. Khim.* **24** 282 (1988)
14. O N Temkin, L G Bruk, A V Zeigarnik *Kinet. Katal.* **34** 445 (1993)
15. A T Balaban (Ed.) *Chemical Applications of Graph Theory* (London: Academic Press, 1976)
16. R B King (Ed.) *Chemical Applications of Topology and Graph Theory* (Amsterdam: Elsevier, 1983)
17. N S Zefirov, S I Kuchanov (Eds) *Primenenie Teorii Grafov v Khimii* (Chemical Applications of Graph Theory) (Novosibirsk: Nauka, 1988)
18. N Trinajstić *Chemical Graph Theory* (Boca Raton, FL: CRC Press, 1992)
19. D Bonchev, D H Rouvray (Eds) *Mathematical Chemistry: Graph Theory and Its Applications to Chemistry* Vols 1, 2 (New York: Gordon and Breach, 1991)
20. D H Rouvray (Ed.) *Computational Chemical Graph Theory* (New York: Nova, 1990)
21. A V Zeigarnik, O N Temkin *Kinet. Katal.* **35** 691 (1994)
22. O N Temkin, A V Zeigarnik, D G Bonchev *J. Chem. Inf. Comput. Sci.* **37** 429 (1995)
23. A T Balaban, in *Graph Theoretical Approaches to Chemical Reactivity* (Eds D Bonchev, O Mekenyan) (Dordrecht: Kluwer, 1994) p. 137
24. O N Temkin, G K Shestakov, Yu A Treger *Atsetilen: Khimiya. Mekhanizmy Reaktsii. Tekhnologiya* (Acetylene: Chemistry. Reaction Mechanisms. Technology) (Moscow: Khimiya, 1991)
25. T T Zung, L G Bruk, O N Temkin *Izv. Akad. Nauk, Ser. Khim.* 1806 (1993)
26. T J Katz, J McGinnis *J. Am. Chem. Soc.* **99** 1903 (1977)
27. S M Brailovskii, O N Temkin, A S Kostyushin, K Yu Odintsov *Kinet. Katal.* **31** 1371 (1990)
28. L V Shchel'tsyn, S M Brailovskii, O N Temkin *Kinet. Katal.* **31** 1361 (1990)
29. I Ugi, J Bauer, K Bley, A Dengler, A Dietz, E Fontain, B Gruber, R Herges, M Knauer, K Reitsam, N Stein *Angew. Chem., Int. Ed. Engl.* **32** 201 (1993)
30. I Ugi, J Bauer, C Blomberg, J Brandt, A Dietz, E Fontain, B Gruber, A v Scholley-Pfab, A Senff, N Stein *J. Chem. Inf. Comput. Sci.* **34** 3 (1994)
31. N S Zefirov, E V Gordeeva *Usp. Khim.* **56** 1753 (1987) [*Russ. Chem. Rev.* **56** 1002 (1987)]
32. R Barone, M Chanon, in *Computer Aids to Chemistry* (Eds G Vernin, M Chanon) (Chichester: Ellis Horwood, 1986) p. 19
33. M Bersohn, A Esack *Chem. Rev.* **76** 269 (1976)
34. E J Corey, W T Wipke, R D Crammer, W J Howe *J. Am. Chem. Soc.* **94** 431 (1972)
35. W T Wipke, H Braun, G Smith, F Choplin, W Sieber *ACS Symp. Ser.* **61** 97 (1977)
36. E J Corey, A K Long, S D Rubenstein *Science* **228** 408 (1985)
37. M Bersohn *Bull. Chem. Soc. Jpn.* **45** 1897 (1972)
38. J Dugunji, I Ugi *Top. Curr. Chem.* **39** 19 (1973)

39. C Jochum, J Gasteiger, I Ugi *Angew. Chem., Int. Ed. Engl.* **19** 495 (1980)
40. C Jochum, J Gasteiger, I Ugi *Z. Naturforsch., B Chem. Sci.* **37** 1205 (1982)
41. I Ugi, M Wochner *J. Mol. Struct.* **42** 229 (1988)
42. M Wochner, J Brandt, A v Scholley, I Ugi *Chimia* **42** 217 (1988)
43. I Ugi, M Wochner, E Fontain, J Bauer, B Gruber, R Karl, in *Concepts and Applications of Molecular Similarity* (Eds M A Johnson, G M Maggiora) (New York: Wiley, 1990) p. 239
44. J Koča, M Kratochvíl, V Kvasnička, L Matyska, J Pospíchal *Synthon Model of Organic Chemistry and Synthesis Design (Lect. Notes Chem.)* Vol. 51 (Berlin: Springer, 1989)
45. V Kvasnička, M Kratochvil, J Koča *Collect. Czech. Chem. Commun.* **48** 2284 (1983)
46. J Koča *Collect. Czech. Chem. Commun.* **53** 1007 (1988)
47. J Koča *Collect. Czech. Chem. Commun.* **53** 3108 (1988)
48. J Koča *Collect. Czech. Chem. Commun.* **53** 3119 (1988)
49. J Koča *J. Math. Chem.* **3** 73 (1989)
50. J Koča *J. Math. Chem.* **3** 91 (1989)
51. V Kvasnička, J Pospíchal *J. Math. Chem.* **5** 309 (1990)
52. V Kvasnička, J Pospíchal *J. Math. Chem.* **3** 161 (1989)
53. V Kvasnička, J Pospíchal *Int. J. Quantum Chem.* **38** 253 (1990)
54. J B Hendrickson, C A Parks *J. Chem. Inf. Comput. Sci.* **32** 209 (1992)
55. S V Dozmorov *React. Kinet. Catal. Lett.* **19** 289 (1982)
56. S H Bertz, in *Chemical Applications of Topology and Graph Theory* (Ed R B King) (Amsterdam: Elsevier, 1983) p. 206
57. S H Bertz, W C Herndon *ACS Symp. Ser.* **306** 169 (1986)
58. J Gastieger, M G Hutchings, P Löw, H Saller *ACS Symp. Ser.* **306** 258 (1986)
59. V A Likholobov, L I Maier, N N Bulgakov, A V Fedotov, L G Shteingauer, in *Materialy V Mezhdunar. Simp. po Svyazi Mezhdu Gomogennym i Geterogennym Katalizom, Novosibirsk, 1986* (Proceedings of the Fifth International Symposium on the Relation between Homogeneous and Heterogeneous Catalysis, Novosibirsk, 1986) p. 231
60. L I Shtokolo, L G Shteingauer, V A Likholobov, A V Fedotov *React. Kinet. Catal. Lett.* **26** 227 (1984)
61. V A Likholobov, L I Mayer, N N Bulgakov, A V Fedotov, L G Shteingauer, in *Homogeneous and Heterogeneous Catalysis* (Eds Yu Yermakov, V Likholobov) (Novosibirsk: Institute of Catalysis, Siberian Branch of the Russian Academy of Sciences, 1986) p. 229
62. L G Shteingauer, L I Mayer, N N Bulgakov, A V Fedotov, V A Likholobov *React. Kinet. Catal. Lett.* **36** 139 (1988)
63. L G Zabolotnaya, Candidate Thesis in Physicomathematical Sciences, Institute of Catalysis, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 1989
64. L I Maier, Candidate Thesis in Chemical Sciences, Institute of Catalysis, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 1990
65. J Gasteiger *Tetrahedron* **35** 1419 (1979)
66. J Gasteiger, P Jacob, U Strauss *Tetrahedron* **35** 139 (1979)
67. R Barone, M Chanon, M L H Green *J. Organomet. Chem.* **185** 85 (1980)
68. I Theodosiou, R Barone, M Chanon *J. Mol. Catal.* **32** 27 (1985)
69. I Theodosiou, R Barone, M Chanon *Adv. Organomet. Chem.* **26** 165 (1986)
70. D Forster *Adv. Organomet. Chem.* **17** 255 (1979)
71. J F Roth, J H Craddock, A Hershman, F E Paulik *Chem. Technol.* **1** 600 (1971)
72. P J Collman, L S Hegedus, J R Norton, P Finke *Principles and Applications of Organotransition Metal Chemistry* (Mill Valley, CA: University Sci. Books, 1987)
73. E Fontain, J Bauer, I Ugi *Chem. Lett.* 37 (1987)
74. E Fontain, K Reitsam *J. Chem. Inf. Comput. Sci.* **31** 97 (1991)
75. E Fontain *Tetrahedron Comput. Method.* **3** 469 (1990)
76. I Ugi, J Bauer, K Bley, A Dengler, E Fontain, M Knauer, S Lohberger *J. Mol. Struct.* **230** 73 (1991)
77. I Ugi, E Fontain, J Bauer *Anal. Chim. Acta* **235** 155 (1990)
78. I Ugi, J Bauer, R Baumgartner, E Fontain, D Forstmeyer, S Lohberger *Pure Appl. Chem.* **60** 1573 (1988)
79. E J Corey, W T Wipke *Science* **166** 178 (1969)
80. E J Corey, X-M Cheng *The Logic of Computer Synthesis* (New York: Wiley, 1989)
81. Y Yoneda *Bull. Chem. Soc. Jpn.* **52** 8 (1979)
82. D Bonchev *Information Theoretic Indices for Characterization of Chemical Structures* (Chichester: Research Studies Press, 1983)
83. R E Valdés-Pérez, Ph. D. Thesis, Carnegie Mellon University, School of Computer Science, 1990
84. R E Valdés-Pérez *Tetrahedron Comput. Method.* **3** 277 (1990)
85. R E Valdés-Pérez *J. Comput. Chem.* **13** 1079 (1992)
86. R E Valdés-Pérez *J. Phys. Chem.* **96** 2394 (1992)
87. R E Valdés-Pérez *J. Comput. Chem.* **14** 1454 (1993)
88. R E Valdés-Pérez *Catal. Lett.* **28** 79 (1994)
89. R E Valdés-Pérez *J. Chem. Inf. Comput. Sci.* **34** 976 (1994)
90. R E Valdés-Pérez *J. Comput. Chem.* **15** 1266 (1994)
91. R E Valdés-Pérez *J. Chem. Inf. Comput. Sci.* **31** 554 (1991)
92. R E Valdés-Pérez *Artif. Intell.* **65** 247 (1994)
93. R E Valdés-Pérez, private communication
94. D Bonchev, O N Temkin, D Kamenski *React. Kinet. Catal. Lett.* **15** 119 (1980)
95. D Bonchev, D Kamenski, O N Temkin *J. Math. Chem.* **1** 345 (1987)
96. E V Gordeeva, D Bonchev, D Kamenski, O N Temkin *Chem. Inf. Comput. Sci.* **34** 436 (1994)
97. A V Zeigarnik, O N Temkin *Kinet. Katal.* **37** (1996) [*Kinet. Catal. (Engl. Transl.)* **37** (1996) (in the press)]
98. G M Ostrovskii, A G Zyskin, Yu S Snagovskii, in *Fizicheskaya Khimiya: Sovremennye Problemy* (Physical Chemistry: Current Problems) (Ed. Ya M Kolotyrkin) (Moscow: Khimiya, 1986) p. 84
99. Yu S Snagovskii, G M Ostrovskii *Modelirovanie Kinetiki Geterogennykh Kataliticheskikh Protsessov* (Kinetic Modelling of Heterogeneous Catalytic Processes) (Moscow: Khimiya, 1976)
100. T Morikawa *Match* **13** 7 (1982)
101. D I Kaminski, O N Temkin, D G Bonchev *Appl. Catal.* **88** 1 (1992)
102. J Happel, P H Sellers *Ind. Eng. Chem., Fundam.* **21** 67 (1982)
103. J Happel, P H Sellers *Chem. Eng. Commun.* **83** 221 (1989)
104. P H Sellers, in *Chemical Applications of Topology and Graph Theory* (Ed. R B King) (Amsterdam: Elsevier, 1983) p. 420
105. V I Dimitrov *Prostaya Kinetika* (Simple Kinetics) (Novosibirsk: Nauka, 1982)
106. R H S Mah, R Aris *Ind. Eng. Chem., Fundam.* **2** 90 (1963)
107. V V Kafarov, V N Pisarenko, A V Solokhin *Zh. Fiz. Khim.* **52** 3143 (1978)
108. M L Mavrovouniotis, G Stephanopoulos *Ind. Eng. Chem. Res.* **31** 1625 (1992)
109. M L Mavrovouniotis *Ind. Eng. Chem. Res.* **31** 1637 (1992)
110. M L Mavrovouniotis, G Stephanopoulos, G Stephanopoulos *Biotechnol. Bioeng.* **36** 1119 (1990)
111. M L Mavrovouniotis, G Stephanopoulos, G Stephanopoulos *Comput. Chem. Eng.* **16** 605 (1992)
112. M L Mavrovouniotis, in *Artificial Intelligence Approaches in Product and Process Engineering* (New York: Academic Press, 1995) Ch. 3
113. T D Salatin, W L Jorgensen *J. Org. Chem.* **45** 2043 (1980)
114. W L Jorgensen, E R Laird, A J Gushurts, J M Fleischer, S A Gothe, H E Helson, G D Paderes, S Sinclair *Pure Appl. Chem.* **62** 1921 (1990)
115. M A Baltanas, G F Froment *Comput. Chem. Eng.* **9** 71 (1985)
116. B O Pol'geim, L G Bruk, S I Shalgunov, O N Temkin, in *Tez. Dokl. VII Vsesoyuz. Konf. "Matematicheskie Metody v Khimii", Kazan', 1991* (Abstracts of Reports at the Seventh All-Union Conference "Mathematical Methods in Chemistry", Kazan', 1991) p. 220
117. S I Shalgunov, L G Bruk, O N Temkin, unpublished results
118. S I Shalgunov, unpublished results
119. C A Tolman *Chem. Soc. Rev.* **1** 337 (1972)