Computer-Assisted Solution of Chemical Problems—The Historical Development and the Present State of the Art of a New Discipline of Chemistry

By Ivar Ugi,* Johannes Bauer, Klemens Bley, Alf Dengler, Andreas Dietz, Eric Fontain, Bernhard Gruber, Rainer Herges, Michael Knauer, Klaus Reitsam, and Natalie Stein

Dedicated to Professor Karl-Heinz Büchel

The topic of this article is the development and the present state of the art of computer chemistry, the computer-assisted solution of chemical problems. Initially the problems in computer chemistry were confined to structure elucidation on the basis of spectroscopic data, then programs for synthesis design based on libraries of reaction data for relatively narrow classes of target compounds were developed, and now computer programs for the solution of a great variety of chemical problems are available or are under development. Previously it was an achievement when any solution of a chemical problem could be generated by computer assistance. Today, the main task is the efficient, transparent, and non-arbitrary selection of meaningful results from the immense set of potential solutions—that also may contain innovative proposals. Chemistry has two aspects, constitutional chemistry and stereochemistry, which are interrelated, but still require different approaches. As a result, about twenty years ago, an algebraic model of the logical structure of chemistry was presented that consisted of two parts: the constitution-oriented algebra of be- and r-matrices, and the theory of the stereochemistry of the chemical identity group. New chemical definitions, concepts, and perspectives are characteristic of this logic-oriented model, as well as the direct mathematical representation of chemical processes. This model enables the implementation of formal reaction generators that can produce conceivable solutions to chemical problems—including unprecedented solutions—without detailed empirical chemical information. New formal selection procedures for computer-generated chemical information are also possible through the above model. It is expedient to combine these with interactive methods of selection. In this review, the Munich project is presented and discussed in detail. It encompasses the further development and implementation of the mathematical model of the logical structure of chemistry as well as the experimental verification of the computer-generated results. The article concludes with a review of new reactions, reagents, and reaction mechanisms that have been found with the PC-programs IGOR and RAIN.

1. Introduction

Chemistry is concerned with the analysis and synthesis of chemical substances, the elucidation of molecular structures, and the reactivity and physical properties of chemical compounds. Apart from the fundamental gain in scientific knowledge, the manifold applications of chemistry justify the endeavors and expenditure of chemical research. The search for compounds with desirable effects, functions, and properties and the development of optimal methods for their synthesis, purification, and quality control are characteristic of research in applied chemistry.

Progress in chemistry is generally due to systematic research. It is evolutionary and takes place in small steps.[1] The rare revolutionary jumps of innovation are produced by new ways of thinking or expertly recognized and systematically exploited, fortunate, incidental observations. Examples are the insight into the significance of conformations[2] by Sachse[3] and Mohr,[4] Meerwein's concept of ionic reaction mechanisms,[5] the discovery of the first organometallic catalysts containing transition elements by Reppe et al.[6a] and Roelen,[7a] the solution of the absolute configuration problem by Bijvoet et al.,[8] and the development of the chemistry of superacids by Olah et al.[9]

For progress in applied chemistry, serendipity and initial misunderstandings can play an important role. Spectacular examples are the introduction of the sulfonamides by Domagk, Mietsch, and Klarer,[10] who thereby revolutionized chemotherapy of bacterial infections, and the discovery of the azole fungizides by Büchel and Plempel[11] that enabled efficient and broad chemotherapy of fungal infections.

The diversity of applications for computers in chemistry reflects the variety in chemical research. They have been used since they have been available and have become indispensable in all areas of chemistry. In 1959 Konrad Zuse,[12] the inventor of the computer, sold the first commercially available machine, the magnetic drum computer Z22, to Bayer AG, who used it partly for scientific purposes.

[*] Prof. Dr. I. Ugi, Dr. J. Bauer, Dr. K. Bley, Dr. A. Dengler, Dipl.-Chem. A. Dietz, Dr. E. Fontain, Dr. B. Gruber, Dr. M. Knauer, Dipl.-Chem. K. Reitsam, Dipl.-Chem. N. Stein
Organisch-chemisches Institut der Technischen Universität München
Lichtenbergstrasse 4, D-W-8046 Garching (FRG)
and
Dr. R. Herges,
Organisch-chemisches Institut der Universität, Henkestr. 42,
D-W-8520 Erlangen (FRG)

Initially computers were exploited exclusively for numerical computations, an application that is still on the increase. The term computational chemistry[13-15] is used for such applications, in which the calculation of molecular energy levels and geometries prevails. Their numerical capabilities make computers an indispensable tool of quantum chemistry and analytical chemistry, which include the elucidation of molecular structure by spectroscopic and X-ray methods. However, they are also useful in the determination of properties of substances, the design of experiments, the collection, evaluation, graphic display, and interpretation of experimental data, as well as in the visualization of molecular structures whose atomic coordinates are obtained from measurements, by quantum chemical computations, or by force-field methods.

To a larger extent these applications of computers in chemistry —mostly computations according to fixed "recipes"— can be automated. With the exception of molecular modeling where both the numerical and graphic capabilities of computers are exploited, an interactive participation of the user is not needed.

Many types of chemical problems can, however, not be solved by numerical computations. The solution of these chemical problems requires another kind of approach, incorporating the logical and combinatorial capabilities of computers. This discipline is called computer chemistry.[16-20] The latter term is not too well chosen, but is now widely used.

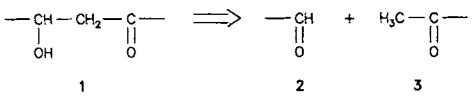
The oldest and best known computer programs for the direct solution of chemical problems serve to design syntheses and elucidate molecular structures from spectroscopic data. The search for synthetic routes or reaction mechanisms and the elucidation of molecular structures from spectroscopic data belong to the "classical" problems of chemistry. The solution of such problems involves finding molecules that have certain characteristic chemical features. Generally, the classical chemical problems have a significant combinatorial aspect, and have many solutions that cannot be obtained by numerical computations, although such computations may be of indirect use.[21] The best solutions are often found by intuition, creative thinking, or trial and error. The computer-assisted solution of chemical problems requires the interactive participation of the user, because there are always many potential solutions and no generally applicable rules. Therefore, a nonarbitrary selection of the practicable solutions is needed. The methods for solving such problems are, as a rule, nondeterministic in nature.

Joshua Lederberg and George Vleduts, who died in 1990, are the founding fathers of computer chemistry. Lederberg[22] initiated the DENDRAL project[23] to develop a computer program that determines the significant structural features of a molecule from spectroscopic data and then assembles its complete structure from its substructures.[24] Although, the DENDRAL project was terminated before all its goals were achieved, it nevertheless represents an important milestone in the development of computer chemistry.[13] It was the first computer program that exploited formal means for the solution of chemical problems. DENDRAL also promoted the notion artificial intelligence.[23]

In 1963, Vleduts[25] proposed the design of computer programs for planning multistep organic syntheses (CAOS: Computer-Assisted Organic Synthesis),[26] that is, for finding a way to synthesize a given target molecule from available starting materials in the best possible way according to the criteria that are applicable. As a pioneer of data banks of chemical reactions,[27] Vleduts was the first to discuss in detail how the reverse of stored chemical reactions, the so-called retroreactions, can be used to generate synthetic routes that lead from the target molecule via precursors to available starting materials. For instance, the structural feature **1** in a target molecule leads to the chapter of aldol condensations in the reaction library and thus to the precursors **2** and **3**.

From the precursors of the target molecule their precursors are obtained in turn, until the starting materials are reached. Thus, a tree of syntheses results that generally con-

Ivar Ugi was born in Arensburg in Estonia in 1930. The family moved to Dillingen an der Donau (Germany) in 1941, where Ivar Ugi completed his Abitur in 1949. In the fall of the same year he began his study of chemistry at the University of Tübingen, which he completed with a doctorate at the University of Munich in 1954 under Prof. Huisgen. He obtained his habilitation in 1959 for his work on pentazoles and isonitriles and continued as a research associate at the university for a further two years. In 1962 he joined the company Bayer in Leverkusen and rose to Director and Chairman of the Kommission für Grundlagenforschung. After six years his spirit of research led him to the USA: in 1968 Ugi became Full Professor of Chemistry at the University of Southern California in Los Angeles. In July 1971 he accepted a chair rich in tradition at the Technische Universität München, the Lehrstuhl I für Organische Chemie. Ivar Ungi is a member of committees for many foundations and has been a full member of the Swedish Royal Society for Science since 1987 and the Estonian Academy of Sciences since 1990. His scientific endeavors have been acknowledged with many awards: in 1964 he received the research prize of the Akademie der Wissenschaft in Göttingen and in 1988 the Philip Morris research prize for the Dugandji–Ugi model which he had developed with the topologist Dugandji. His life's work was honored in 1992 with the Emil-Fischer-Medaille der Gesellschaft Deutscher Chemiker not only for the development of mathematical models of the logical structure of chemistry, but also for the many new preparative methods arising from his isonitrile chemistry, four-component condensation, and stereochemical models, which open several new fields of research. These synthetic methods give strong impetus in particular to the chemistry of peptides, nucleotides, and β-lactam antibiotics.



1 2 3

tains many pathways from which one has to be selected. Corey[28] introduced the term retrosynthetic analysis for this procedure.

Bilateral synthesis design is an alternative to the retrosynthetic approach. When both the target molecule and a suitable set of starting materials are known, it suffices to find pathways that lead from the starting materials to the target molecule.

Computers can only be used for the solution of chemical problems when sufficient knowledge and experience are available for model concepts and reasoning by analogy. In the exploration of new domains in chemistry and in areas of incipient research, novel insights, visions, new ideas, and fortunate coincidences are still indispensable. Computers are here, at best, of indirect use, especially as a chemist would probably reject an "imaginative" computer-generated proposal as nonsensical artifact, even if it were completely realizable, instead of checking it in the laboratory.

Since computers do not have any creative intelligence, they are neither able to generate revolutionary ideas nor capable of pioneering scientific achievements. However, as an "amplifier of intelligence" a computer can, within the given body of knowledge, accelerate the progress of science, and also give impetus to research work that would not be feasible without computer-assistance. Even if a computer were able to generate all the conceivable solutions for a given chemical problem, the meaningful and creative solutions could not be selected without the participation of a qualified user.

2. Computer-Assisted Synthesis Design

2.1. Empirical Synthesis Design Assisted by Reaction Libraries

The programs for retrosynthetic design strategy based on data from reaction libraries LHASA,[29, 30] Wipke's SECS,[31] and Gelernter's SYNCHEM,[32] as well as the Leverkusener Peptide Synthesis Design Program, which operates along the bilateral principle,[33, 34] belong to the first computer programs for the solution of chemical problems. Since then, many groups have tried to develop computer programs with various success for the design of syntheses[18, 35].

Since the beginning of organic synthesis, chemists have planned syntheses by retrosynthetic reasoning,[36] mentally dissecting the target molecule while bearing in mind known synthetic reactions. Sir Robert Robinson[37] employed retrosynthetic reasoning particularly systematically. The disconnection method as recommended by Warren[38] also exercises retrosynthetic reasoning.

The early synthetic chemists did not attach a name to their procedure of synthesis design. Since the introduction of the terms synthon, transform, and retrosynthesis (also called antithesis), as well as the transform arrows (⇒) by Corey,[28] publications often contain a retrosynthetic analysis performed after completion of the synthesis. Multistep syntheses are certainly all based on some plan; however, the complex multistep synthetic pathways that are successfully executed in the laboratory cannot yet be planned to the last detail. They evolve from the set of proposed syntheses by trial and error and by successive adjustment of the design to what is feasible.

The empirical retrosynthetic programs are typical expert systems[39] with a knowledge basis and a set of rules. The empirical retrosynthetic programs generate synthetic pathways according to the perceived structural features of the target by application of the transforms, rules, and schemes. The precursors of a given target molecule are generated from it by breaking and making bonds according to stored information on chemical reactions and heuristic rules (see **1**–**3**). The process is iterated for each precursor until available starting materials are reached.

An empirical retrosynthetic program that is not confined to a narrow field of applications and a small repertoire of chemistry requires a large reaction library and a powerful computer. The creation and updating of large error-prone reaction libraries and the incorporation of synthesis-related experience into empirical synthesis design programs is time-consuming and costly. Reactions that are sequences of individual steps are particularly difficult to handle, because redundant information is stored. An advantage of empirical retrosynthetic programs is that they exploit existing chemical experience and therefore yield plans of syntheses that are combinations of known reactions and thus likely to succeed in the laboratory. The reaction library also determines the limits of the program, which is unable to break new territory. Moreover, since it is impossible to store all known reactions, reaction libraries provide to a more or less arbitrary choice of reactions.

Conceivable precursors can exist at any level of synthesis design. Thus, a rapidly expanding tree of synthetic pathways is generated. It is necessary—as in chess—to select particularly promising precursors with a view to finding the next precursors. Within the framework of programs for retrosynthetic analysis, this selection is based on heuristic strategies and selection rules.

The empirical retrosynthesis programs reached a very high level of perfection about 10 years ago. The most advanced example is CASP,[40] a modification and extension of SECS,[31] which was developed by large German and Swiss chemical companies at great expense. It is hardly conceivable that, apart from details, synthesis design programs based on reaction libraries will significantly surpass the present state of the art.

CASP was not sufficiently accepted by potential users, probably because its extremely large reaction library needs a powerful computer. It therefore cannot be directly accessed by bench chemists, but requires a computer specialist as a middleman. It also seems that chemists prefer to use stored information directly instead of having it manipulated by an empirical synthesis design program.

2.1.1. Which Are the Strategic Bonds in Ring Systems?

Many interesting target molecules contain complex polycyclic systems for which suitable strategies need to be planned

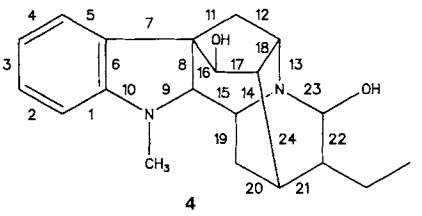
first.[41] Corey et al.[42,43] developed heuristic strategies for the synthesis of carbocyclic systems, which were particularly important computer-assisted retrosynthetic analysis. Their rules refer to "strategic" bonds that are preferentially broken and made, and can be summarized as follows (see Section 4.3.2):

1. A bond is strategic if it belongs to a four-, five-, six-, or seven-membered primary ring. A ring is primary if it does not contain two or more smaller rings.
2. A bond is strategic if it is directly connected to another ring (exo to another ring), unless this ring has only three members.
3. A strategic bond should belong to the ring that contains the highest number of bridged positions (bridgehead atoms).
4. If a ring of seven or more members is formed when a bond is broken, this bond is not strategic.
5. Bonds of an aromatic system are not strategic.
6. If a bridge contains a chiral center, none of its bonds are strategic, unless directly connected to the chiral center.

Bonds between carbon and heteroatoms are treated differently. Only rules 4, 5, and 6 apply, and rule 2 if a three-membered ring is involved.

2.1.2. *Difficulties in Identifying Strategic Bonds*

If the rules for strategic bonds are applied to bridged systems with four or more rings (according to Frèrejacques counting[44]) the results are sometimes less than satisfactory, especially when heteroatoms are present. This is because no hierarchy was defined for these rules, and they were conceived without sufficient attention to heteroatoms. As a consequence, in many cases several bonds are recognized as "strategic", with equal priority. An example is ajmaline (**4**),[45] for which six strategic bonds (bonds 9, 10, 15, 16, 23, and 24) are found according to the rules. Breaking bonds 9



4

and 10 does not simplify the molecular structure, while breaking only one bond does not lead to the formation of an accessible precursor, so other strategic bonds have to be determined. However, this does not necessarily lead to a simplification of the problem, because a burgeoning number of bonds qualify as strategic when the complexity of the ring system decreases. This leads to ever more precursors that have to be processed. Thus, breaking bond 16 of **4** produces seven new strategic bonds.

As mentioned above, the dissection of a bond does not necessarily lead to a "simplified" molecule. Corey's rules also lack a criterion for terminating ring fragmentations. Termination of the procedure after dissection of all bridges between rings would be advantageous; thus, nonbridging bonds such as 9 and 10 in **4** would not be considered.

Heuristic rules are not equally valid in all areas of chemistry, and usually their domains of validity cannot be clearly defined. The automated use of heuristic rules is therefore rather arbitrary and may lead to errors.[21]

Moderate amounts of data can be examined interactively by the user. This is widely applicable, and in computer-assisted synthesis design the interactive selection of synthetic routes by a knowledgeable chemist will yield better results than any automated, heuristic selection procedure. In automated retrosynthetic analysis, synthetic routes with favorable final steps will invariably be preferred over those with less favorable final steps.[46]

2.2. Semiformal Synthesis Design and Prediction of Reactions

The most severe limitations in "creativity" of the present semiformal synthesis design programs are their automated heuristic selection procedures; they are neither transparent nor well-founded.[21] Therefore, the full set of conceivable solutions can not be taken into account.

In principle, semiformal synthesis design programs can carry out all kinds of retrosynthetic analyses. However, they do not have selection procedures that can scan very large sets of solutions. Accordingly, the semiformal approach is limited to relatively small synthetic problems that fall within the domain of validity of the applicable rules and the underlying schematic procedures of the programs. Without a suitable formalism no interactive ordering and classification of the conceivable solutions is possible. For semiformal synthesis design programs, the selection procedures for precursors and retroreactions can still be substantially improved, but no major progress in generating the synthetic pathways seems possible.

In contrast to the empirical synthesis design programs, the semiformal and in particular the formal synthesis design programs have considerable innovative abilities. The semiformal programs cannot intentionally be used to "invent" new reactions (in contrast to the formal programs, e.g. IGOR; see Section 4.4.3), but can invent new types of molecules and chemical reactions when they are needed as a part of a synthetic route that is generated.

To our knowledge, no new compounds or reactions that have been predicted by a semiformal synthesis design program have been subsequently verified by experiment.

The program CAMEO of Jorgensen et al.[47] was developed for the prediction of reaction products, but it can also be used for the design of syntheses. CAMEO is based on the idea that most organic reactions can be represented by a combination of a few mechanistic elementary reactions (see also ref. [48]). The required chemical information is incorporated into individual program modules. Since CAMEO generates chemical reactions from their mechanistic steps, it is more capable of innovation than the retrosynthesis programs based on reaction libraries.[49]

Hendrickson[50] introduced the so-called half-reactions that can be combined to full chemical reactions according to valency considerations, without employing reaction mechanisms. Moreau[51] developed the semiformal synthesis design program MASSO on this principle. Shortly afterwards, Hen-

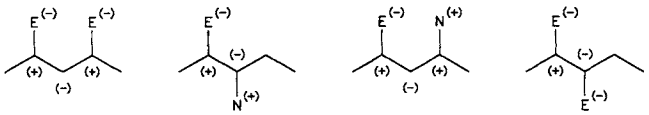
drickson et al.[52] completed SYNGEN, a similar program that is able to generate convergent syntheses from, at most, four subunits of the target molecule.

Modification and extension of the feasibility study CICLOPS[53] led to the first versions of the synthesis design program EROS,[54] which are of the formal type. Later versions of EROS[55] are essentially semiformal.

The selection processes of EROS for synthetic routes are heuristic in nature. They are based on assumptions on the lability of bonds and on physical data, which restricts the scope of reactions by defining, certain bonds as "breakable". In the course of multistep syntheses, often more than ten bonds are broken or made. Under the assumption that one to three bonds are broken or made per reaction, there are more than 7×10^6 synthetic routes that differ in the sequence of their operations, and that are by no means always equivalent (they would only be equivalent if all operations were independent of each other).

The semiformal synthesis design program TOSCA[56] is most remarkable because it is based on Evans' concept of consonant and dissonant molecules. In a lecture at the University of California at Los Angeles (UCLA) on 6th May 1971, D. A. Evans demonstrated the importance of consonant and dissonant structural features of molecules. Evans documented this concept in a paper with the title "Consonant and Dissonant Relationships—An Organizational Model for Organic Syntheses".[57] Although this paper was never published, its contents became well known in the chemical community, because in 1972 Evans distributed many copies among interested colleagues.

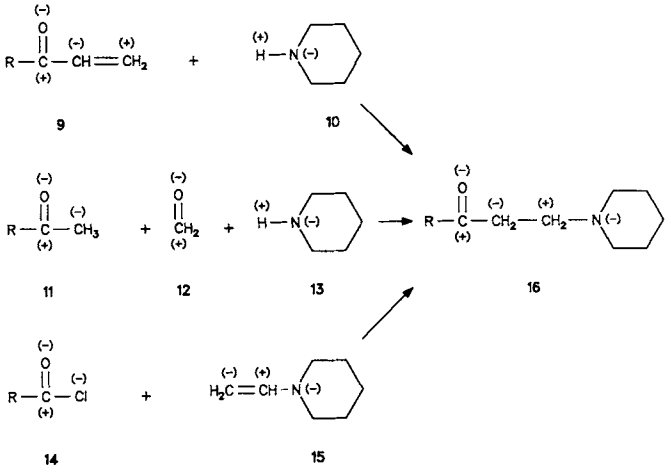
Evans' concept is based on Lapworth's polarity patterns of organic molecules that contain heteroatoms.[58] Evans termed these charge affinity patterns. Molecules like **5** and **6** are called consonant. They have alternating sites of positive and negative charge affinity, which can react as a nucleophile or an electrophile respectively. Molecules like **7** and **8** to



5 **6** **7** **8**

which no such patterns can be assigned are called dissonant. The syntheses of **16** from **9** and **10**, **11**, **12**, and **13**, and **14** and **15** are examples of consonant syntheses.

Evans showed that the conversion of consonant molecules into dissonant molecules and vice versa by ionic reactions requires "inversion operations", which correspond to Umpolung as defined by Seebach.[59] Recognition of consonant and dissonant structural features in the starting material and products of a synthesis is very important for the design of syntheses because the need for inversion operations must be taken into account. We use the Evans–Lapworth schemes in our bilateral synthesis design program RAIN (Reaction And Intermediate Networks).[60–63] With his consonance/dissonance concept Evans has made a far-reaching contribution to computer-assisted synthesis design.



9 **10** **11** **12** **13** **16** **14** **15**

2.3. Synthesis Design with Formal Foundations

2.3.1. What Does "Formal" Mean?

The disadvantages of programs for empirical, semiformal synthesis design can only be avoided by a comprehensive mathematical theory of chemistry that can be used to generate and to select by formal means all molecular systems and chemical reactions that must be accounted for in the selection of a given chemical problem.[21]

Different disciplines use the term formal differently. D. Hilbert and Bernays[64] called mathematics "the science of the formal systems". In informatics formal descriptions require a language (syntax), which may be mathematics, and a mathematically expressed meaning (semantics) together with definitions of rules for transformations and proofs. The validity and domain of a formal algorithm must be proven like a theorem. In chemistry, the term formal is not defined precisely, and is used quite randomly for all kinds of generalizations and classifications, as well as for more or less abstract representations and algorithms. As an empirical science, chemistry is not subject to any rigorously formal approaches in the sense understood in mathematics and informatics. However, our group attempts to put computer-assistance in chemistry onto a formal basis, as far as this is possible. Whenever this is not feasible, we avoid the use of heuristic rules with unpredictable consequences, leaving the required decisions to interactive user participation. The formal description of chemistry requires an adequate, detailed translation of chemistry, including its dynamic aspect, into the language of mathematics, and also a transparent translation of mathematical representations of chemical objects and facts back into the language of chemistry. This is accomplished through a mathematical model of the logical structure of chemistry that represents the molecular systems as well as the chemical changes that they may undergo (see Section 3).

Such a mathematical model can be implemented in computer programs whose applications are not restricted to a single type of problems and that can generate and select, through a formal algorithm, solutions of chemical problems without files of stored, detailed chemical information.[21] Such computer programs are well suited to explore new chemical terri-

tory, because they can propose molecules and chemical reactions that are without precedent, in particular when in dialogue with a creative, knowledgeable, and experienced chemist.

With suitable computer programs all of the combinatorially possible solutions of chemical problems, including those without precedent, can be taken into account. This is, however, of little use if the generated data cannot be ordered and the desired information cannot be extracted. To some extent, this is possible by formal means, but the interactive participation of the user is still required. His intuition and expertise are not only needed to formulate the problem, but also to select the best solutions.

The widespread opinion that a computer does not produce anything that has not been previously provided as data sounds plausible. It is, however, only strictly true for data banks and expert systems that rely solely on stored data.

2.3.2. *Mathematization of Chemistry*

Quantum chemistry has contributed a great deal to the mathematization of chemistry via physics. Quantum chemistry, a branch of molecular theoretical physics, is indispensable for understanding chemistry and for computing the numerical, measurable properties of well-defined molecular systems. The goal of mathematical chemistry is the mathematization of chemistry without the intermediacy of physics and the direct solution of chemical problems by qualitative mathematical methods. The traditional approaches for treating chemical problems by qualitative formulations of discrete mathematics are confined to a static perspective. Such mathematical chemistry does not reach beyond a group theoretical visualization and interpretation of the chemical constitution of molecules and of families of isomers[65-68] and the graph theoretical classification and enumeration of isomers and products of isomerizations.[69-77]

Neither quantum chemistry, which is often called theoretical chemistry, nor traditional mathematical chemistry are suitable as a theoretical basis for the solution of chemical problems with a strong combinatorial aspect, such as the search for molecules or chemical reactions that meet given chemical requirements. The computer-assisted solution of chemical problems by formal means requires global mathematical modeling of chemistry beyond the treatment of individual chemical objects. The model must represent the relations between the objects and the logical structure of chemistry as a whole. Constitutional chemistry and stereochemistry belong together, and yet they are so different that they require distinct mathematical perspectives and approaches. A global mathematical model of chemistry must therefore consist of two parts, one that represents constitutional chemistry and one that describes stereochemistry.

The theory of the be- and r-matrices[78] is an algebraic model of the logical structure of constitutional chemistry (see Section 3.2). It is also cited[79-81] as the "Dugundji–Ugi Model" or "Dungundji–Ugi Theory", abbreviated to the DU model. Besides the algebraic DU model, a graph theoretical model that was published later by Kvasnička et al.[80,81] can serve as a mathematical foundation of formal synthesis-design programs. The logical structure of stereochemistry can be represented by the theory of the Chemical Identity Group (CIG theory),[82] which is based on the notion of permutational isomerism.[76]

In contrast to traditional mathematical chemistry, the above models not only refer to, but even emphasize the dynamic aspect of chemistry. Within the framework of the DU model, chemical reactions are described by transformations of be-matrices. The be-matrices represent constitutional formulas which, in turn, correspond to graphs of the chemical constitution of molecules. In the CIG theory, the set-valued maps represent dynamic processes by which the stereochemical features of molecules are changed. These mathematical devices for the direct modeling of chemical processes are the theoretical foundation of the computer-assisted, deductive solution of a wide variety of chemical problems without reference to detailed empirical chemical information.

In the DU model and the CIG theory, the direct translation of chemistry into mathematics and vice versa is more important than the actual mathematical basis. Mathematics thus becomes part of the language of chemistry bringing its semantics and syntax, just as chemical formulas have been for decades.

2.4. Programs for Formal Synthesis Design

Through the DU model the objects of chemistry (molecules, ensembles of molecules (EM), chemical reactions) can be represented by objects of mathematics (matrices). Thus, chemical problems can be translated into mathematical problems, whose solutions are the solutions of the chemical problems.

This model started the development of formal algorithms, reaction generators, and computer programs for the deductive solution of chemical problems by mathematical means. The formal algorithms can not only generate solutions for many types of chemical problems but are also useful in the classification and selection of these solutions.[21]

The development of programs for formal synthesis design began with the feasibility study CICLOPS.[53] The purpose of CICLOPS was to ascertain whether chemical problems could, in principle, be solved with the DU model. In this respect the CICLOPS study was successful. However, it was also found that CICLOPS was bound to fail under the conditions of practical synthesis design of problems incorporating the combinatorial aspect. Gasteiger and Jochum[54] developed the first versions of EROS from CICLOPS, by reducing the reaction generator (see below) of CICLOPS from approximately 100 potential to the three most important r-classes (see Section 3.2.2) of chemical reactions and by adding a selection procedure for chemical reactions. In analogy to the selection procedure of Stevens and Brownscombe,[83] the choice is made on the basis of reaction enthalpy estimations.

Reaction matrices can be assembled from basic elements that correspond to the elementary reactions of mechanisms.[78] These can be used to create reaction generators that build chemical reactions from their elementary steps. The synthesis design program ASSOR[48] contains a reaction generator of this type and is therefore able to account for the mechanistic aspects of chemical reactions.

The DU model can be used in many ways as the theoretical foundation of monolateral synthesis design programs that

generate synthetic routes either from the target molecule in a retrosynthetic mode, or from the starting material in a synthetic mode. However, the possibilities that the DU model provides for the bilateral design of syntheses[84, 85] seem to be more promising. The obvious combinatorial advantages of bilateral synthesis design have also been recognized by other authors. Johnson et al.[86] developed a bilateral synthesis-design program on the basis of LHASA, and recently Hendrickson and Parks[87] combined SYNGEN with the synthesis program FORWARD into a bilateral synthesis-design program.

In bilateral synthesis design, the synthetic routes are spread from both ends, the starting materials and target. At the moment, we are developing a system of computer programs for bilateral synthesis design that operates in three stages.[85, 88, 89] The partitioning of the problems by such a system provides substantial combinatorial advantages. First, a set of suitable starting materials is selected for the target molecule from a list of available compounds by the substructure correlation program CORREL-S (see Section 4.3.2). Then the co-products (ancillary products formed besides the target molecule) are determined (program STOECH, see Section 4.3.3). The result is a target EM that is isomeric (in terms of Section 3.1) to the EM of starting materials. In a third step, the computer program RAIN (see Section 4.4) generates a network of synthetic routes that connect both ends of the synthesis.

Other formal synthesis-design programs are FLAMINGOES,[90] PEGAS,[91] and MAPOS.[92] Hippe[93] tried to combine the advantages of the formal and empirical approaches in their programs SCANSYNTH, SCANMAT, and SCANPHARM. So did Johnson et al.[94] in their system SYNLMA.

3. The Logical Structure of Chemistry

3.1. The Hierarchy of Isomerisms

Chemistry is an empirical science. Nevertheless, its objects have a consistent logical structure, because all molecules are constructed according to uniform principles.[95] This logical structure is a system of equivalence relations. The most important relations are the various types of isomerism and the interconvertibility of isomeric molecules and ensembles of molecules. The logical structure is determined by the valence properties of the chemical elements and a few general principles. When chemistry is compared to a language, the chemical facts, objects, phenomena, and events correspond to the vocabulary, while the logical structure of chemistry corresponds to the grammar.

The backbone of the logical structure is the classification of molecules according to the various types of isomerism.[96] Isomers contain the same number and type of atoms and accordingly have the same empirical formula. The term "isomer" can be applied to molecules as well as to substances. From the existence of distinct isomers A. von Humboldt inferred that molecules must have an intrinsic structure. Otherwise substances with the same element composition would not be distinguishable.[97]

Those molecular systems that consist of the same collection of atoms—or more exactly, the same collection of atomic cores (i.e., nucleus plus electrons of the inner shells) and valence electrons—are isomeric EMs. The EMs may consist of one or more molecules. This extension of the notion of isomerism from molecules to EMs is one of the conceptual foundations of a mathematical representation of the logical structure of chemistry.[78]

The chemical constitution of molecules corresponds to the atomic neighborhood relations defined by covalent bonds. Isomers that differ by their chemical constitutions are called constitutional isomers.

By their classical definition stereoisomers are molecules with the same chemical constitution but different relative spatial arrangements of their constitutient atoms.[82, 98] This definition is of limited usefulness, since many stereoisomers are nonrigid molecules that can not be adequately represented by rigid geometric models. A generally valid definition of stereoisomers that also accounts for flexible molecules is possible on the basis of the concept of chemical identity described in the next paragraph. This equivalence relation can also account for nonrigid chiral molecules.[73, 98]

Molecules are chemically identical if they interconvert spontaneously under the given observation conditions and belong to the same uniform chemical compound. Two films showing the time-dependent changes of geometry of a molecule and of a chemically identical molecule will not be the same. However, the sets of molecular frames that are shown in the two films are identical. The same set will be obtained if a snapshot of a sufficiently large number of molecules of the corresponding compound is taken.

If molecules with the same chemical constitution are not chemically identical they are stereoisomers.[73, 82, 98] Stereoisomerism as defined thus is a principle of classification in its own right and is subordinate to constitutional isomerism.

3.2. An Algebraic Model of the Logical Structure of Constitutional Chemistry

The logical structure of constitutional chemistry follows from the following statements that were used by J. Dugundji and I. Ugi[78] as the axiomatic foundations of the theory of the be- (*b*ond and *e*lectron) and r-matrices (*r*eaction) published in 1973:

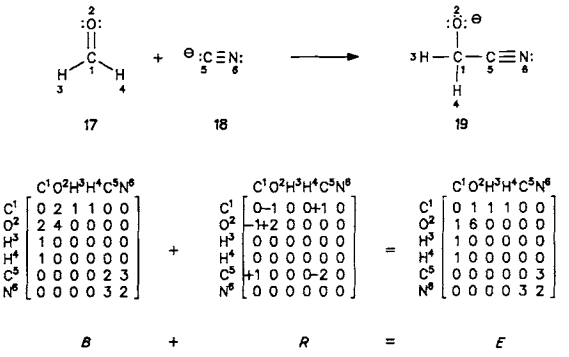
> Molecules consist of atomic cores and valence electrons, held together by covalent bonds. A covalent bond corresponds to a pair of valence electrons that simultaneously belongs to two adjacent atoms. A chemical reaction is the conversion of an EM into an isomeric EM by redistribution of valence electrons. During a chemical reaction the atomic cores and the total number of valence electrons remain the same.

The logical structure of constitutional chemistry can be illustrated by the chemistry of a fixed collection $A = \{A_1, \ldots A_n\}$ of atoms. Since any collection of atoms can serve as A, a model of the logical structure of the chemistry of A is valid for all chemistry.

The chemistry of A is given by the family of isomeric EM(A). EM(A) is an EM which contains every atom of A precisely once.[93]

Within the framework of the DU model, an EM of n atoms is described by its symmetric $n \times n$ be-matrix. The ith row

(and column) of a be-matrix is assigned to the atom A_i ($1 \leq i \leq n$). The entry b_{ij} ($= b_{ji}$, $i \neq j$) of the ith row and the jth column of a be-matrix $\boldsymbol{B}$ is the formal bond order of the covalent bond between the atoms A_i and A_j. The diagonal entry b_{ii} is the number of free valence electrons at the atom A_i. The entries $b_{12} = b_{21} = 2$ of the be-matrix of EM(**17**, **18**) signify that the C atom no. 1 and O atom no. 2 are connected by a covalent double bond, while $b_{22} = 4$ represents four free valence electrons at the O atom no. 2 (Scheme 1).



Scheme 1. A reaction and its matrix description.

The rows and columns of the be-matrices represent the number and position of the electrons (valence schemes) of the relevant atoms. It follows that all rows (columns) of be-matrices of stable EMs must correspond to allowable valence schemes of the represented chemical elements. They are the valence boundary conditions of the be-matrices.

The adjacency and connectivity matrices[99] that are used in chemical documentation differ from the be-matrices in their diagonal entries. These matrices are merely tables of bonds that represent chemical constitution. In contrast, the be-matrices do not only represent single molecules but also multimolecular EMs, and are genuine mathematical objects. The algebra of the be- and r-matrices forms a free additive abelian group.[78]

The redistribution of valence electrons by which the starting materials $EM_{\boldsymbol{B}}$ of chemical reactions are converted into their products $EM_{\boldsymbol{E}}$ is customarily indicated by "electron-pushing arrows", which are represented by r-matrices $\boldsymbol{R}$. The chemical reaction $EM_{\boldsymbol{B}} \rightarrow EM_{\boldsymbol{E}}$ corresponds to the additive transformation of the be-matrices $\boldsymbol{B}$ of the educts by the reaction matrix $\boldsymbol{R}$ into the be-matrix $\boldsymbol{E}$ of the products. $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ is the fundamental equation of the DU model.

The addition of matrices proceeds entry by entry, that is, $b_{ij} + r_{ij} = e_{ij}$. Since there are no formal negative bond orders or negative numbers of valence electrons, the negative entries of $\boldsymbol{R}$ must be matched by positive entries of $\boldsymbol{B}$ of at least equal values.

3.2.1. Chemical Distance

When the n^2 entries of the be-matrices are interpreted as coordinates of points in an n^2-dimensional euclidean space, the algebraic model of the logical structure of chemistry can be visualized as a geometric model.[78, 100] The EMs correspond to be-points and chemical reactions are described by connecting r-vectors. The L_1 length $d(\boldsymbol{B}, \boldsymbol{E})$ of an r-vector is the sum of the absolute values of the differences of the coordinates of the be-points $P(\boldsymbol{B})$ and $P(\boldsymbol{E})$. The L_1 distance ("taxi driver distance")[101] between the two be-points $P(\boldsymbol{B})$ and $P(\boldsymbol{E})$ is called the chemical distance (CD) between $EM_{\boldsymbol{B}}$ and $EM_{\boldsymbol{E}}$. It is twice the number of valence electrons that are redistributed when $EM_{\boldsymbol{B}}$ and $EM_{\boldsymbol{E}}$ are interconverted (see Table 1).[100]

$$d(\boldsymbol{B},\boldsymbol{E}) = \sum_{ij} |b_{ij} - e_{ij}| = \sum_{ij} |r_{ij}| \qquad (1)$$

When the lattice of points representing a family of all isomeric EM with paired valence electrons is viewed from any of its EMs, then the be-points are surfaces of concentric L_1 shells whose radii differ by four units. The radius of the largest sphere is the total number of valence electrons of the underlying collection of atoms. This aesthetic geometric picture of the logical structure of chemistry is a "map" of the

Table 1. Corresponding representations of the logical structure of constitutional chemistry.

Chemical Representation	Algebraic Representation	Geometric Representation
constitutional formula of $EM_{\boldsymbol{B}}(A)$ ($A = \{A_1, \ldots A_n\}$)	symmetric $n \times n$ be-matrix $\boldsymbol{B} = \langle b_{ij} \rangle$ of $EM_{\boldsymbol{B}}(A)$.	be-point $P(\mathbf{B})$ with the coordinates $(b_{11}, \ldots b_{1n}, \ldots b_{n1}, \ldots b_{nn})$ in an n^2-dimensional euclidian space.
up to $n!$ permutations of atomic indices in $EM_{\boldsymbol{B}}(A)$	up to $n!$ equivalent be-matrices $\boldsymbol{B}'$ can be obtained by row (or column) permutations of $\boldsymbol{B}$ according to $\boldsymbol{B}' = \boldsymbol{P} \cdot \boldsymbol{B} \cdot \boldsymbol{P}^{-1}$. $\boldsymbol{P}$ is a $n \times n$ permutation matrix.	b-cluster of the up to $n!$ equivalent be-points $P(\boldsymbol{B}')$ of $EM_{\boldsymbol{B}}(A)$.
chemical reaction $EM_{\boldsymbol{B}}(A) \rightarrow EM_{\boldsymbol{E}}(A)$ by redistribution of valence electrons	the additive transformation of $\boldsymbol{B}$ in $\boldsymbol{E}$. The entries r_{ij} of $\boldsymbol{R}$ represent the redistribution of valence electrons.	r-vectors in the n^2-dimensional space which connects the be-points, $P(\boldsymbol{B})$ and $P(\boldsymbol{E})$, of the participating EMs, $EM_{\boldsymbol{B}}(A)$ and $EM_{\boldsymbol{E}}(A)$.
the number of redistributed valence electrons in a reaction $EM\boldsymbol{B}(A) \rightarrow EM_{\boldsymbol{E}}(A)$ (corresponds to the CD)	$d(\boldsymbol{B}, \boldsymbol{E}) = \Sigma\lvert r_{ij}\rvert$; that is, the sum of the absolute values of the entries of the r-matrix $\boldsymbol{R} = \boldsymbol{E} - \boldsymbol{B}$.	L_1 distance ("taxi driver distance") between the be-points $P(\boldsymbol{B})$ and $P(\boldsymbol{E})$.
the chemistry of A that is given by the family of all $EM(A)$ and the mutual interconversions of $EM(A)$	the groups of be-matrices of $EM(A)$ with their transformations under given boundary conditions.	sets of points which correspond to the elements of the family of all $EM(A)$ and the connecting r-vectors.
chemical reactions preferentially proceed by a minimal redistribution of valence electrons.	the atom-onto-atom correlations $EM_{\boldsymbol{B}}(A)$ and $EM_{\boldsymbol{E}}(A)$ with minimal $d(\boldsymbol{B}, \boldsymbol{E})$ are preferred.	preferred r-vectors for $EM_{\boldsymbol{B}}(A) \rightarrow EM_{\boldsymbol{E}}(A)$ lead from one point of the $\boldsymbol{B}$-cluster to the "nearest" point of the $\boldsymbol{E}$-cluster.

energy minima of the potential energy surface of the given collection of atoms.[102]

The solutions of the basic equation $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ of the DU model correspond to the solutions of a great variety of chemical problems. Accordingly, the DU model can serve as the universal theoretical foundation of computer programs for the deductive solution of constitution-related chemical problems. Furthermore, it is suitable as a basis of strictly formal, transparent procedures for the selection of chemically meaningful solutions from very large sets of conceivable solutions.[21] Here the Principle of Minimal Chemical Distance[78, 103-107] plays an important role. This principle is a quantitative version of the classical qualitative principle of minimum structure change:[108] The interconversion of isomeric EM by chemical reactions preferentially proceeds by redistribution of the minimum number of valence electrons.

The minimal CD corresponds to one or more atom-to-atom maps of the EM, because the CD between isomeric EMs depends on the correlation of their atoms. In a broad sense this is also true for sequences of reactions, that is, the CD of reaction sequences is rarely more than four units above the minimum of CD.

The principle of minimal CD can not only select the "shortest" pathways of chemical reactions but can also help to determine reactive centers, that is, those atoms whose covalent bonds and free valence electrons are immediately affected by the reaction. The complete set of reactive centers is called the core of the reaction. The bonds that are broken or made during a reaction are determined by the atom-to-atom maps of the interconnected EMs.

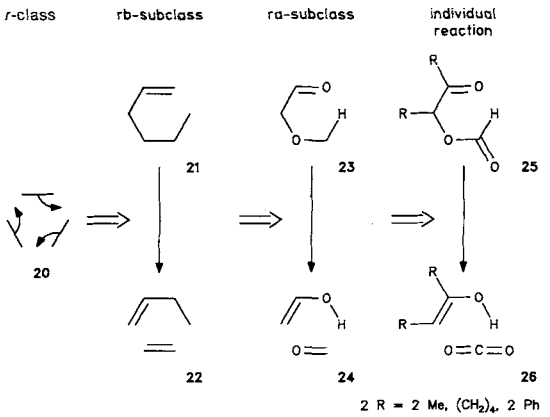
Various distance functions and measures of similarity have been published for molecular systems. In essence, they all rely on the fact that chemical reactions correspond to vectors (see Table 1) and have metric properties.[78] Other examples, besides CD,[100] are reaction distance that was defined differently by Kvasnička et al.[80, 81, 109] and by Hendrickson et al.,[87] synthetic proximity of Johnson et al.,[110] and adjacency distance, the sum of the absolute values of the t-matrices (differences of adjacency matrices) of Fontain.[111]

3.2.2. The Hierarchic Classification of Chemical Reactions

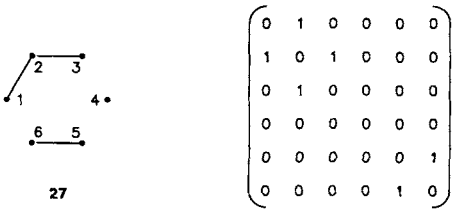
A hierarchic classification of chemical reactions by similarity classes[100] follows from the principle of minimal CD (see Scheme 2).

The position of a chemical reaction, for example **25** → **26**, in this hierarchic classification system is determined by stepwise neglect of its characteristic features. First, the reaction is reduced to its core of reactive centers by omitting all substructures that do not directly participate in the reaction. The result is the ra-subclass of the reaction, whose members have the same reaction core but different invariant molecular substructures.

The next step of abstraction is to neglect the differences between the chemical elements of the atoms of the core. This leads to the rb-subclasses (basic reactions). Their members are characterized by the same arrangement of covalent bonds between the atoms of the core (see **27**) and can be represented by the so-called intact be-matrix (Scheme 3).[79]



Scheme 2. The procedure for the hierarchic classification of chemical reactions. The individual reaction **25** → **26** is reduced to the electron redistribution scheme **20**.



Scheme 3. Intact be-matrix of the rb-subclass **21** → **22**.

Reactions with the same electron redistribution scheme and therefore the same irreducible r-matrix[100, 112, 113] belong to the same r-class. The r-matrix is converted into its irreducible r-matrix by removing all rows and columns that only contain zeros. The rows (and columns) of the irreducible r-matrix belong to the reactive centers of the reactants. The hierarchic classification of chemical reactions ends with the CD-classes.[79, 100, 112-114] Reactions belong to the same CD-class if the same number of valence electrons is redistributed during the course of the reaction and thus the same CD is covered.

Scheme 2[100, 114] illustrates the hierarchic classification of chemical reactions for the example of the extrusion reaction **21** → **22**. The rb-subclass is represented by the bonding scheme **27** and the corresponding 6×6 intact be-matrix of Scheme 3.

The classification described above does not only open new ways of documenting chemical reactions,[79, 100, 112, 113] but also plays an important role in the computer program IGOR (*I*ntermediate *G*eneration of *O*rganic *R*eactions).[114-116] It is used to select computer-generated chemical reactions and to assess their degree of novelty.[88, 117] The degree of novelty is determined by assigning classes and subclasses of the reaction in question and ascertaining the level of hierarchy above which no published reactions can be found. The higher this level is, the higher the degree of novelty of the reaction.[88, 113, 114, 117] The reaction **25** → **26** of Scheme 2 was discovered with computer-assistance. It belongs to the extrusion reactions **21** → **22**[114] (see Section 4.1.4.3) that are characterized by the bond system **27**, and that is contained in r-class **20** with 12 rb-subclasses.

No representatives of ra-class **23** → **24** are known. Accordingly the reaction **25** → **26**[114, 118] is novel up to the level of the ra-subclasses.

Almost twenty years ago, Brownscombe and Stevens[83] reported a computer program that was capable of generating any elements of the rb-subclass of extrusion reactions by permuting the chemical elements of the core of the reaction. Reaction **25** → **26** could have already been found with that computer program.

4. The Munich project

The applications of the DU model are not restricted to the design of syntheses. After the CICLOPS study[53] was terminated in 1974 (see Section 2.4) a comprehensive plan, the Munich project, was devised. Its aim was to extend and improve the DU model and to exploit it in as many ways as possible as a formal basis for the computer-assisted solution of chemical problems.[119] During its implementation, the Munich project was changed so much that only its basic ideas are left. However, it initiated the development of computer programs for the solution of chemical problems on a broad front. The Munich project, whose goals have nearly been reached, consists of the following subprojects:

a) Improvement and extension of the DU model, in particular for EMs with multicenter bond systems of delocalized electrons.

b) Development of a computer-oriented mathematical theory of stereochemistry that is able to account for the stereochemical aspect of chemical reactions.

c) Development of a software infrastructure for programs according to d).

d) Development of computer programs for the deductive solution of chemical problems. These programs should not only use the DU model for the generation of solutions of problems, but also for screening and selecting these solutions.

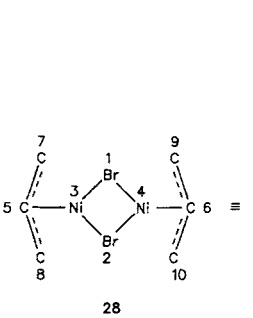
e) Testing and use of programs according to d) and improvement of these programs on the basis of gathered experience.

f) Experimental realization and verification of the results of e).

Progress on individual steps of the Munich project has been described together, as far as possible, with details of the computer programs, which often included the source codes, in order to ensure reproducibility. Here we give a survey of the subprojects of the Munich project, in the order a)–f), covering its historical development and current status.

4.1. Extension of the DU model

For a long time neither we nor others succeeded in extending the DU model as planned in a). Recently, however, a formalism was found through which EMs with multicenter bonds and delocalized valence electron systems (DE systems)[95] can be represented. The xbe-matrix (extended be-matrix) of an EM corresponds to its be-matrix which is extended by additional rows and columns. The be-matrix with n rows and columns refers to the localized covalent bonds whose formal bond orders and free valence electrons can be assigned to the individual atoms. The additional rows and columns with the indices $n + k$ belong to DE systems, for example, multicenter bond systems or delocalized π-electron systems. The off-diagonal entries $b_{i,\,n+k} = b_{n+k,\,i} = 1$ of the kth additional rows and columns indicate that the atom A_i participates in the $(k - n)$th DE system. The diagonal entries $b_{n+k,\,n+k}$ are the numbers of valence electrons that belong to the $(k - n)$th DE system. The xbe-matrix of π-allyl nickel bromide (**28**)[120] may serve as an example in Scheme 4; for the sake of simplicity, the CH bonds and the corresponding entries are omitted.



4	0	1	1	0	0	0	0	0	0	0	0
0	4	1	1	0	0	0	0	0	0	0	0
1	1	8	0	0	0	0	0	0	0	1	0
1	1	0	8	0	0	0	0	0	0	0	1
0	0	0	0	0	0	1	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	1	1	1	0
0	0	0	0	1	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	0	0	0	0	1
0	0	1	0	1	0	1	1	0	0	4	0
0	0	0	1	0	1	0	0	1	1	0	4

Scheme 4. Structural formula of **28** showing numbering and the xbe-matrix representing it. The row and column of the DE systems are indicated by dashed lines.

In analogy to the original DU model, chemical reactions are represented by addition of xr-matrices[95] to the xbe-matrices. The 18 theorems of the DU model are equally valid for the algebra of the xbe- and the xr-matrices, which is named the model of the chemistry of *d*elocalized *e*lectron (DE) systems.

In order to adapt this system to computers, a new data structure was introduced. At the moment, a reaction generator for reactions in which DE systems may also participate is being developed.[121] Here the particular properties of the DE systems must be accounted for by formulation of suitable boundary conditions.

4.2. Computer-Oriented Formalization of Stereochemistry—the Theory of the Chemical Identity Group and Accumulations

Stereochemistry is the science of the spatial structure of molecules and its observable consequences. The differences in the formation, reactions, and properties of stereoisomers are the central issues of stereochemistry. Stereoselective syntheses belong to the most attractive current topics of organic chemistry.[1] Therefore, the stereochemical aspect of computer-assisted chemistry is of particular interest. The large numbers of combinatorial possibilities create huge amounts of data in solving stereochemical problems, which can only be handled if the mathematical structures behind the problems are fully recognized and exploited. It is advantageous to introduce some suitable new concepts and approaches for this purpose.

4.2.1. *Traditional Approaches*

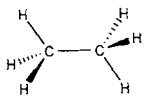
Corey et al.[122] supplemented the transforms of the individual reactions with detailed stereochemical information. This approach is very cumbersome; just the treatment of the stereochemistry of six-membered rings requires a large program of its own. Wipke and Dyott[123] proposed to solve stereochemical problems by assuming sufficiently rigid reactants that are determined by a procedure based on steric bulk. Hanessian et al.[124] assign chiral starting materials to chiral target molecules through their program CHIRON. In the early stages of the Munich project the stereochemical features of EMs with stereogenic subunits with coordination numbers ≤ 4 were represented by parity vectors.[125]

None of these approaches leads to a generally applicable method for the computer-assisted solution of stereochemical problems. The comprehensive computer-oriented formal treatment of stereochemistry requires a fundamentally different approach.

4.2.2. *A Nongeometric Alternative*

In traditional stereochemistry the facts and phenomena are interpreted and predicted on the basis of rigid geometric models. Thus, stereochemistry is reduced to elementary geometry. Molecular geometries and point group symmetries are drastically overemphasized. Widely applicable computer-assisted methods for the solution of stereochemical problems, however, require a theory that treats the static and dynamic aspects in a uniform way.

Many molecules are not rigid. Since their shapes vary with time, they are in general not adequately represented by geometric models. For instance, at 20 °C the methyl groups in ethane **29** rotate around the C–C bond with a frequency of $10^9\ s^{-1}$.[126, 127] This molecule can not be represented by a rigid geometric model.



29

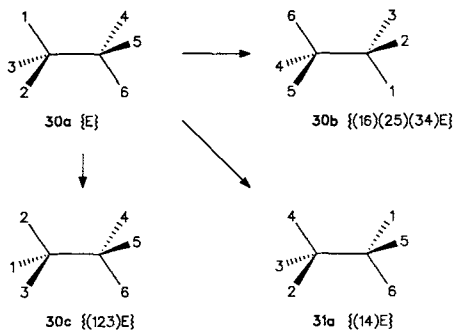
The first prerequisite for a comprehensive and uniform theoretical treatment of stereochemistry is that the traditional definition of stereoisomers should be replaced by a definition that is also able to account for flexible stereoisomers. This definition should be based on the notion of chemical identity (see Section 3.1) and should not refer explicitly to any geometry.

Permutation isomerism, as defined in 1970,[76] and the theory of the chemical identity group (CIG theory), which was published in 1984,[82] have played an important role in the formalization of stereochemistry. The CIG theory is a group theoretical model of the logical structure of stereochemistry that not only accounts for the energy-related and geometric properties of the molecules, but also includes the valid observation conditions. It avoids explicit reference to rigid geometric models and their point group symmetries. In contrast to the traditional applications of group theory (see Section 2.3.2), the CIG theory does not only cover molecular objects, but also stereochemical relations and processes between molecular systems.

4.2.3. *The Basic Ideas of the CIG Theory*

A molecule can be dissected in the imagination into a molecular skeleton and a set of ligands. The distinct molecules that differ by positioning of the ligands at the skeletal sites, are the permutation isomers. The set of all permutation isomers that can be obtained from a reference isomer is called a family of permutation isomers.

Let m be a snapshot of a molecule. The conceptual separation of m into a molecular skeleton and a set L of ligands yields a reference model. The reference model belongs to an isomer X which serves as the reference isomer. In contrast to a molecule, a model in the present sense has a fixed spatial orientation. The chemical identity of a molecule is independent of its spatial orientation. There are permutations of ligands that convert one model into another that belongs to the same molecule (see Scheme 5). The chemical identity of X is preserved by all permutations of ligands through which models of X are converted into other models of X, for example, permutations that may be interpreted as rotations of the molecule as a whole.



Scheme 5. Ligand permutations of **30a**, the reference model of an arbitrarily substituted ethane molecule.

If all ligands in the set L are distinguishable, the identity-preserving permutations generally form a group $S(E)$. This is the CIG of the reference isomer X that is represented by the model E. $S(E)$ is a subgroup of the symmetric permutation group $Sym(L)$ of L. The distinct permutation isomers of X are represented by the left cosets $\lambda S(E)$ ($\lambda \in Sym(L)$) of $S(E)$ in $Sym(L)$.[*]

For a molecule with n ligands $Sym(L)$ consists of $n!$ permutations. This is the cardinality of the family of permuted models $P(E)$. The maximum number of chemically distinct permutation isomers is $n!/|S(E)|$, because all cosets have the same cardinality $|S(E)|$. The left cosets of the CIG in $Sym(L)$ or any of their elements can serve as nomenclature descriptors for the permutation isomers.[76] The left coset of a CIG corresponds to the permutation isomer of the reference isomer, whose models are obtained from the reference model by ligand permutations from the CIG, followed by a permutation from the considered left coset.

[*] For readers who are less familiar with mathematical concepts and symbols there is an appendix with brief explanations.

Permutations are customarily written as cycles,[128] for example, (123); the vector [123] of numbers is permuted into [231] by mapping $1 \rightarrow 2$; $2 \rightarrow 3$; $3 \rightarrow 1$. The combination (12)(123) of the permutations (12) and (123) corresponds to the action of (12) on the result of (123), that is (12)(123) transforms [123] into [132] and is thus equivalent to (23).

For example, the reference model **30a** of the ethane derivative **30** is converted into the model **30b** by the permutation (16)(25)(34) of the ligands; this corresponds to a rotation by 180° of **30a** around an axis that is perpendicular to the plane of the paper. Note that this rotation of the model **30a** does not superimpose the ethane skeleton onto itself, according to D_{3d} point group symmetry, since the different ligands destroy the symmetry of the skeleton (see Chapter 3 of ref. [82]).

Since the chemical identity of a molecule is independent of its spatial orientation, **30a** and **30b** are chemically identical. The model **30c** also belongs to **30** if the internal rotation around the C–C axis belongs to the spontaneous intramolecular motions, under the given observation conditions. This permutation can be represented by (123).

The CIG (2) contains all ligand permutations that preserve the chemical identity of **30**. The ligand permutations of

$$S(30a) = \{(), (123), (132), (456), (465), (123)(456), (123)(465), (132)(456), (132)(465), (14)(26)(35), (15)(24)(36), (16)(25)(34), (142635), (143526), (152436), (153624), (162534), (163425)\} \quad (2)$$

$S(30a)$ generate further models of $S(30a) \cdot$ **30a** from any model that belongs to $S(30a) \cdot$ **30a**. The ligand permutation (14) converts **30a** into **31a** which represents a permutation isomer of **30**. Likewise the ligand permutation (14) converts any model of **30** into a model of **31**. They can also be obtained by any other permutation of $S(30a)$, followed by any permutation of (14) $\cdot$ $S(30a)$. **30** and **31** are distinct permutation isomers if the ligands 1 and 4 are distinguishable.

In the following sections we will describe set-valued maps (SVM) and accumulations. These operations are very useful for representing permutation isomers with indistinguishable ligands and isomerizations which convert permutation isomers into each other.

4.2.4. Implementation and Applications of the CIG Theory

Within the framework of the CIG theory stereochemical equivalence relations are represented by surjective set-valued maps (SVM) of permutations onto sets of permutations that may be unions of cosets or Wigner subclasses[129] of subgroups of $Sym(L)$. The direct consideration of the dynamic aspect of stereochemistry is the most important contribution of the CIG theory to the mathematical treatment of stereochemistry.

The SVM were introduced into chemistry as a part of the CIG theory. They are a most versatile device for the solution of stereochemical problems that can be formulated as equivalence relations between permutation isomers. Such equivalence relations correspond to equivalence spaces that are generated by SVM.

Let M be a set which is partitioned into equivalence classes in two distinct ways (Fig. 1). The SVM of a class B from the partition $\mathscr{B}$ yields the union of those equivalence classes

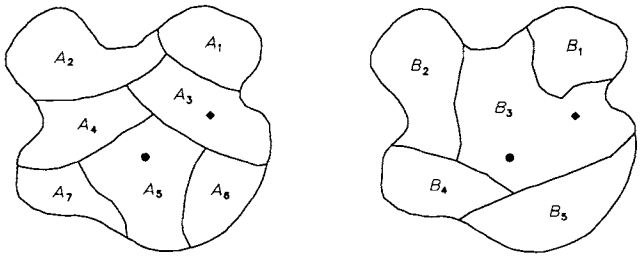


Fig. 1. Two partitions $\mathscr{A}$ and $\mathscr{B}$ of set M.

from partition $\mathscr{A}$ whose intersection with B is non-empty [Eq. (3)].

$$\mathrm{SVM}(\mathscr{A}, B) = \bigcup_{A \cap B \neq \{\}, A \in \mathscr{A}} A \quad (3)$$

For example, ◆ is an element of class A_3 and ● is an element of A_5. These elements also belong to class $B_3 \in \mathscr{B}$. All elements of class A_3 are equivalent to ◆. The elements of class A_5 are equivalent to ●. The elements ◆ and ● are also equivalent since they both belong to class B_3.

Equivalence relations are reflexive, symmetric, and transitive (p. 14 of ref. [128]). The equivalence classes A_3 and A_5 (and any further equivalence classes of $\mathscr{A}$ that intersect B_3) form a single equivalence class. We have the SVM given by (4).

$$(\mathscr{A}, B_3) = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6 \quad (4)$$

In recent years the CIG theory has been modified and extended. It is now more suitable as a foundation for the computer-assisted solution of a great variety of stereochemical problems. In particular, (equivalence) accumulations[130, 131] represent significant progress in practical computer-assisted stereochemistry.

The CIG and its cosets form a space of equivalence classes in $Sym(L)$. Since, by definition, cosets are disjoint, each coset can be interpreted as an equivalence class that represents a permutation isomer. One permutation is sufficient to represent the whole family of permutation isomers. Thus, the amount of data to be processed for stereochemical problems is reduced to a small fraction.

We consider a equivalence space A of molecular models and some independent information I about further equivalences of models, for example, from equivalences of ligands or interconvertibilities of models. According to the additional information I, there exists an equivalence space U which differs from A in that some of the equivalence classes of A merge into a single equivalence class. The information I is an equivalence relation corresponding to group theoretical relations between the individual models. Within the framework of the CIG theory, I can be expressed by a few generating elements that follow directly from the chemical properties of the molecules. An example for such information is the coset space of the CIG $S(30a)$ in $Sym(L)$ mentioned in Section 4.2.3: For each coset $\lambda S(30a)$, $\lambda S(30a) \cdot$ **30a** is an equivalence class of chemically identical models. All pairs of models in the equivalence class are elements of the equivalence relation I, that is, for all models $m, n \in \lambda S(30a) \cdot$ **30a** we have $(m, n) \in I$.

For a given set of models M that is partitioned according to the equivalence space A, and the information I, the accu-

mulation Acc(A, I) is defined according to (5)–(7) (+ represents the transitive closure[132] of a relation):

$$\mathrm{Acc}(A,I) := \{[m]_R \mid m \in M\} \tag{5}$$

$$\text{where} \quad R = (I \cup a \times a \mid a \in A)^+ \tag{6}$$

$$\text{and} \quad [m]_R = \{n \in M \mid (m,n) \in R\} \tag{7}$$

Ligands are considered to be equivalent, if their chemical interpretations are the same and their permutation does not affect the chemical identity of molecules. The permutations of equivalent ligands form the ligand equivalence group Σ. The right cosets of Σ in $Sym(L)$ represent those models that are potentially distinct if all ligands were distinguishable, but are chemically identical because of the equivalence of some of the ligands.

A ligand permutation that converts a reference model into a model of a distinct permutation isomer, followed by a ligand permutation from Σ, corresponds to a ligand permutation that belongs to a right coset of Σ.

Thus, the right coset space of Σ in $Sym(L)$ also contains the following information: the models λE in each equivalence class $\Sigma\mu$ are mutually chemically identical.

If this equivalence class $\Sigma\mu$ intersects with the equivalence classes $\lambda S(E)$ and $\sigma S(E)$ of the coset space of the CIG of E, permutations ρ and τ exist such that $\rho \in \Sigma\mu \cap \lambda S(E)$ and $\tau \in \Sigma\mu \cap \sigma S(E)$. Since $\rho \in \lambda S(E)$, the model ρE is chemically identical to all models $\lambda S(E)E$, and τE is chemically identical to all models $\sigma S(E)E$, because $\tau \in \sigma S(E)$. On the other hand, ρE and τE are chemically identical, because ρ and τ belong to the same right coset $\Sigma\mu$, that is the molecules that are represented by ρE and τE are chemically identical, because models ρE and τE differ only by a permutation of equivalent ligands.

Because as an equivalence relation chemical identity is reflexive, symmetrical, and transitive, all models belonging to $\lambda S(E)E$ are also chemically identical to τE (because ρE is chemically identical to all of those models). The right coset $\Sigma\mu$ contains the information that all models $\Sigma\mu E$ are chemically identical. Thus, the representatives $\lambda S(E)E$ and $\sigma S(E)E$ (and all further equivalence classes $\xi S(E)E$, with $\xi S(E) \cap \Sigma\mu \neq \{\}$) of the permutation isomers that were initially considered to be distinguishable, merge into the class of permuted models described by (8). This class models represents a single permutation isomer. Since such SVM do not take into account the full right coset space of Σ, in some cases the results may be incomplete. A coset $\lambda S(E)$ can exist whose intersection with $\Sigma\mu$ as well as with a further right coset $\Sigma\omega$ is non-empty. Then the SVM of $\Sigma\mu E$ does not yield the complete set of chemically identical models: The expression (9) states that the $\lambda S(E)$ models are chemically identical with all models from (10) and also with all models from (11).

$$\bigcup_{\xi S(E) \cap \Sigma\mu \neq \{\}} \xi S(E) \cdot E = \mathrm{SVM}(\{\lambda S(E) \cdot E \mid \lambda \in Sym(L)\}, \Sigma\mu \cdot E) \tag{8}$$

$$\lambda S(E) \cap \Sigma\mu \neq \{\} \wedge \lambda S(E) \cap \Sigma\omega \neq \{\} \tag{9}$$

$$\bigcup_{\xi S(E) \cap \Sigma\mu \neq \{\}} \xi S(E) \cdot E \tag{10}$$

$$\bigcup_{\xi S(E) \cap \Sigma\omega \neq \{\}} \xi S(E) \cdot E \tag{11}$$

Accordingly, all models that belong to any one of the above molecules are chemically identical. Such problems are avoided by the accumulation (12) by forming the transitive envelope (13).

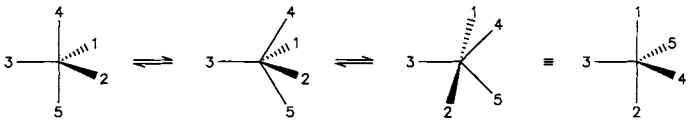
$$\mathrm{Acc}(\{\lambda S(E) \cdot E \mid \lambda \in Sym(L)\}, \{\Sigma\mu \cdot E \times \Sigma\mu \cdot E \mid \Sigma\mu \cdot E \in \{\Sigma\mu \cdot E \mid \mu \in Sym(L)\}\}) \tag{12}$$

$$R = (\{\Sigma\mu \cdot E \times \Sigma\mu \cdot E \mid \Sigma\mu \cdot E \in \{\Sigma\mu \cdot E \mid \mu \in Sym(L)\}\} \cup \{\lambda S(E) \cdot E \times \lambda S(E) \cdot E \mid \lambda S(E) \cdot E \in \{\lambda S(E) \cdot E \mid \lambda \in Sym(L)\}\})^+ \tag{13}$$

Thus, finally, a family of permutation isomers with a ligand set containing equivalent ligands is represented as a union of the double cosets $\Sigma\lambda S(E)$ ($\lambda \in Sym(L)$). In these cases the customary methods for the enumeration and classification of isomers[70, 72, 74] must be used with caution. In contrast to these methods, the CIG theory also indicates which isomers are involved, which isomers merge into a single isomer due to ligand equivalences, and which chiral isomers in turn yield chiral isomers.

The accumulation is a generalization of the SVM. It takes all of the available information I into account to determine the distinct permutation isomers. In this case this is the whole right coset space of Σ. Besides equivalence of ligands there are other reasons why permutation isomers that were initially regarded as distinguishable are to be classified as chemically identical.

Isomerizations of a permutation isomer X into a permutation isomer of the same family that proceed via a permutation isomer Y of a different family (see Scheme 6) are of particular interest. Let E and F be the reference models of X



Scheme 6. The reference reaction for the generation of graphs of the Berry pseudorotation and the turnstile rotation.

and Y. The reaction $E \rightleftharpoons F$ is used as the reference reaction, that is, it is assumed that any mutual interconversion between λE and λF (for all $\lambda \in Sym(L)$) is possible. When the permutation isomers of X are considered, the left coset spaces of the CIG $S(E)$ and $S(F)$ are generated. The left coset space of $S(E)$ corresponds to the distinct permutation isomers of X, as long as no isomerization takes place. The left coset space of $S(F)$ in $Sym(L)$ yields the information on the permutation isomers of X that are interconvertible by isomerization and that can thus be viewed as chemically identical. The procedure for determining the distinct permutation isomers of X under the given isomerization conditions corresponds to the procedure for the investigation of the equivalence classes of $S(E)$ when some of the ligands are equivalent.

The reaction schemes (or reaction graphs) that represent all interconversions of permutation isomers are obtained by the SVM of the left cosets $\lambda S(E)$ and $\lambda S(F)$ of the two CIGs $S(E)$ and $S(F)$.[82]

The generation of the graph of all Berry pseudorotations[133] and turnstile rotations[82, 134, 135] of the family **32** of permutation isomers from the reference reaction **32a** $\rightleftharpoons$ **32b** by SVM of the left cosets of $S(32a)$ and $S(32b)$ may

serve as an illustration (Scheme 6). For this, each of the individual members of the family **32** are represented by a left coset of $S(32a)$ and also by a left coset of $S(32b)$. A SVM of the two left coset spaces is established.

The isomerization processes involving two or more families of permutation isomers like **32** and **33** can be represented by (equivalence) accumulations, if permutation isomers with the same set of ligands participate in the reference reaction. When **32a**$\rightleftharpoons$**33a** is the reference reaction, a graph of the Berry pseudorotation results, in which the reactants **32** are connected by transition states **33** (see p. 132 of ref. [82]).

Ligand permutations that belong to the same Wigner subclass[129] of the relevant CIG correspond to permutation isomerizations with analogous intramolecular motions, that is the same reorganization mechanism.[96, 134] The union of left cosets of a CIG that intersect with a Wigner subclass of this CIG represent a Musher mode.[82, 136, 137] The latter corresponds to those permutation isomers that are formed directly from a reference isomer by a given reorganization mechanism (or an equivalent mechanism). Berry pseudorotations and turnstile rotations yield equivalent results and thus belong to the same Musher mode. They are represented by the same isomerization graph (see p. 132 of ref. [82]).

Since the result of an (equivalence) accumulation is in turn a space of equivalence classes the procedure of accumulations can be restarted with further information (for example additional ligand equivalencies or isomerization processes). Finally an equivalence class space is obtained whose cardinality yields information about the number of distinct permutation isomers, and, through inclusion relations, also about chirality and hyperchirality.[138]

Since the accumulation algorithm is applied in steps, isomerization graphs can be obtained. Thus, accumulations are well suited for the analysis and description of stereoselective reactions. If the ligands 1 and 2 as well as 3 and 4 of **32** are chemically equivalent, then the equivalence space (14) is subjected to a further accumulation by the right coset space of the ligand equivalence group (15). The result is the new equivalence group (16).

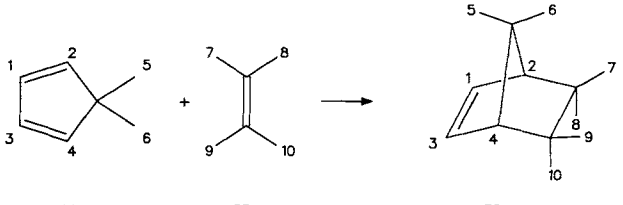
$$A = \mathrm{Acc}(\{\lambda S(32a) \cdot \mathbf{32a} \,|\, \lambda S(32b) \cdot \mathbf{32b} \in \{\lambda S(32b) \cdot \mathbf{32b} \,|\, \lambda \in Sym(L)\}\}) \tag{14}$$

$$\Sigma = \{(), (12), (34), (12)(34)\} \tag{15}$$

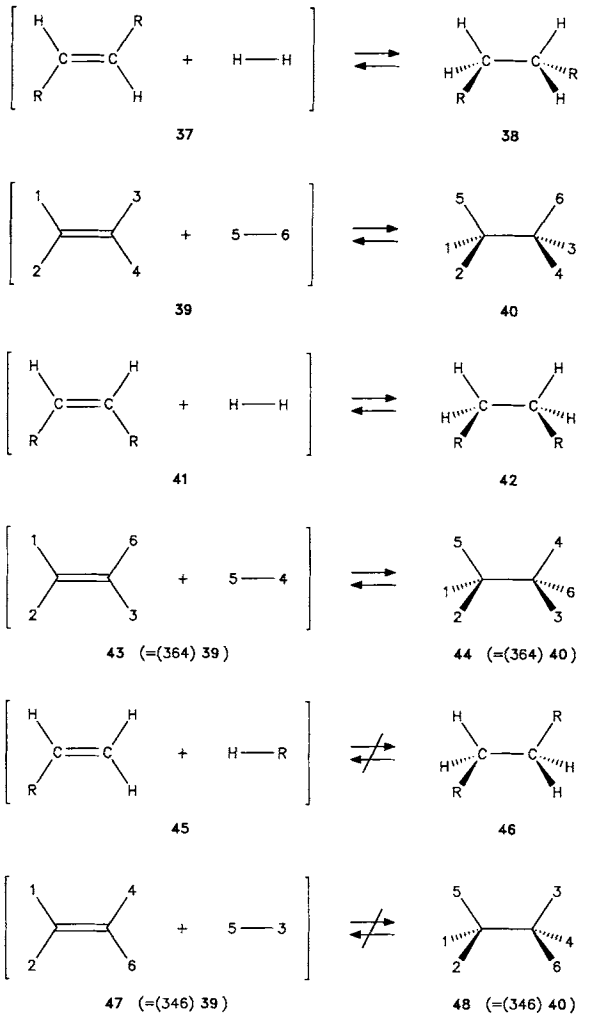
$$B = \mathrm{Acc}(A, \{\Sigma\mu \cdot E \times \Sigma\mu \cdot E \,|\, \Sigma\mu \cdot E \in \{\Sigma\mu \cdot E \,|\, \mu \in Sym(L)\}\}) \tag{16}$$

The graph of Berry pseudorotations/turnstile rotations changes accordingly (see p. 133 of ref. [82]).

Analogously it is also possible to determine which stereoisomers can be formed by a ligand-preserving reaction that affects the chemical constitution of the reactants. The formation of one of the stereoisomers can serve as the reference reaction, and the change in chemical constitution is treated as a transition from one family of permutation isomers to another. The representation of all Diels–Alder reactions that **34** and **35** may undergo in analogy to the reference reaction of Scheme 7 is an example.[98]



34 **35** **36**



37 **38** **39** **40** **41** **42** **43** (=(364) **39**) **44** (=(364) **40**) **45** **46** **47** (=(346) **39**) **48** (=(346) **40**)

Scheme 7. The *cis–trans* isomerization of alkenes.

The applications of accumulations described above emphasize the wide variety of problems that can be treated by SVM and reaction schemes.[94, 95, 97, 124]

In addition, even the isomerizations which interconvert the members of two families of permutation isomers X and Y (with reference models E and F) that have different sets of ligands can be treated by accumulations. In such cases a reference isomerization $E \rightleftharpoons F$ does not necessarily imply that all processes $\lambda E \rightleftharpoons \lambda F$ must take place. The *cis–trans* isomerizations of an alkene that proceed via an alkane may serve as a simple example (Scheme 8).

It is assumed that internal rotations around the C–C axis of the alkane take place freely under the given observation conditions. It follows that the models (346)**40** (= **48**), (364)**40** (= **44**) and **40**, are interconvertible, that is, "equivalent" in the present sense.

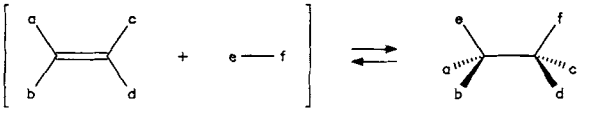
The CIG $S(40)$ contains the permutations (346) and (364). They belong to different left cosets of $S(39)$. Accordingly, the intersections $S(39) \cap (346)S(39)$, $S(39) \cap (364)S(39)$, and $(346)S(39) \cap (364)S(39)$ are empty. However, $S(40)$ intersects with each of these left cosets of $S(39)$: $() \in S(40) \cap S(39)$, $(346) \in S(40) \cap (346)S(39)$ and $(364) \in S(40) \cap (364)S(39)$.

It follows that **39**, (346)**39** (= **47**) and (364)**39** (= **43**) are chemically identical. This is certainly true for **39** and (364)**39**, because **37** can be converted into **40** by the hydration route **39** → **40** → (364)**40** → (364)**39**, followed by an internal rotation and subsequent formation of **41** by dehydrogenation.

The *cis–trans* isomers that were initially rated as distinguishable now become interconvertible and thus equivalent (according to the left coset space of $S(39)$ the internal rotation about the C–C bond is not allowable in this case).

In contrast to **39** ⇌ **40** and (364)**39** ⇌ (364)**40** the reaction (346)**39** ⇌ (346)**40** does not represent a hydrogenation/dehydrogenation. Therefore, even though **46** is equivalent to **38** and **42** it cannot be converted into **45**, and (346)**39** ⇌ (346)**40** does not correspond to a model reaction of a hydrogenation/dehydrogenation. In this case a straightforward determination of the intersections is not suitable for establishing the mutual interconvertibilities.

In the description of intermolecular relations chemical reality can be accounted for by so-called filters, which "filter" the feasible interconversions from the set of conceivable or theoretically possible interconversions $\lambda E \rightleftharpoons \lambda F (\lambda \in Sym(L))$. These "allowable" individual isomerizations $\lambda E \rightleftharpoons \lambda F$ are distinguished by characteristic placements of the ligands at the skeletal sites, which are specified as follows: In a reference model M ligand 1 is located at the skeletal site a(M), ligand 2 at the skeletal site b(M), etc. In the hydrogenation/dehydrogenation of Scheme 7, the assignment of the skeletal sites is shown in Scheme 8, in which **39a** = E and **40a** = F are the reference models.



39a (E) **40a** (F)

Scheme 8. Labeling of skeletal sites for the isomerization of Scheme 7.

The *syn* mechanism of hydrogenation/dehydrogenation requires that the skeletal sites e(F) and f(F), and also e(E) and f(E) are occupied by H atoms. The remaining skeletal sites may, in principle, bear any ligands. The only condition is that the corresponding skeletal sites, for example, a(E)/a(F), b(E)/b(F), must bear the same ligands. The filter for this isomerization is formulated in (17).

$$\Phi := \{a(E) = a(F) \wedge b(E) = b(F) \wedge c(E) = c(F) \wedge d(E) = d(F) \wedge e(E) = e(F) = H \wedge f(E) = f(F) = H\} \quad (17)$$

The statement ⟨skeletal site⟩ = ⟨skeletal site⟩ expresses that the two specified skeletal sites must carry the same ligands (e.g., a(E) = a(F)). The condition ⟨skeletal site⟩ = ⟨skeletal site⟩ = ⟨ligand⟩ states that a specific ligand must be present (e.g., e(E) = e(F) = H). Thus, the individual isomerizations $E \rightleftharpoons F$ and $(364)E \rightleftharpoons (364)F$ meet the criterion of the above filter, whereas $(346)E \rightleftharpoons (346)F$ does not.

There are also isomerizations that interconnect members of families which do not have the same set of ligands. Let E and F be the reference models of the participating compounds X and Y. The ligand set $L(E) = \{1, 2, 3, 4\}$ belongs to model E while $L(F) = \{1, 2, 3, 4, 5\}$ belongs to F. Evidently, the permutation (15) only acts on F, because E does not have a fifth ligand. Furthermore, ligand 4 may represent different chemical residues in E and F (e.g., 4(E) = Me; 4(F) = Et). Therefore, the chemical meaning of the permutation may differ depending on the model to which it is applied. In such cases interconvertible isomers of X and Y cannot be determined by intersecting the coset spaces of $S(E)$ and $S(F)$. Filters must be employed to pick the allowable isomerizations. This is an advantage of accumulations[130, 131] over SVM and reaction schemes.[82, 100] The automated elucidation of relations (e.g. interconvertibility) between permutation isomers (which may belong to different families) merely according to the non-empty intersections of equivalence classes restricts the range of applications of reaction schemes and SVM significantly. This restriction is overcome by accumulations which exploit any relation between models in families of permutation isomers.

For four years, we have been developing computer programs to solve stereochemical problems on the basis of the present mathematical concepts.[130, 131]

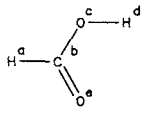
4.3. The Software Infrastructure

A suitable software infrastructure needed to be created for the computer programs for the deductive solution of chemical problems on the basis of the DU model and the CIG theory.

4.3.1. Canonical Indexing of Atoms

The algorithm and computer program CANON[139, 140] were developed for internal documentation. The algorithm CANON refers to the atomic numbers and coordination numbers of atoms as well as their covalently bound neighbors.

The principle of CANON is illustrated by a simple example (Scheme 9).



49

1. Atomic indices:	a: 3	b: 2	c: 1	d: 3	e: 1
1. Atomic descriptors:	a: 3.2	b: 2.113	c: 1.23	d: 3.1	e: 1.2
2. Atomic indices:	a: 5	b: 3	c: 2	d: 4	e: 1
2. Atomic descriptors:	a: 5.3	b: 3.125	c: 2.34	d: 4.2	e: 1.3

Scheme 9. An example to illustrate the principle of CANON.

CANON begins by arbitrarily labeling the atoms. In the example of **49**, the letters a–e serve as these arbitrary labels. The first atomic indices are assigned by starting with 1 for the atom with the highest atomic number and proceding in order with decreasing atomic numbers (O:1; C:2; H:3). The first atomic descriptors are formed from the first atomic in-

dices of the considered atom and in numerical order the indices of its covalently bound immediate neighbors (e.g., b: 2.113; c: 1.23; e: 1.2). Subsequently, the order of the first atomic descriptors taken as decimal numbers (lexicographic order) is used to determine the second atomic indices (e.g., b: 3; c:2; e: 1). The second atomic descriptors are formed analogously to the first atomic descriptors from the second atomic indices. Since the lexicographic order of the second atomic descriptors already corresponds to the second atomic indices, the latter are the final atomic indices according to CANON for this example. If this were not so, a further cycle would begin with the computation of the next atomic descriptors. If constitutionally equivalent atoms are present, any representative of their equivalence class can be selected arbitrarily without introducing ambiguity to the indexing procedure.

With CANON atomic indices are unequivocally assigned to atoms in molecules as well as to the corresponding rows (and columns) of the be-matrices with due consideration of chemical constitution and constitutional symmetries. Redundancies are thereby avoided.[21] For our purposes CANON is superior to the Morgan algorithm that is used by the Chemical Abstracts Service.[141] Since CANON recognizes constitutionally equivalent atoms, it is also suitable for ligand indexing in stereochemistry[140, 142] and for predicting the chemical shift patterns of NMR spectra. When stereochemical problems are solved with computer-assistance, CANON-based ligand indexing has some significant advantages over the system of CIP rules.[143] In contrast to CANON, the system of CIP rules takes into account the formal bond orders instead of the coordination numbers of the atoms. CANON is simpler and unambiguous in all cases. In addition, CIP is based on a comparison of the individual ligands instead of an analysis of the whole chemical constitution of the considered molecules. The computer program CANON will soon be available for interested users.[144]

4.3.2. Recognition and Correlation of Substructures

For the computer-assisted solution of chemical problems the presence or absence of certain specified substructures in the generated molecules is often required. Therefore, the substructure search algorithm CABASS[145] based on central atoms and a computer program were developed. CABASS serves as a functional subunit in IGOR2.[115]

Substructure search systems, such as the DARC system of Dubois et al.,[146] often operate through a screening procedure and a subsequent atom-onto-atom mapping. The first procedures for atom-onto-atom mappings were published by the Chemical Abstracts Service[147] and by Sussenguth.[148] In its screening procedure, the DARC system uses spherical fragments that are coded as byte strings and embedded in tree structures for storage in a data bank. The search for substructures exploits this tree structure. The method of spherical fragmentation is also used by Lynch et al.[149] for the generation of screening systems for substructure search in structural data files. When Friedrich[150] analyzed the customary substructure search methods[146, 151] he found that they are not suitable for the correlation of molecules according to common substructures. This is particularly the case when large files of data must be analyzed, because computing time and memory space increase more steeply than a polynominal curve with the size of the problem.[152]

Recently Ihlenfeldt[153] developed a computer program for structure-oriented correlation of molecules. Although this program is hampered by the disadvantages of the common procedures, it is capable of processing moderate amounts of data quite efficiently.

Fortunately, the complexity of molecular graphs is generally only moderate. Thus, it is possible to devise algorithms to correlate molecules through their substructures so that the need for memory space and computing time increases at worst polynomially (degree ≤ 5) with size and number of molecules that are to be considered. These algorithms are based on stepwise fragmentation of molecules and imbed the resulting fragments in a network of father–son relations which includes all the listed molecules. Based on this concept J. Friedrich[150] tested CORREL, a substructure correlation program for bilateral synthesis design. Improvements of CORREL led to CORREL 2,[154] which is more suitable for routine use. Recently, CORREL-S was employed for bilateral synthesis design;[155] it is based on a selection of substructures.

A special version of CORREL has been used for documentation and sequence matching of peptides[156] and is now being extended to include nucleotide sequences, as well as their correlation with peptide sequences.

The basic design of CORREL, published[150] with its source code, stimulated the implementation of several substructure search and correlation programs of similar design like RESY,[157] KOWIST,[158] HTSS,[159] S4,[151, 160] and a program by Klopman.[161, 162] The latter program as well as RESY and KOWIST are particularly suitable for structure–activity studies.[158]

Strategic Bonds and Substructures Relevant for Syntheses

For bilateral synthesis design it is necessary to find suitable starting materials for the synthesis of given target molecules. Here the largest common substructures are used as a guideline. CORREL-S was developed for this purpose.[89, 155]

When a data bank of substructures relevant for synthesis is assembled, a set of rules about "breakable" bonds is required to reduce the size of the file in a meaningful way.[163] To establish this set, Corey's rules about strategic bonds were modified, hierarchically ordered, and provided with a termination criterion: the procedure is terminated when all nonaromatic rings have been dissected. A strategic bond must fulfill the following criteria:

1. The bond does not belong to a carbocyclic aromatic system. This does not apply to heteroarenes, which are often assembled in the course of a synthesis.
2. A bond is ranked by the number of rings its rupture will open. The higher this number, the higher the bond's preference; the number of rings is determined by CANON[139] and enumerated according to the Frèrejacques formula.[44]
3. If there are more than two bonds that qualify by rule 2, those that generate the fewest new rings are given priority.
4. Bonds at heteroatoms are strategic.
5. A bond is preferred if it is *exo* to other rings.
6. The bond must involve at least one carbon atom with more than two neighboring non-hydrogen atoms. This essentially corresponds to rule 5 when polycyclic structures are

present. However, rule 6 becomes relevant when only isolated single rings are left over.

7. A multiple bond is strategic.

8. The bond is a direct or next neighbor to a bond according to rule 4 or 7.

With these rules it is possible to process simple and complex ring systems. The preference rules help to restrict the number of strategic bonds and to establish the priorities for bond rupture in order to reach substructures relevant for the synthesis in available starting materials as soon as possible.

4.3.3. Stoichiometric Completion of Truncated Reaction Equations

The co-products of the target molecule of a synthesis, its stoichiometric complements in the target EM, are found with the computer program STOECH.[164] STOECH is based on a matrix formalism that was used in the elucidation of the reaction mechanism[165] of the four component condensation (4CC; α-addition of iminium ions and anions to isocyanides, followed by secondary reactions;[166] Ugi reaction[167]). Stoichiometric completion of truncated reaction equations[168] is now also of interest for the systematic documentation of chemical reactions based on their hierarchic classification (see Section 4.3.5).

4.3.4. Determination of Minimal Chemical Distance and the Corresponding Atom-onto-Atom Mappings

The repeated determination of the chemical distances (CDs) between isomeric EMs is a prerequisite for the optimization of reaction networks.[21, 100, 169] The associated atom-onto-atom mappings are needed in hierarchic reaction documentation.[79, 100, 112] Since the computer program PMCD[105] for the approximate determination of the minima of CD, which is based on a heuristic "branch-and-bound" algorithm,[170] does not always yield satisfactory results the computer program PEMCD[106] for the exact determination of the minima of CD was tested. A reaction core with 16 contiguous reaction centers is the upper limit for the present version of the PEMCD. We are therefore presently developing a more powerful computer program as a sucessor of PEMCD.

The CD between EMs with up to 100 reactive centers can be determined with a program that was recently implemented by Fontain.[107] This program is based on the "genetic algorithms".[171] Unfortunately this CD-minimization program does not yield all atom-onto-atom mappings that belong to the minima of CD.

4.3.5. Reaction Documentation

In the customary commercially available reaction documentation systems, chemical reactions are represented by their starting materials and products. This does not suffice to solve the reaction documentation problem satisfactorily.[172] Until recently ORAC[173] outstripped the available reaction documentation systems, but it is unfortunately no longer available. Since the deficiencies of traditional reaction documentation are known, and since reaction documentation belongs to the most important, yet unsolved problems of computer chemistry, many new approaches have been mooted in this field.[174, 175] Therefore within the Munich project, a hierarchically ordered system has been developed[79, 112] that refers directly to the processes of the chemical reactions as such; up to now this system has not found appreciable acceptance. The main reason is that the considered reactions must be stoichiometrically balanced, that is, they must be representable as isomerizations of EMs. In the chemical literature, reactions are often published in truncated form, where some of the participating reactants are omitted and would have to be stoichiometrically completed. This is now possible through the recently published program STOECH (see Section 4.3.3).

4.3.6. Graphic Output of Results

Since the results of the computer-assisted deductive solution of chemical problems are be-matrices, the graphic output program MDRAW[176] was developed. This converts the be-matrices into the graphic constitutional formulas that are more familiar to chemists. ARGOS, a similar program that was tested by J. Bauer and E. Fontain, is part of the user interface of IGOR and RAIN, and it has also been used routinely for several years by the Beilstein Institute for the conversion of connectivity lists into constitutional formulas.

4.4. The Multipurpose Programs IGOR and RAIN[*]

4.4.1. Formal Reaction Generators

The fundamental equation $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ of the DU model can be solved from a given be-matrix $\boldsymbol{B}$ by determining those pairs $(\boldsymbol{R}, \boldsymbol{E})$ which fulfill $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ under the given boundary conditions. We call these solutions the b-solutions. They are found by *r*eaction *g*enerators (RG) of the type RGB.[169, 177, 178] The equation $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ can also be solved from a given r-matrix $\boldsymbol{R}$. These r-solutions correspond to the pairs $(\boldsymbol{B}, \boldsymbol{E})$ for which $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ is fulfilled and are obtained by reaction generators of the type RGR.[114–116]

The two complementary types of solutions of the equation $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ and the respective RGs correspond to the two basic types of computer programs for the solution of chemical problems on the basis of the DU model. The centerpiece of the present version IGOR2[115] of the computer program IGOR[114–116] is an RGR. Recently RAIN[60–62, 169, 177, 178] also became generally available; the "engine" of RAIN is an RGB.

The RG of IGOR and RAIN are guided by *t*ransition *t*ables (TTs).[62, 63, 177, 178] A standard TT or a TT that is defined by the user is assigned to each chemical element that is taken into account. The allowable transitions are recorded in these TTs (see Fig. 2).

The computer programs IGOR and RAIN operate with strictly formal reaction generators, and are thus not restricted to the solution of a particular type of problem, such as retrosynthetic analysis. In combination with suitable auxil-

[*] IGOR and RAIN are available from the authors on request.

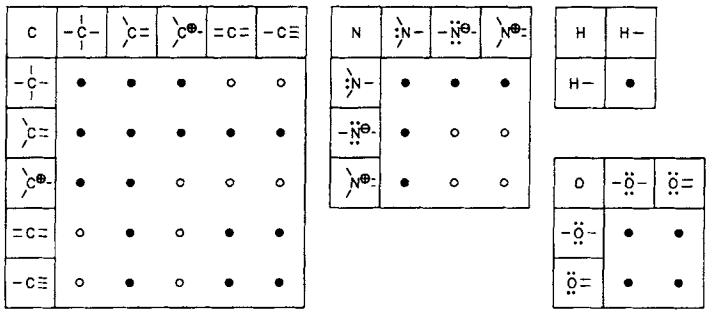


Fig. 2. Examples of transition tables. • = allowed transition, o = forbidden transition.

iary programs, both IGOR and RAIN can deal with a wide variety of chemical problems and deserve to be called multipurpose problem-solving programs. The solutions that they generate can belong to known chemistry, or they can be entirely without precedent, because these programs are independent of detailed empirical chemical information.

The first TT-guided RG (TRG) was the TRGR of IGOR, which solves the equation $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ from an r-matrix $\boldsymbol{R}$ and a list of the TTs of all relevant chemical elements. The user selects a suitable collection of TTs for each row/column pair of the potential solutions ($\boldsymbol{B}$, $\boldsymbol{E}$). The TRGR checks whether the entries of the TTs are compatible with $\boldsymbol{R}$ to provide a reduced collection of TTs, which are used to generate the be-matrix $\boldsymbol{B}$, row by row, column by column. From the rows and columns of $\boldsymbol{B}$, $\boldsymbol{R}$ generates rows and columns of $\boldsymbol{E}$. Now all row/column combinations of ($\boldsymbol{B}$, $\boldsymbol{E}$) are verified through use of the TTs. The novelty of each generated pair ($\boldsymbol{B}$, $\boldsymbol{E}$) is checked by CANON[139] in order to avoid redundancies. Subsequently the pairs of matrices are analyzed by CABASS[145] for forbidden or required substructures. The acceptable solutions are finally represented graphically on the screen or printed.

IGOR's TRGR generates pairs of starting materials and products according to a given irreducible r-matrix $\boldsymbol{R}$, which represents the redistribution of electrons (see Section 3.2.2). Thus, IGOR uses given "electron-pushing" patterns to "invent" chemical reactions. The r-matrices of such electron-pushing patterns can be produced by an independent computer program.[179] The reactions that are generated by IGOR are subjected to an interactive selection procedure that follows the hierarchic classification system for chemical reactions.[114]

In the structure-generating mode, a zero matrix $\boldsymbol{R} = \{\}$ is used as an r-matrix. IGOR then produces be-matrices $\boldsymbol{B}$ of molecules or EMs with $\boldsymbol{B} = \boldsymbol{E}$. Sets of molecules with the specified structural features can be obtained by restricting IGOR's output.[114, 177, 180, 181]

An RGB generates all EMs into which a given EM is directly convertible. Since an RGB may also use $-\boldsymbol{R}$ instead of $\boldsymbol{R}$, it can also generate the EMs from which the input EM can be obtained. When incorporated into a retrosynthetic program, an RGB can play the same role as a reaction library in combination with a structure-perceiving module, that is, the RGB finds the precursors of a given target molecule and their precursors, etc.

A TRGB operates in two steps: the first step is to determine the allowable valence schemes of those chemical elements that are assigned to the rows and columns of $\boldsymbol{E}$. They are obtained from the rows and columns of $\boldsymbol{B}$ through the TT. Subsequently, all $\boldsymbol{E}$s are generated that fulfill $\boldsymbol{B} + \boldsymbol{R} = \boldsymbol{E}$ under the given boundary conditions.

RAIN's TRGB[60, 63, 178] generates isomeric EMs from an $EM_{\boldsymbol{B}}$ that is represented by its be-matrix $\boldsymbol{B}$. The isomeric EMs correspond to the products that can be directly formed from $EM_{\boldsymbol{B}}$, or to starting materials for the $EM_{\boldsymbol{B}}$. RAIN is capable of simultaneously generating two trees of successive chemical reactions that consist of sequences of isomeric EMs. The "geometric" properties of the family of all isomeric EMs are exploited to direct the trees so that they grow from $EM_{\boldsymbol{B}}$ and $EM_{\boldsymbol{E}}$ and meet as soon as possible to form a contiguous network of reaction pathways; this bilateral generating process guides each reaction pathway to meet by checking the CD of the intermediate EMs[21, 100, 107, 182, 183] and is terminated as soon as an EM is reached from both sides. Thus, RAIN produces reaction networks that connect isomeric EMs.[60, 63, 177, 184, 185] By interactively specifying the lower and upper bounds and other characteristics of the reaction pathways (number and structural features of the intermediates, number of redistributed electrons per reaction step, etc.), the user can determine the nature of the network.[21] Depending on whether only the valence schemes of stable compounds are admitted, or also those of unstable intermediates, the networks that are generated contain only stable molecules or also short-lived intermediates.

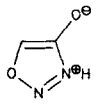
Accordingly, RAIN may be used for mono- and bilateral synthesis design,[84, 88, 89] for planning and predicting reactions, and also for the elucidation of biosyntheses and complex reaction mechanisms (see Section 4.4.4).

Recently, Valdés-Pérez[186] presented the system MECHEM that is capable of generating networks of reaction pathways by symbolic computation. However, the latter program can only take empirical formulas into account, and not constitutional formulas.

4.4.2. Structure Generation by IGOR and RAIN

In their structure-generating modes, IGOR and RAIN produce all constitutional formulas that meet the specified conditions. Thus, for instance, all 23 EMs with the formula $(CH)_8$ and all 18 1,3-dipoles[187] of the elements C, N, and O are generated.[114] In order to demonstrate how the number of isomers increases when the C and H atoms of hydrocarbons are partially replaced by heteroatoms, the number of constitutional formulas of compounds with the empirical formula $C_2H_2N_2O_2$ (nonionic; no triple bonds or cumulated double bonds in rings; ammonium or iminium N^+ or amide N^- and O^- allowable as electronically charged atoms) were enumerated by IGOR and RAIN. Both programs found 1806 EMs, whereas there are only six hydrocarbons C_4H_4. When these molecules are generated by RAIN, the network of all constitutional isomers was generated from two atoms of each of the elements C, H, N, and O. For comparison nonionic $C_2H_2N_2O_2$ isomers were also generated that have a five-membered ring and formally one positive and one negative electrical charge. Sydnone **50** and 51 of its analogues result.

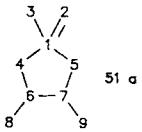
At present the customary reagents for the synthesis of oligonucleotides are derivatives of phosphorous acid[188] because, in general, the alternative phosphoric acid derivatives



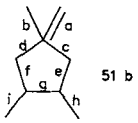
50

do not react sufficiently quickly with nucleoside derivatives. However, five-membered cyclic phosphoric acid derivatives react 10^5–10^7 times faster with nucleophiles or apicophiles than their acyclic analogues.[189, 190] Since 1973, the research groups of F. Ramirez and I. Ugi endeavoured to exploit this fact for syntheses by attempting to develop highly reactive five-membered phosphorylating reagents based on P^V for oligonucleotide syntheses.[191] These reagents would have some advantages over customary P^{III} reagents.[192]

After some initial success[193] the development stagnated, mainly because new ideas were lacking. Only after Ugi et al.[180] decided to review all of the candidate P^V compounds as defined by **51a** and **51b** with the aid of IGOR was progress again rapid.



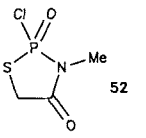
51 a



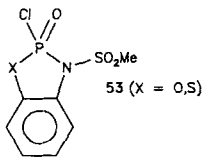
51 b

The following atoms were admitted for **51a**: 1 = P; 2 = O; 3 = X; 4, 5 = O, N, S; 6 = C, N, O, S; 7 = sp^2-C; 8, 9 = C, Cl, N, O, S. H atoms are placed according to valence rules. The covalent bonds are distributed as indicated by **51b**: a double bond; b, c, d: single bond; e, f, g: single or double bond; h, i: single or no bond. IGOR found 278 formulas **51** that complied with the restricting definitions.[180]

Five of these candidates were chosen to be investigated experimentally.[194] Already, two promising phosphorylating reagents, **52**[192] and **53**,[195] have resulted from these studies.

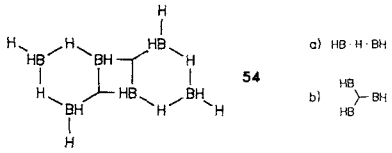


52



53 (X = O,S)

Lipscomb used the rules that he had established for the structure of boron hydrides to postulate structure **54** for B_6H_{14}, containing open (a) and closed three-center bonds (b).[196] Recent NMR data indicate that two kinds of B atoms (based on first-sphere neighbors) are present.[197] The formulas of Figure 3 generated by RAIN for B_6H_{14} agree with the observations.



54

Computer-assisted structure elucidation à la DENDRAL employing analytical data is a particularly promising application of IGOR and RAIN. The experimental evidence can be used as restricting conditions when molecular structures are generated. All constitutional formulas that are compatible with the measured data are obtained.

The graph theoretical representation of chemical constitution has already been successful within the DENDRAL pro-

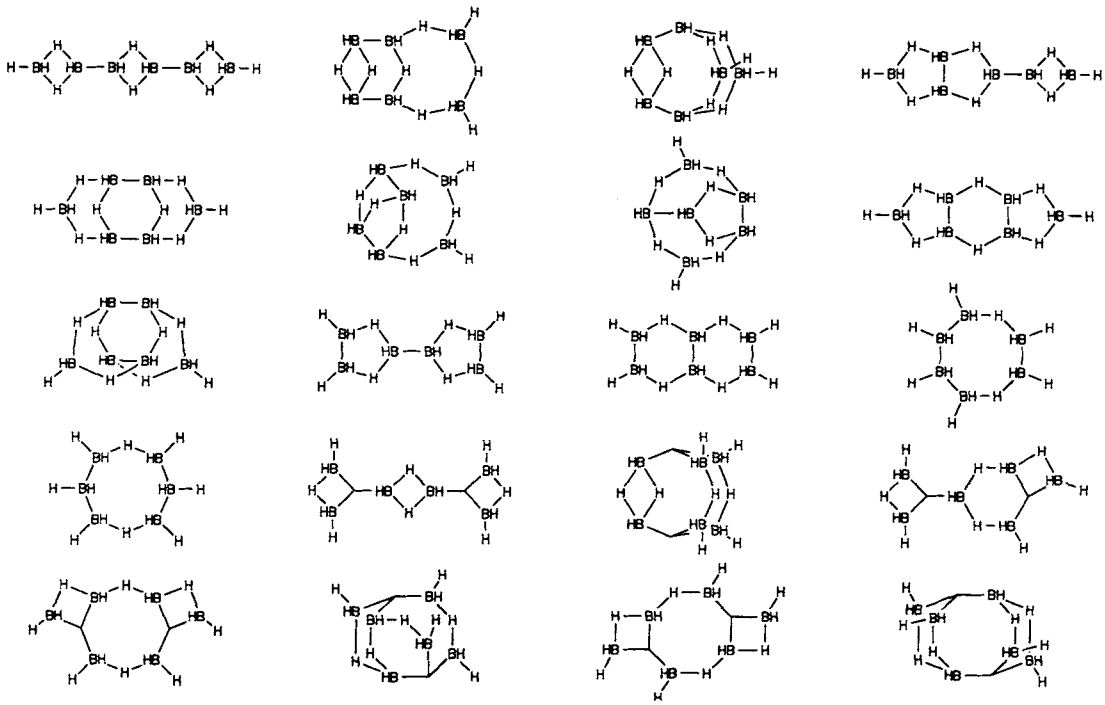
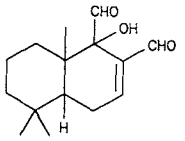


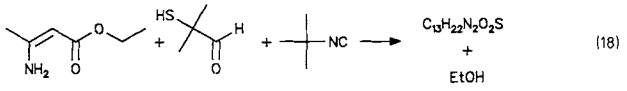
Fig. 3. Proposed structures of B_6H_{14} generated with RAIN.

ject as a mathematical foundation of the computer programs CONGEN[198] and GENOA,[199] and also later in the comparable programs RASTR,[200] MOLGRAPH,[201] ESSESA,[202] and GEN,[203] which generate families of constitutional formulas. GENOA was the first functioning program within this category. The computer-assisted elucidation of chemical constitution of walburganal **55** with the aid of GENOA is one of the most impressive applications of DENDRAL.[24]



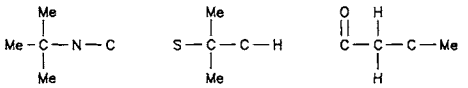
55

RAIN and GENOA generate the same set of 42 constitutional formulas, when the empirical formula $C_{15}H_{22}O_3$ of walburganal **55** and the substructures that are required or forbidden by the spectroscopic data are input. A substance with the empirical formula $C_{13}H_{22}N_2O_2S$ results from the reaction of ethyl β-aminocrotonate **56** and α-mercaptoisobutyraldehyde **57** with *tert*-butyl isocyanide **58**.[204]



(18)

56 **57** **58**



59 **60** **61**

Assuming that the substructures **59**–**61** in the starting materials are also present in the product, RAIN proposes the constitutional formulas of Figure 4.

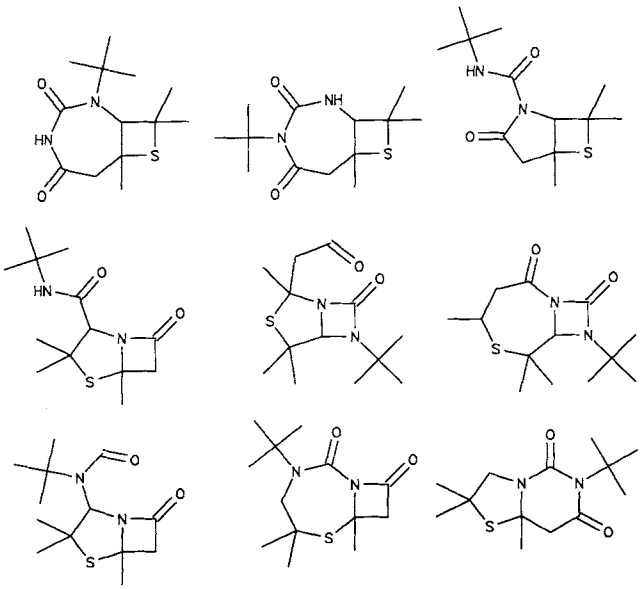


Fig. 4. Structures generated by RAIN as proposed products of reaction (18).

During routine runs, RAIN automatically recognizes the chemical equivalence of prototropic tautomers and resonance structures. For example, the tautomers and resonance structures of Figure 5 are obtained for the nucleobase cytosine.

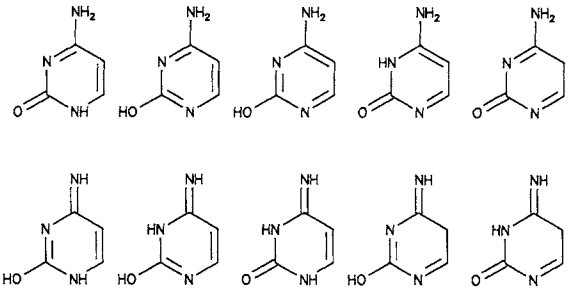


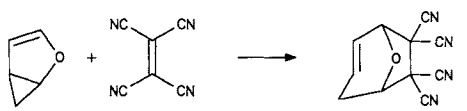
Fig. 5. The tautomers and resonance structures generated by RAIN for cytosine.

4.4.3. Generating Reactions with IGOR

Recently, Barton[205] published a very remarkable article on the systematic search for new reactions without computer-assistance. In this section the computer-assisted discovery of new reactions is described.

About ten years ago, R. Herges was given the task—as a part of his doctoral thesis—to find unprecedented reactions with the aid of the computer program IGOR,[116] which was then still under development, and to verify these reactions in the laboratory. The experience and suggestions gathered in the course of the investigations substantially improved the efficiency and user-friendliness of IGOR.

One of the results was that the chemist's participation varies strongly.[206] Computers only gave the inspiration for the development of cycloadditions of homodienes.[207] The systematic investigation of the r-class of the pericyclic 6-center 6-electron reactions demonstrated that in this 6,6 r-class of 13 basis reactions, only **25** → **26** with C atoms in the core of the reaction had not been studied much, although it was plausible. The reaction **62** + **63** → **64** was selected as an example according to heuristic criteria (ring strain, polarization, molecular geometry, etc.); this reaction succeeded in the laboratory.[114, 118]



62 **63** **64**

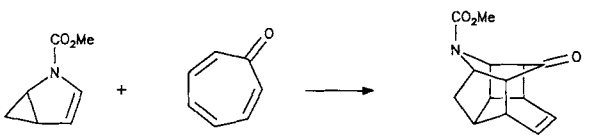
In this context, our experience was that it is very difficult to ensure the unprecedentedness of a reaction in terms of the chemical literature. The data banks of molecular structures (CAS online, Beilstein online, etc.), which contain data on almost all compounds that have ever been synthesized, are easy to research. The commercially available data banks of chemical reactions, however, are not sufficiently structured and far from complete, and precedence must be determined by searching for all potential starting materials and products of the considered reaction.

H. Prinzbach informed us that Fowler[208] carried out the cycloaddition of *N*-methylcarbonylhomopyrrole **65** to dimethyl acetylenedicarboxylate in 1971. This reaction is related to **62** + **63** → **64** because the reactants **62** and **65** are similar. We had overlooked this reaction.

The extrusion reaction **25** → **26** (Scheme 2) is a further example of an experimentally verified reaction that was found by IGOR through the hierarchic classification system of chemical reactions. Hydrogen was chosen as the group to be transferred, since hydrogen transfer reactions generally proceed with particular ease; a carbonyl oxygen atom appeared to be favorable as an sp^2 center. All other reactive centres of **25** → **26** were allowed to be C or O. Furthermore, the chemically meaningful condition was introduced that the extruded molecule must be CO_2. This led to four types of reactions, of which three were already known. One, however, was not, namely the pyrolysis of α-formyloxy ketones **25** to **26**. Three examples of this reaction were successfully carried out in the laboratory[114, 118] (see Scheme 2).

This reaction is of moderate novelty, because the rb-subclass of the extrusion reactions was put in. Accordingly, at best, a further extrusion reaction could be discovered as a "new" reaction. Nevertheless, reaction **25** → **26** is still interesting, novel, and even of preparative value. In the synthesis of cyclic ketones by acyloin condensation[209] it is superior to the customary reduction of acyloins by Zn/HCl.[210] Obviously, the above reaction could also have been found without computer-assistance, but it is interesting and surprising that an probable reaction in the intensively studied field of extrusion reactions had been overlooked. This demonstrates that the systematic search for new reactions is more successful with computer-assistance than without.

There are many conceivable reactions from the CD-class 8 (see Section 3.2.2) that are novel up to the level of the ra-subclasses.[114] Some years ago, a systematic search for pericyclic reactions whose basic reactions are without precedence was conducted with IGOR. The CD-class 20 (reactions with five redistributed electrons) was the first in which a new experimentally realizable pericyclic reaction was found[88, 117] (**65** + **66** → **67**).

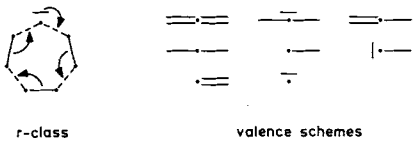


65 **66** **67**

Further reactions found through IGOR are a carbene rearrangement,[211] and a method for the synthesis of isocyanides containing electron-withdrawing groups.[206]

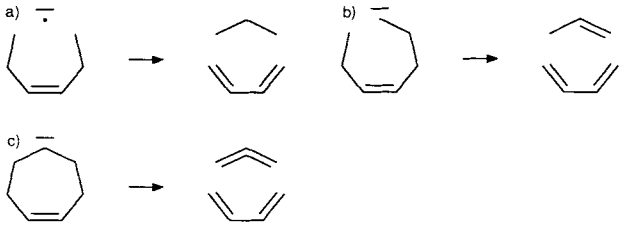
Besides systematic screening of classes of reactions for new reactions, a systematic search can be conducted for reactions that may be used for the synthesis of a given class of compounds. Herges and Hoock[212, 213] scanned the 7-center, 8-electron pericyclic reactions for syntheses of 1,3-dienes. The restrictions were given in the irreducible r-matrix and the valence schemes of the participating atoms (Scheme 10).

Of a total of 72 basic reactions generated (when triple bonds are admitted 470 basic reactions are generated), only three



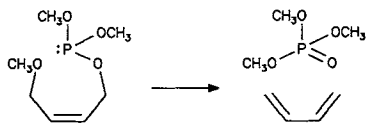
Scheme 10. Pericyclic 7-center, 8-electron reactions.

(Scheme 11) are suitable for the synthesis of 1,3-dienes (a butadiene fragment is produced in these cases).



Scheme 11. Potential reactions for the synthesis of 1,3-dienes according to Scheme 10. The arrow in a) indicates an atom possessing a lone pair of electrons (symbolized by a short line).

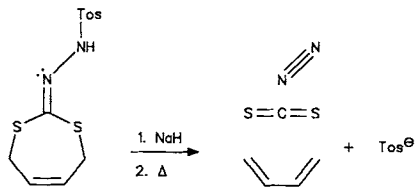
Only reaction a) is known through some variants involving heteroatoms. One example is the reduction of 1,4-dichloro-2-butene with Zn. To our knowledge, the basis reactions b) and c) have not been published. The variant **68** → **69** of basis



68 **69**

reaction b) was carried out recently, and yields 40% of butadiene in the gaseous phase at 0.40 Torr and 350 °C. It is not useful for syntheses, but is a novel type of 1,4-elimination.

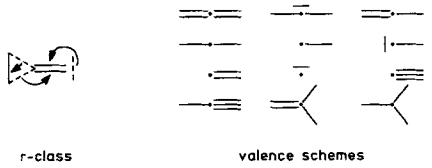
The example **70** → **71**[212] was verified as a representative of c).



70 **71**

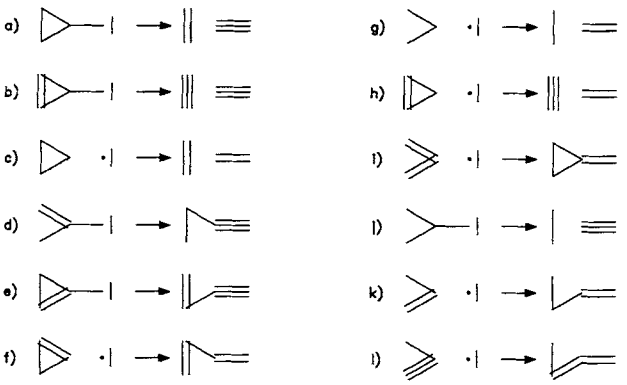
Not only pericyclic reactions can be generated in this way, but also reactions from other areas of organic chemistry. This can be seen in the example in Scheme 12 of a search for a novel carbene reaction.[214]

The first four basis reactions of Scheme 13 are illustrated in published examples. The fragmentations of cyclopropyl and 2-cyclopropyl carbene (a and b) have been thoroughly investigated.[215, 216] Also, reactions of carbenes with strained



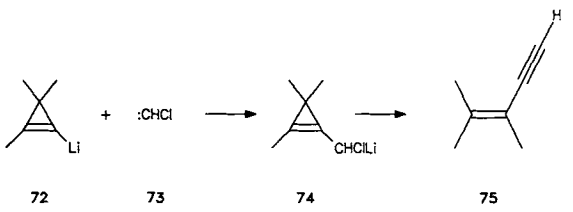
Scheme 12. Carbene reactions.

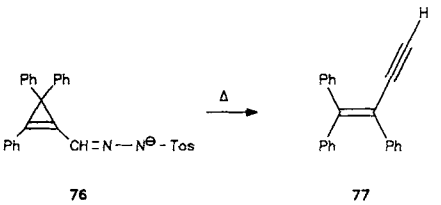
three-membered rings (c)[217] and the carbene rearrangement (d)[218] are already known. Herges[214] experimentally realized the unprecedented basis reaction (c) for two different carbene precursors (**72** + **73** → **74** → **75**; **76**–**77**).



Scheme 13. Potential carbene reactions. Lone pairs of electrons are symbolized by short lines.

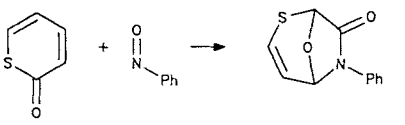
Recently, using IGOR2, Fisher, Juarez-Brambila, Goralski, Wipke, and Singaram[219] found and experimentally verified a novel rearrangement of α-aminoalkylboranes to the corresponding β-dialkylaminomonoalkylboranes.





4.4.4. *Generation of Reaction Networks with RAIN*

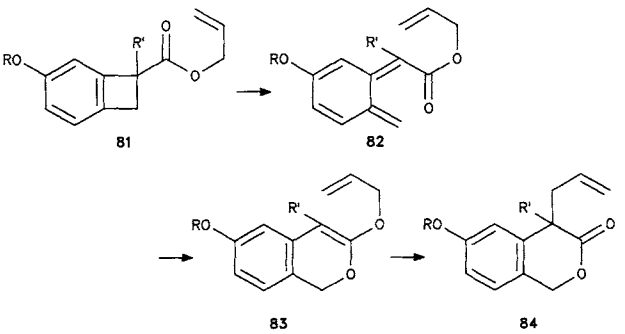
In the summer of 1982, a reaction mechanisms contest took place at the University of California at Los Angeles. The topic of this contest was the proposal of a plausible reaction mechanism for the newly published Streith–Defoin reaction **78** + **79** → **80**.



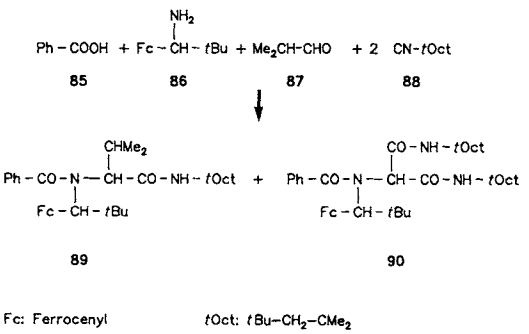
M. Jung, who organized the contest, asked I. Ugi to make some computer-generated proposals for the mechanism of for the Streith–Defoin reaction. We succeeded only after RAIN was available.[21, 60, 63] The Streith–Defoin reaction[220, 221] played an important role as a "fitness bicycle" in the development of RAIN. The network that RAIN generates under suitably chosen boundary conditions[63] contains the four best proposals of that contest; one of these corresponds to the reaction mechanism that has since been elucidated.[221]

Ried and Dietrich's indazole synthesis,[222] the thiazole synthesis by Hantsch,[223] and the Favorskii rearrangement[224] belong to the first applications of RAIN. Initially, a surprisingly large network of conceivable reaction mechanisms was found for the Favorskii rearrangement—116 intermediates at six levels and 43 intermediates at its "widest" level. Besides the published reaction mechanism,[224] this network contained many other mechanisms that are compatible with the known experimental results. When a suitable set of restrictions is input to the present version of RAIN, it generates a single reaction mechanism for the Favorskii rearrangement—the published one.

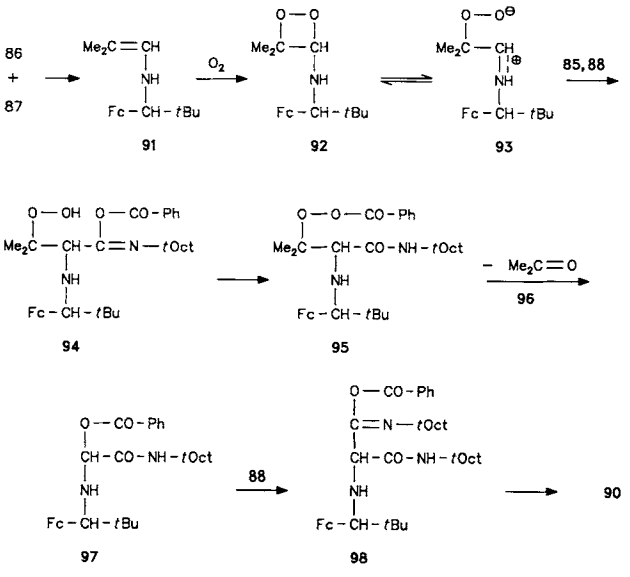
RAIN finds only one reaction mechanism for the rearrangement of benzocyclobutene **81** into isochroman **84**[62] that was studied by Kametani et al.[225] The intermediates were **82** and **83**.



Recently the mechanism of an undesirable side reaction of the four component condensation was elucidated with RAIN.[184, 185] The formation of **89** and the malonamide



derivative **90** from **85**–**88** served as a model reaction. Experimental evidence in combination with suggestions from RAIN leads to the assumption that **90** is formed by the reaction mechanism depicted in Scheme 14 via the intermediates **91**–**98**.



Scheme 14. Mechanism for the formation of **90**.

The elucidation of the mechanism of the hypothetical prebiotic synthesis of adenine from five molecules of hydrogen cyanide[226] is still open-ended, since the experimental evidence is not enough to decide which of the conceivable mechanistic alternatives proposed by RAIN is valid.

5. Perspectives

Nowadays computers belong to the standard instrumentation of chemical research, and it is safe to predict that computers will increasingly contribute to progress in chemistry in ever more diverse ways. The importance of numerical applications in chemistry, in particular in quantum chemistry and chemometrics but also in the planning, executing, and evaluating of measurements and in molecular modeling, will continue to soar. The use of computers in the documentation of chemistry-related data and information will develop vigorously in volume and efficiency.

It is foreseeable that computer programs for the direct solution of chemical problems will become routine in chemical research. In this context, formal algorithms will play an increasingly important role, even in expert systems that rely on stored detailed information. Empirical data will primarily be used for comparison with the results of computer programs that operate on a formal basis.

As a consequence of inquisitiveness of the potential users, computer programs with innovative capabilities will be preferred if the offered software has a convenient user interface and is suitable for widely available hardware. It is also important that such programs operate in an interactive mode, because an intelligent and experienced user wants to participate in the problem-solving process. With the aid of a suitable computer program it is, in principle, possible to generate all the conceivable solutions of a given chemical problem, including the most imaginative ones, but these solutions are worthless if they are hidden under enormous amounts of insignificant data; the meaningful and nonarbitrary selection of the solutions cannot be accomplished without the participation of a qualified user. The computer is able to provide assistance in arriving at new chemical ideas, but chemists must recognize their value and select them from a large number of conceivable alternatives. Thus, innovation may switch from the generation of proposals to their evaluation and selection, but new chemistry cannot be produced without the participation of human creativity. Progress in computer chemistry is based on new insights, ways of reasoning, formalisms, algorithms, software techniques, and also advances in computer hardware. Since the development of computers is still vigorous, it is too early to evaluate the future importance of single projects and tendencies. Much of what now looks impressive will be forgotten tomorrow, and what now seems to be excentric or irrelevant may turn out to be the beginning of a useful and important development.

We would like to thank the Deutsche Forschungsgemeinschaft, the Volkswagen-Stiftung, the Alexander-von-Humboldt-Stiftung, the Commission of the European Community, the Bundesministerium für Forschung und Technologie, and the Fonds der Chemischen Industrie for the generous financial support of our work.

Received: July 6, 1991
Revised: July 14, 1992 [A 894 IE]
German version: *Angew. Chem.* **1993**, *105*, 210

Mathematical Concepts and Notations

Boolean Algebra (let a and b be Boolean values)

$\mathbb{B}$ — set of Boolean values {O, L}

$\wedge$ — conjunction (logical and): connective $\wedge : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ according to the table

$\wedge$	O	L
O	O	O
L	O	L

$\vee$ — inclusive disjunction (logical or): connective $\vee : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ according to the table

$\vee$	O	L
O	O	L
L	L	L

$\Rightarrow$ — implication ($a \Rightarrow b$: "if a then b"): connective $\Rightarrow$: $\mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ according to the Table

$\Rightarrow$	O	L
O	L	L
L	O	L

$\Leftrightarrow$ — equivalence ($a \Leftrightarrow b$: "a is equivalent to b", "a has the same value as b"): connective $\Leftrightarrow$: $\mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ according to the Table

$\Leftrightarrow$	O	L
O	L	O
L	O	L

$\neg$ — negation (logical not): $\neg$: $\mathbb{B} \rightarrow \mathbb{B}$ with $\neg(0) = L$ and $\neg(L) = O$.

Quantifiers

$\exists$	existential quantifier ("there exists at least one ...")
$\forall$	universal quantifier ("for all ...")

Sets (let A, B, M and N be sets)

$\in$	$m \in M$ signifies that m is an element of M.
$\cap$	The intersection $A \cap B$ of two sets contains the elements that belong to both sets A and B.
$\cup$	The union $A \cup B$ of two sets A and B contains those elements that belong either to A or to B.
$\setminus$	The difference $A \setminus B$ of two sets A and B contains those elements of A that do not belong to B.
$\{m: P(m)\}$	The set of all objects, for which the predicate P holds. For example, the expression $\{n: n = 2m;\ m \in \mathbb{N}\}$ describes the set of all even natural numbers.
$\lvert M \rvert$	Cardinality: the number of elements in a set M
$M \times N$	The set of all ordered pairs (m, n) that can be formed from the elements $m \in M$ and $n \in N$ of two sets M and N.
$\mathrm{SVM}(\mathscr{A}, B) = \bigcup_{A \cap B \neq \{\}, A \in \mathscr{A}} A$	The union of all sets A that belong to the family of sets $\mathscr{A}$ and have a nonempty intersection with the set B.

Relations, Mappings

binary relation A binary relation R associates pairwise the elements of two sets X and Y. Accordingly, R is the set of ordered pairs $X \times Y$, expressed by $(x, y) \in R$, or in infix terminology: xRy. A binary relation R may be defined by a predicate P that is valid for all pairs $R = \{(x, y) \in X \times Y: P(x, y)\}$; otherwise all pairs are stated explicitly.

transitive closure The transitive closure R^+ of a (binary) relation R is defined as follows:
(1) $(a, b) \in R \Rightarrow (a, b) \in R^+$
(2) $(a, b) \in R^+ \wedge (b, c) \in R \Rightarrow (a, c) \in R^+$
(3) R^+ contains only what follows from (1) and (2).
R^+ is the smallest transitive relation that comprises R.

equivalence relation Special case of a binary relation $R \subseteq X \times Y$ for which the sets X and Y are identical and which is also endowed with the following three properties:
(1) reflexivity: $\forall x \in X: xRx$ (i.e. each element is in relation with itself).
(2) symmetry: $\forall x, y \in X: xRy \Rightarrow yRx$ (i.e. a relation between elements x and y implies a relation between y and x).
(3) transitivity: $\forall x, y, z \in X: xRy \wedge yRz \Rightarrow xRz$ (i.e. if there is a relation between x and y, and between y and z, then there is also a relation between x and z).

Due to the symmetry of an equivalence relation R on X, for some pair $(x, y) \in R$, the phrase "x has a relation R to y" is generally replaced by "x and y have a relation R"; for any x the set $[x] = \{y: xRy\}$ is the equivalence class of x with regard to R. For z the equivalence class $[z]$ is either identical with $[x]$ if $z \in [x]$, or the equivalence classes are disjoint (i.e. without common elements). The partition $\{[x]: x \in X\}$ of X that is induced by the set of all equivalence classes is called a space of equivalence classes or a quotient set.

$[m]_R$ The equivalence class of m with regard to the equivalence relation R, that is, all elements that are equivalent to m according to the property that is represented by R. Whenever it is necessary to declare the equivalence of some permuted models within a family with regard to a given property, the equivalence relations are of interest in the theory of the CIG.

permutations In the present article permutations are denoted by lowercase Greek letters. A permutation of n elements is a bijective map $\lambda: L \rightarrow L$ of a set L onto itself. A permutation $\lambda(l_a) = l_b, \lambda(l_b) = l_c, \ldots, \lambda(l_s) = l_a$ can be written in cycle notation as $(l_a l_b \ldots l_s)$. All further elements $l \in L \setminus \{l_a, l_b, \ldots, l_s\}$ are mapped to themselves by λ.

Groups

group A group $G = (M, \bullet)$ consists of a set M and an operation $\bullet$ in M with the following properties:
(1) closure: The combination of any two elements $a, b \in M$ yields again an element of M.
(2) associativity: For all elements $a, b, c \in M$ we have $a \bullet (b \bullet c) = (a \bullet b) \bullet c$.
(3) existence of an identity element: There exists an identity element $e \in M$ such that we have for all elements $a \in M$: $a \bullet e = e \bullet a = a$.
(4) existence of an inverse element: For any element $a \in M$ there exists an element a^{-1} such that: $a \bullet a^{-1} = a^{-1} \bullet a = e$.

In the context of a group the operation symbol $\bullet$ is omitted and one writes "ab" instead of "$a \bullet b$". The cardinality of a group is also called its order. A subgroup of the group $G = (M, \bullet)$ is a group $U = (N, \bullet)$ with $N \subseteq M$.

coset The set aU that belongs to an element $a \in G$ is called a left coset of U; it comprises all elements au with $u \in U$. If the multiplication by a takes place from the right hand side, the right coset Ua result. If the direction of the multiplication follows from the context or is irrelevant, the term coset is used. The set of the (left/right) cosets of the subgroup U in the group G form a (left/right) coset space. A coset space of U in G is a partitioning of G. Any coset space is a space of equivalence classes.

transversals A minimal set T with the property $TU = G$ is called a complete left transversal of U in G, and analogously as a complete right transversal, if $UT = G$. If it follows from the context whether a left or right transversal is meant, the term transversal or system of representatives is used. If U is nontrivial—that is, U does not only contain the identity element—the transversal is ambiguous. The complete transversal of U in G contains precisely one representative from each coset of U in G. Any element of a coset may serve as its representative. The term leading elements of cosets is often used for representatives.

generators In order to define a group G, it suffices to specify the nature of the group operation $\bullet$ and a few elements that, together with their inverses, can be combined to yield G. The latter elements are called the generators of the group G. If the elements of a group are permutations, the term permutation group is used. The set P of all permutations on a set L of n symbols, together with the consecutive execution of permutations as the group operation $\bullet$, form a group $Sym(L) = (P, \bullet)$ with $n!$ permutations; it is called the symmetric group of degree n. Often it is not relevant which elements of L are permuted, but how many. Thus one writes $S_{|L|}$ instead of $Sym(L)$. In the theory of the CIG, permutation group have direct applications. Any family of permuted models $P(E)$ with a set of ligands L is isomorphic to the symmetric group $Sym(L)$, since the action of distinct permutations from $Sym(L)$ on the reference model E leads to distinct models in $P(E)$; any model from P(E) can be obtained by the action of a permutation of $Sym(L)$ on the reference model.

conjugates For $\lambda \in G$ the product $\lambda U \lambda^{-1}$ is called the $(\lambda -)$ conjugate of U in G. The conjugate of a group is itself a group. For $\mu \in G$, the set $\{\lambda\mu\lambda^{-1}: \lambda \in U\}$ is called a Wigner subclass of U in G. The set $\{\{\lambda\mu\lambda^{-1}: \lambda \in U\}: \mu \in G\}$ of the Wigner subclasses forms a space of equivalence classes of G.

[1] D. Seebach, *Angew. Chem.* **1990**, *102*, 1363; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 1320.
[2] a) E. L. Eliel, N. L. Allinger, S. J. Angyel, G. A. Morrison, *Conformational Analysis*, Interscience, New York, **1965**; b) M. Hanack, *Conformational Theory*, Academic Press, New York, **1965**; c) G. Chiurdoglu, *Conformation Theory*, Academic Press, New York, **1971**.
[3] H. Sachse, *Ber. Dtsch. Chem. Ges.* **1890**, *23*, 1323; *Z. Phys. Chem.* **1892**, *10*, 203.
[4] E. Mohr, *J. Prakt. Chem.* **1918**, *98*, 315.
[5] H. Meerwein, K. van Emster, *Ber. Dtsch. Chem. Ges.* **1922**, *55*, 2500.
[6] a) W. Reppe, O. Schlichting, K. Klager, T. Toepel, *Justus Liebigs Ann. Chem.* **1948**, *560*, 1; for a retrospective comment see [6b]; b) G. Schröder, *Cyclooctatetraen*, Verlag Chemie, Weinheim, **1965**.

[7] a) O. Roelen, *Angew. Chem.* **1948**, *60*, 213; for reviews see [7b]; b) J. Falbe, *Carbon Monoxide in Organic Synthesis*, Springer, Heidelberg, **1970**; W. A. Herrmann, Kontakte (Darmstadt) **1991**, (1), 22.
[8] J. M. Bijvoet, A. F. Peerdeman, A. J. Bommel, *Nature* **1951**, *168*, 271; J. M. Bijvoet, *Endeavour* **1955**, *14*, 7.
[9] G. A. Olah, G. K. S. Prakash, J. Sommer, *Superacids*, Wiley, New York, **1985**.
[10] G. Domagk, *Klin. Wochenschr.* **1937**, *16*, 1412; F. Mietsch, *Ber. Dtsch. Chem. Ges. B* **1938**, *71*, 15; *Chem. Zentralbl.* **1938**, 120.
[11] K. H. Büchel, *Bayer. Landwirtsch. Jahrb.* **1981**, *58*, 27; K. H. Büchel, M. Plempel, *Chron. Drug Discovery* **1983**, *2*, 235.
[12] K. Zuse, *Der Computer, mein Lebenswerk*, Springer, Heidelberg, **1984**.
[13] *Computer Applications in Chemical Research and Education* (Eds.: J. Brandt, I. Ugi) Hüthig, Heidelberg, **1989**.
[14] *Computational Methods in Chemistry* (Ed.: J. Bargon), Plenum, New York, **1980**.
[15] M. D. Johnston, Jr., *Computational Chemistry*, Elsevier, Amsterdam, **1988**; D. M. Hirst, *A Computational Approach to Chemistry*, Blackwell, Oxford, **1990**; *Modern Techniques in Computational Chemistry: MOTECC-90* (Ed.: E.Clementi), ESCOM, Leiden, **1990**.
[16] *Computer Aids to Chemistry* (Eds.: G. Vermin, M. Chanon), Ellis-Horwood, Chichester, **1986**; B. Adler, *Computerchemie – eine Einführung*, VEB Leipzig, **1986**.
[17] M. A. Ott, J. H. Noordik, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 239.
[18] J. Zupan, *Algorithms for Chemists*, Wiley, New York, **1985**.
[19] *PC for Chemists* (Eds.: J. Zupan, D. Hadzi), Elsevier, Amsterdam, **1990**.
[20] I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, M. Knauer, K. Reitsam, N. Stein, *Labor 2000* (special issue of *Labor Praxis*) **1992**, 170.
[21] I. Ugi, E. Fontain, J. Bauer, *Anal. Chim. Acta* **1990**, *235*, 155.
[22] J. Lederberg, *Proc. Natl. Acad. Sci. USA* **1990**, *53*, 134; *How DENDRAL was conceived and born* (ACM Symposium on the History of Medical Informatics 1987), National Library of Medicine, p. 270.
[23] R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*, McGraw-Hill, New York, **1980**.
[24] N. A. B. Gray, *Computer-assisted Structure Elucidation 2*, Wiley-Interscience, New York, **1986**.
[25] G. Vleduts, *Inf. Storage Retr.* **1963**, *1*, 117.
[26] A. Dengler, E. Fontain, M. Knauer, N. Stein, I. Ugi, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 262.
[27] G. Vleduts, K. A. Finn, *Proc. Dept. of Mechanization and Automation of Information Work*, Acad. Sci. USSR, Moskau, **1960**, p. 66.
[28] E. J. Corey, *Pure Appl. Chem.* **1967**, *14*, 19; see also *Angew. Chem.* **1991**, *103*, 469; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 455.
[29] E. J. Corey, W. T. Wipke, *Science* **1991**, *166*, 178; E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, **1989**.
[30] E. J. Corey, A. K. Long, S. D. Rubenstein, *Science* **1985**, *228*, 4089; see also E. J. Corey, A. K. Long, G. I. Lotto, S. D. Rubenstein, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 304.
[31] W. T. Wipke, D. Rogers, *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 71.
[32] H. Gelernter, N. S. Sridharan, A. J. Hart, S. C. Yen, F. W. Fowler, H. J. Shue, *Top. Curr. Chem.* **1973**, *41*, 113; H. Gelernter, J. R. Rose, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492.
[33] The computer program for planning peptide syntheses, developed by G. Kaufhold and I. Ugi at Bayer AG in Leverkusen (FRG) is described in [34].
[34] I. Ugi, *Rec. Chem. Prog.* **1969**, *30*, 389; *Intra-Sci. Chem. Rep.* **1971**, *5*, 229.
[35] M. Bersohn, *Computer Assisted Organic Synthesis* (*ACS Symp. Ser.* **1977**, *61*); S. Sasaki, K. Funatsu, *Tetrahedron Comput. Methodol.* **1988**, *1*, 39; R. Barone, M. Arbelot, M. Chanon, *ibid.* **1988**, *1*, 3; E. S. Blurock, *ibid.* **1989**, *2*, 207; A. Weise, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 228; I. Dogane, T. Takabatake, M. Bersohn, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 291; L. Baumer, I. Compagnari, G. Sala, G. Sello, *ibid.* **1992**, *111*, 297; P. Hamm, P. Jauffret, G. Kaufmann, *ibid.* **1992**, *111*, 317.
[36] E. Hjelt, *Geschichte der Organischen Chemie*, Vieweg, Braunschweig, **1916**.
[37] a) R. Robinson et al., *J. Chem. Soc.* **1916**, *109*, 1029, 1039; *ibid.* **1917**, *111*, 958; *ibid.* **1918**, *113*, 639; *ibid.* **1919**, *115*, 943; *ibid.* **1921**, *120*, 545; b) Sir R. Robinson, *The Structural Relations of Natural Products*, Clarendon Press, Oxford, **1955**.
[38] S. G. Warren, *Organic Synthesis: The Disconnection Approach*, Wiley, New York, **1982**.
[39] P. H. Winston, *Artificial Intelligence*, Addison-Wesley, Reading, MA, USA, **1984**; B. A. Hohne, T. H. Pierce, *Expert system Applications in Chemistry* (*ACS Symp. Ser.* **1989**, *408*).
[40] H. W. Brown, *Chem. Ind. London* **1988** (5), 43.
[41] G. M. Downs, V. J. Gillet, J. D. Holliday, M. F. Lynch, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 172.
[42] E. J. Corey, G. Petersson, *J. Am. Chem. Soc.* **1972**, *94*, 460.
[43] E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, G. Petersen, *J. Am. Chem. Soc.* **1975**, *97*, 6116.
[44] M. Frèrejaques, *Bull. Soc. Chim. Fr. Mem.* **1939**, *5*, 1008.
[45] R. B. Woodward, *Angew. Chem.* **1956**, *68*, 13.
[46] M. Hicks, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 352.
[47] M. G. Bures, W. L. Jorgensen, *J. Org. Chem.* **1988**, *53*, 2504.
[48] W. Schubert, *MATCH* **1979**, *6*, 213.
[49] G. Grethe, T. E. Moock, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 505.
[50] J. B. Hendrickson, *Top. Curr. Chem.* **1976**, *62*, 49.
[51] G. Moreu, *Nouv. J. Chim.* **1978**, *2*, 187.
[52] J. B. Hendrickson, A. G. Toczko, *Pure Appl. Chem.* **1988**, *60*, 1583; J. B. Hendrickson, *Anal. Chim. Acta* **1990**, *235*, 103; *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 323.
[53] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie, I. Ugi in *Computer Representation and Manipulation of Chemical Information* (Eds.: W. T. Wipke, S. R. Heller, R. J. Feldman, E. Hyde), Wiley, New York, **1974**, p. 129.
[54] J. Gasteiger, C. Jochum, *Top. Curr. Chem.* **1978**, *74*, 93.
[55] J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki, *Top. Curr. Chem.* **1987**, *137*, 19; J. Gasteiger, W. D.-Ihlenfeldt, P. R. Röse, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 270.
[56] R. Doenges, B. T. Groebel, H. Nickelsen, J. Sander, *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 425.
[57] D. A. Evans, *unpublished manuscript*, **1972**.
[58] A. Lapworth, *J. Chem. Soc.* **1921**, *120*, 543.
[59] D. Seebach, *Angew. Chem.* **1979**, *91*, 259; *Angew. Chem. Int. Ed. Engl.* **1979**, *18*, 239.
[60] E. Fontain, J. Bauer, I. Ugi, *Chem. Lett.* **1987**, 37.
[61] E. Fontain, *Tetrahedron Comput. Methodol.* **1990**, *3*, 469.
[62] E. Fontain, *Dissertation*, Technische Universität München, **1987**.
[63] E. Fontain, J. Bauer, I. Ugi, *Z. Naturforsch. B* **1987**, *42*, 297.
[64] D. Hilbert, P. Bernays, *Grundlagen der Mathematik*, Teubner, Berlin, **1934**.
[65] A. Cayley, *Ber. Dtsch. Chem. Ges.* **1875**, *8*, 172; F. Herrmann, *ibid.* **1880**, *13*, 792.
[66] *Chemical Applications of Graph Theory* (Ed.: A. T. Balaban), Academic Press, London, **1976**.
[67] *Chemical Applications of Topology and Graph Theory* (Ed.: R. B. King), Elsevier, Amsterdam, **1983**.
[68] R. E. Merrifield, H. E. Simmons, *Topological Methods in Chemistry*, Wiley, New York, **1989**.
[69] H. R. Henze, C. M. Blair, *J. Am. Chem. Soc.* **1931**, *53*, 3077.
[70] A. Polyá, *Acta Math.* **1937**, *68*, 145.
[71] *The Permutation Group in Physics and Chemistry* (Ed.: J. Hinze), Springer, Heidelberg, **1979**.
[72] A. Kerber, K.-J. Thürlings, *Symmetrieklassen von Funktionen and ihre Abzähltheorie* (Bayreuther Math. Schriften), Bayreuth, **1983**.
[73] J. Dugundji, R. Kopp, D. Marquarding, I. Ugi, *Top. Curr. Chem.* **1978**, *75*, 165.
[74] E. Ruch, W. Hässelbarth, B. Richter, *Theor. Chim. Acta* **1970**, *19*, 288; W. Hässelbarth, E. Ruch, *ibid.* **1973**, *29*, 259; W. Hässelbarth, E. Ruch, D. J. Klein, T. H. Seligman in *Group Theoretical Methods in Physics* (Eds.: R. T. Sharp, B. Kolman), Academic Press, New York, **1977**, p. 617.
[75] J. Dugundji, J. Showell, R. Kopp, D. Marquarding, I. Ugi, *Isr. J. Chem.* **1980**, *20*, 20.
[76] I. Ugi, D. Marquarding, H. Klusacek, G. Gokel, P. D. Gillespie, *Angew. Chem.* **1970**, *82*, 741; *Angew. Chem. Int. Ed. Engl.* **1970**, *9*, 703.
[77] I. Ugi, B. Gruber, N. Stein, A. Demharter, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 485.
[78] J. Dugundji, I. Ugi, *Top. Curr. Chem.* **1973**, *39*, 19.
[79] J. Brandt, J. Bauer, R. M. Frank, A. von Scholley, *Chem. Scr.* **1981**, *18*, 53; J. Brandt, A. von Scholley, *Comput. Chem.* **1983**, *7*, 51.
[80] J. Koča, M. Kratochvil, V. Kvasnička, L. Matyska, J. Pospichal, *Lect. Notes Chem.* **1989**, *51*; E. Hladka, J. Koča, M. Kratochvil, V. Kvasnička, L. Matyska, J. Pospichal, V. Potuček, *Top. Curr. Chem.*, in press.
[81] V. Kvasnička, M. Kratochvil, J. Koča, *Coll. Czech. Chem. Commun.* **1983**, *48*, 2284; J. Koča, *ibid.* **1988**, *53*, 1007, 3108, 3119; *J. Math. Chem.* **1989**, *3*, 73, 91; V. Kvasnička, J. Pospichal, *ibid.* **1989**, *3*, 161; *Int. J. Quantum Chem.* **1990**, *38*, 253; V. Kvasnička, J. Pospichal, *J. Math. Chem.* **1990**, *5*, 309.
[82] I. Ugi, J. Dugundji, R. Kopp, D. Marquarding, *Lect. Notes Chem.* **1984**, *36*.
[83] T. F. Brownscombe, R. V. Stevens, private communication to I. Ugi; T. F. Brownscombe, *Dissertation*, Rice University, Houston, TX, **1972**.
[84] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum in [14], p. 275.
[85] I. Ugi, J. Bauer, K. Bley, A. Dengler, E. Fontain, M. Knauer, S. Lohberger, *J. Mol. Struct. (Theochem)* **1991**, *230*, 73.
[86] A. P. Johnson, C. Marshall, P. N. Judson, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 310; *J. Chem. Inform. Comput. Chem.*, in press; A. P. Johnson, C. Marshall, *ibid.*, in press.
[87] J. B. Hendrickson, C. A. Parks, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 209.
[88] I. Ugi, J. Bauer, R. Baumgartner, E. Fontain, D. Forstmeyer, S. Lohberger, *Pure Appl. Chem.* **1988**, *60*, 1573.
[89] I. Ugi, E. Fontain, M. Knauer, N. Stein, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 262.

[90] N. S. Zefirov, E. V. Godeeva, *Usp. Khim.* **1954**, *54*, 1753; N. S. Zefirov, *Acc. Chem. Res.* **1987**, *20*, 237; S. S. Tratch, I. I. Baskin, N. S. Zefirov, *Zh. Org. Khim.* **1988**, *24*, 1121; N. S. Zefirov, S. S. Tratch, *Anal. Chim. Acta* **1990**, *235*, 115.
[91] L. Matyska, J. Koča, *J. Chem. Inf. Comput. Sci.*, in press.
[92] L. Matyska, J. Koča, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 380.
[93] Z. Hippe in [13], p. 165.
[94] P. Y. Johnson, I. Burnstein, J. Crary, M. Evens, T. Wang, *Expert Systems Application in Chemistry* (*ACS Symp. Ser.* **1988**, 408).
[95] I. Ugi, N. Stein, M. Knauer, B. Gruber, K. Bley, R. Weidinger, *Top. Curr. Chem.* **1993**, *166*, 199; R. Weidinger, *Diplomarbeit*, Universität Passau, **1991**.
[96] J. Gasteiger, P. D. Gillespie, D. Marquarding, I. Ugi, *Top. Curr. Chem.* **1974**, *48*, 1.
[97] E. O. von Lippmann, *Chem.-Ztg.* **1909**, *1*, 1; R. Grund, A. Kerber, R. Laue, *Bayreuther Universitätsblätter*, in press.
[98] I. Ugi, *Chimia* **1986**, *40*, 340.
[99] L. Spialter, *J. Chem. Doc.* **1964**, *4*, 261.
[100] I. Ugi, M. Wochner, E. Fontain, J. Bauer, B. Gruber, R. Karl in *Concepts and Applications of Chemical Similarity* (Eds.: M. A. Johnson, G. M. Maggiora), Wiley, New York, **1990**, p. 239.
[101] J. Dugundji, *Topology*, Allyn and Bacon, Boston, **1966**.
[102] W. Preuss, *Int. J. Quantum Chem.* **1969**, *3*, 123, 131.
[103] M. Wochner, *Dissertation*, Technische Universität München, **1984**.
[104] M. Wochner, I. Ugi, *J. Mol. Struct. (Theochem)* **1988**, *165*, 229.
[105] C. Jochum, J. Gasteiger, I. Ugi, *Angew. Chem.* **1980**, *92*, 503; *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 495; C. Jochum, J. Gasteiger, I. Ugi, J. Dugundji, *Z. Naturforsch. B* **1982**, *37*, 376.
[106] M. Wochner, J. Brandt, A. von Scholley, I. Ugi, *Chimia* **1988**, *42*, 217.
[107] E. Fontain, *Anal. Chim. Acta* **1992**, *265*, 227.
[108] H. Kolbe, J. Liebig, *Ann. Chem. Pharm.* **1850**, *75*, 211; *ibid.* **1850**, *76*, 1.
[109] J. Pospíchal, V. Kvasnička, I. Gutman, *Collect. Sci. Pap. Kragujevac*, **1991**, *12*, 109–125; V. Baláz, J. Pospíchal, V. Kvasnička, *Discrete Appl. Math.* **1992**, *35*, 1–19.
[110] A. P. Johnson, C. Marshall, *J. Chem. Inf. Comput. Sci.*, in press.
[111] E. Fontain, *Dissertation*, Technische Universität München, **1987**.
[112] I. Ugi, J. Brandt, A. von Scholley, S. Minker, M. Wochner, H. Schönmann, B. Straupe, *Hierarchisch strukturierte Speicherung and Ermittlung von chemischen Reaktionen* (Forschungsbericht, Information and Dokumentation, BMFT-FB-ID85-005), FIZ Karlsruhe, **1985**.
[113] J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi in *Computer Assisted Organic Synthesis* (*ACS Symp. Ser.* **1977**, *61*).
[114] J. Bauer, R. Herges, E. Fontain, I. Ugi, *Chimia* **1985**, *39*, 43.
[115] J. Bauer, *Tetrahedron Comput. Methodol.* **1989**, *2*, 269.
[116] J. Bauer, I. Ugi, *J. Chem. Res. Synop.* **1982**, 298; *J. Chem. Res. Miniprint* **1982**, 3101; J. Bauer, *Dissertation*, Technische Universität München, **1981**.
[117] D. Forstmeyer, J. Bauer, E. Fontain, R. Herges, R. Herrmann, I. Ugi, *Angew. Chem.* **1988**, *100*, 1618; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 1558; J. Bauer, E. Fontain, D. Forstmeyer, I. Ugi, *Tetrahedron Comput. Methodol* **1988**, *1*, 129.
[118] R. Herges, *Dissertation*, Technische Universität München, **1984**.
[119] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, *Angew. Chem.* **1979**, *91*, 99; *Angew. Chem. Int. Ed. Engl.* **1979**, *18*, 111.
[120] E. O. Fischer, G. Bürger, *Z. Naturforsch. B* **1961**, *16*, 77; G. Wilke, B. Bogdanović, P. Hardt, P. Heimbach, W. Keim, M. Kröner, W. Oberkirch, K. Tanaka, E. Steinrücke, D. Walter, H. Zimmermann, *Angew. Chem.* **1966**, *78*, 157; *Angew. Chem. Int. Ed. Engl.* **1966**, *5*, 159.
[121] N. Stein, proposed *Dissertation*, Technische Universität München.
[122] E. J. Corey, W. J. Howe, D. A. Pensak, *J. Am. Chem. Soc.* **1974**, *96*, 7724.
[123] W. T. Wipke, T. M. Dyott, *J. Am. Chem. Soc.* **1974**, *96*, 4825, 4834.
[124] S. Hanessian, J. Franco, G. Gagnon, D. Laramée, B. Laroade, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 413.
[125] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie, I. Ugi, *Tetrahedron* **1974**, *30*, 1845.
[126] C. E. Wintner, *Strands of Organic Chemistry*, Holden-Day, San Francisco, **1979**, p. 9.
[127] H. C. Longuet-Higgins, *Mol. Phys.* **1963**, *6*, 445.
[128] S. McLane, G. Birkhoff, *Algebra*, McMillan, New York, **1967**, p. 91; T. A. Whitelaw, *Introduction to Abstract Algebra*, Blackie, Glasgow, **1988**, p. 124.
[129] E. P. Wigner, *Proc. R. Soc. A* **1971**, *322*, 181.
[130] B. Gruber, *Dissertation*, Technische Universität München, **1992**; A. Dietz, *Dissertation*, Technische Universität München, **1993**.
[131] B. Gruber, A. Dietz, I. Ugi, as yet unpublished.
[132] J. E. Hopcroft, J. D. Ullman, *Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie*, 2nd ed., Addison-Wesley, Bonn, **1990**, p. 8.
[133] R. S. Berry, *J. Chem. Phys.* **1960**, *32*, 933.
[134] P. Gillespie, P. Hoffmann, H. Klusacek, D. Marquarding, S. Pfohl, F. Ramirez, E. A. Tsolis, I. Ugi, *Angew. Chem.* **1971**, *83*, 493; *Angew. Chem. Int. Ed. Engl.* **1971**, *10*, 503.
[135] P. Lemmen, R. Baumgartner, I. Ugi, F. Ramirez, *Chem. Scr.* **1988**, *28*, 451.
[136] M. Gielen, N. van Lautem, *Bull. Soc. Chim. Belges* **1970**, *79*, 679; J. I. Musher, *J. Am. Chem. Soc.* **1972**, *94*, 5662.
[137] J. Brocas, M. Gielen, R. Willem, *The Permutational Approach to Dynamic Stereochemistry*, McGraw-Hill, New York, **1983**.
[138] J. Dugundji, D. Marquarding, I. Ugi, *Chem. Scr.* **1976**, *9*, 74.
[139] W. Schubert, I. Ugi, *J. Am. Chem. Soc.* **1978**, *100*, 37; *Chimia* **1979**, *33*, 183.
[140] See [96], Chapter 8.
[141] H. L. Morgan, *J. Chem. Doc.* **1965**, *5*, 107.
[142] N. Stein, I. Ugi, as yet unpublished.
[143] R. S. Cahn, C. K. Ingold, *J. Chem. Soc.* **1951**, 612; R. S. Cahn, C. K. Ingold, V. Prelog, *Experientia* **1966**, *78*, 413; V. Prelog, G. Helmchen, *Angew. Chem.* **1982**, *94*, 614; *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 567; H. Dodziuk, M. M. Mirowicz, *Tetrahedron Asymmetry* **1990**, *3*, 171.
[144] E. Fontain, as yet unpublished.
[145] A. Dengler, I. Ugi, *Comput. Chem.* **1991**, *15*, 103; *J. Chem. Res. Synop.* **1991**, 162; *J. Chem. Res. Miniprint* **1991**, 1279.
[146] J. E. Dubois, *J. Chem. Doc.* **1973**, *13*, 8; R. Attias, J. E. Dubois, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 2.
[147] W. E. Cossum, G. M. Dyson, M. F. Lynch, R. N. Wolfe, *Mimeographed Report*, Chemical Abstracts Service, **1963**.
[148] E. H. Sussenguth, Jr., *J. Chem. Doc.* **1965**, *5*, 36.
[149] G. W. Adamson, J. Cowell, M. F. Lynch, A. H. W. McLure, W. G. Town, A. M. Yapp, *J. Chem. Doc.* **1973**, *13*, 153.
[150] J. Friedrich, *Dissertation*, Technische Universität München, **1979**; J. Friedrich, I. Ugi, *J. Chem. Res. Synop.* **1980**, 70; *J. Chem. Res. Miniprint* **1980**, *1301*.
[151] M. G. Hicks, C. Jochum, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 191.
[152] A. V. Aho, J. E. Hopcroft, P. Ullma, *The Design and Analysis of Computer Algorithms*, Addison-Wesley, Reading, **1974**.
[153] W. D. Ihlenfeldt, *Dissertation*, Technische Universität München, **1991**.
[154] K. Bley, *Dissertation*, Technische Universität München, **1991**; E. Fontain, *Diplom thesis*, Technische Universität München, **1986**; K. Bley, I. Ugi, as yet unpublished.
[155] M. Knauer, *Dissertation*, Technische Universität München, **1992**.
[156] H. Thalhammer, *Diplom thesis*, Technische Universität München, **1991**.
[157] B. Raspel, G. Roden, B. Woost, C. Zirz in *Software-Entwicklung in der Chemie 3* (Ed.: G. Gauglitz), Springer, Heidelberg, **1989**, p. 33; W. T. Donner, in [13], p. 9.
[158] E. Meyer, E. Sens, *Anal. Chim. Acta* **1988**, *210*, 135.
[159] P. Bruck in *Chemical Structures - The International Language of Chemistry* (Ed.: W. A. Warr), Springer, Heidelberg, **1988**, p. 113.
[160] M. Hicks, C. Jochum, *Anal. Chim. Acta* **1990**, *235*, 87; M. G. Hicks in *Software-Entwicklung in der Chemie 3* (Ed.: G. Gauglitz), Spinger, Heidelberg, **1989**, p. 9.
[161] G. Klopman, *J. Am. Chem. Soc.* **1984**, *106*, 7315; *Mutat. Res.* **1984**, *126*, 139, 227; *J. Comput. Chem.* **1985**, *6*, 28.
[162] G. Klopman, *J. Math. Chem.* **1991**, *7*, 187.
[163] M. Knauer, I. Ugi, as yet unpublished.
[164] A. Dengler, I. Ugi, *J. Math. Chem.* **1992**, *9*, 1.
[165] I. Ugi, G. Kaufhold, *Justus Liebigs Ann. Chem.* **1967**, *709*, 11.
[166] I. Ugi, *Angew. Chem.* **1962**, *74*, 9; *Angew. Chem. Int. Ed. Engl.* **1962**, *1*, 8.
[167] I. Ugi, S. Lohberger, R. Karl in *Comprehensive Organic Synthesis: Selectivity of Synthetic Efficiency 2* (Eds.: B. M. Trost, C. H. Heathcock), Pergamon Press, Oxford, **1991**, Chapter 4.6.
[168] G. A. Blakley, *J. Chem. Educ.* **1982**, *59*, 728; C. Behnke, J. Bargon, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 228.
[169] J. Bauer, E. Fontain, I. Ugi, *Anal. Chim. Acta* **1988**, *210*, 123.
[170] R. E. Burkhard, *Methoden der ganzzahligen Optimierung*, Springer, Wien, **1972**; R. E. Burkhard, K. H. Stratmann, *Nav. Res. Logistics Qu.* **1978**, *25*, 129.
[171] D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley, Reading, **1989**.
[172] *Modern Approaches to Chemical Reaction Searching* (Ed.: P. Willett), Gower, Aldershot, **1986**; D. Bawden, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 212.
[173] A. P. Johnson, *Chem. Br.* **1985**, *21*, 59; E. Zass, S. Müller, *Chimia* **1986**, *40*, 38; J. H. Borkent, F. Oukes, J. H. Noordik, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 148.
[174] W. T. Wipke, G. Vladutz, *Tetrahedron Comput. Methodol.* **1990**, *3*, 83.
[175] C. Tonnelier, P. Jauffret, T. Hanser, G. Kaufmann, *Tetrahedron Comput. Methodol.* **1990**, *3*, 351.
[176] K. Bley, J. Brandt, A. Dengler, R. Frank, I. Ugi, *J. Chem. Res.Synop.* **1991**, 261; *J. Chem. Res. Miniprint* **1991**, 2601.
[177] I. Ugi, J. Bauer, E. Fontain in *Personal Computers for Chemists* (Ed.: J. Zupan), Elsevier, Amsterdam, **1990**, p. 135.
[178] E. Fontain, K. Reitsam, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 96.
[179] We have access to an unpublished program developed by J. Bauer, which generates r-matrices from their defined mathematical properties.
[180] I. Ugi, *Chem. Scr.* **1986**, *26*, 205.
[181] J. Bauer, E. Fontain, I. Ugi, *MATCH* **1992**, *27*, 31.
[182] I. Ugi, *Proc. Estononian Acad Sci.* **1989**, *38*, 225; *ibid.* **1990**, *39*, 193.

[183] E. Fontain, *J. Chem. Inf. Comput. Sci.*, **1992**, *32*, 748.
[184] I. Ugi, J. Bauer, K. Bley, A. Dengler, E. Fontain, M. Knauer, S. Lohberger, *J. Mol. Structure (Theochem)* **1991**, *230*, 73.
[185] S. Lohberger, E. Fontain, I. Ugi, G. Müller, J. Lachmann, *New J. Chem.* **1991**, *15*, 913; *Acta Crystallogr. Sect. C* **1991**, *47*, 2444.
[186] R. E. Valdés-Pérez, *Tetrahedron Comput. Methodol.* **1990**, *5*, 277.
[187] R. Huisgen *Angew. Chem.* **1963**, *75*, 604, 742; *Angew. Chem. Int. Ed. Engl.* **1963**, *2*, 565, 633; *J. Org. Chem.* **1976**, *41*, 403.
[188] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859; L. J. McBride, M. H. Caruthers, *ibid.* **1983**, *24*, 245.
[189] D. Marquarding, F. Ramirez, I. Ugi, P. D. Gillespie, *Angew. Chem.* **1973**, *85*, 99; *Angew. Chem. Int. Ed. Engl.* **1973**, *12*, 91.
[190] F. Westheimer, *Acc. Chem. Res.* **1968**, *11*, 70.
[191] F. Ramirez, S. Glaser, P. Stern, P. D. Gillespie, *Angew. Chem.* **1973**, *85*, 39; *Angew. Chem. Int. Ed. Engl.* **1973**, *12*, 66; F. Ramirez, J. F. Marecek, I. Ugi, *Synthesis* **1975**, 99.
[192] W. Richter, R. Karl, I. Ugi, *Tetrahedron* **1990**, *46*, 3167.
[193] F. Ramirez, I. Ugi, *Phosphorus Sulfur* **1976**, *1*, 231.
[194] I. Ugi, P. Jacob, B. Landgraf, C. Rupp, P. Lemmen, U. Verfürth, *Nucleosides Nucleotides* **1988**, *7*, 605; I. Ugi, N. Bachmeier, R. Herrmann, P. Jacob, R. Karl, P. Lemmen, W. Richter, U. Verfürth, *Phosphorus Sulfur Silicon* **1990**, *51/52*, 57.
[195] P. Jacob, W. Richter, I. Ugi, *Liebigs Ann. Chem.* **1991**, 519; R. Karl, P. Lemmen, W. Richter, I. Ugi, B. Werner, *Synthesis*, submitted.
[196] W. N. Lipscomb, *Boron Hydrides*, Benjamin, New York, **1966**, p. 57.
[197] B. Brellochs, H. Binder, *Angew. Chem.* **1988**, *100*, 270; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 262.
[198] R. E. Carhart, D. H. Smith, H. Brown, C. Djerassi, *J. Am. Chem. Soc.* **1975**, *97*, *5755*.
[199] R. E. Carhart, D. H. Smith, N. A. B. Gray, J. G. Nourse, C. Djerassi, *J. Org. Chem.* **1981**, *46*, 1708.
[200] V. V. Serov, L. A. Gribov, M. E. Elyashberg, *J. Mol. Struct. (Theochem)* **1985**, *129*, 183.
[201] A. Kerber, R. Laue, D. Moser, *Anal. Chim. Acta* **1990**, *235*, 221.
[202] H. Huixiao, X. Xinguan, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 116.
[203] S. Bohanec, J. Zupan, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 531.
[204] I. Ugi, E. Wischhöfer, *Chem. Ber.* **1962**, *95*, 136.
[205] D. H. R. Barton, *Aldrichimica Acta* **1990**, *23*, 3.
[206] R. Herges, *Tetrahedron Comput. Methodol.* **1988**, *1*, 15; K. M. T. Yamada, M. W. Markus, G. Winnewesser, W. Joentgen, R. Kock, E. Vogel, H.-J. Altenbach, *Chem. Phys. Lett.* **1989**, *160*, 113.
[207] R. Herges, I. Ugi, *Angew. Chem.* **1985**, *97*, 596; *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 594; *Chem. Ber.* **1986**, *119*, 829.
[208] F. W. Fowler, *Angew. Chem.* **1971**, *83*, 148; *Angew. Chem. Int. Ed. Engl.* **1971**, *10*, 135.
[209] S. M. McElvain, *Org. React. (NY)*, **1948**, *4*, 256.
[210] M. Stoll, *Helv. Chim. Acta* **1947**, *30*, 1837.
[211] R. Herges, *J. Chem. Inf. Comput. Sci.* **1990**, *3*, 377.
[212] R. Herges, C. Hoock, *Science* **1992**, *255*, 711.
[213] R. Herges in *Proc. 4th Chem. Congr. North Am.*, Division of Chemical Information, New York, **1991**, in press.
[214] R. Herges, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 37.
[215] J. Arct, U. H. Brinker, *Methoden Org. Chem. (Houben-Weyl) 4th ed. 1952–Vol. E19b*, **1989**, p. 337.
[216] H. Dürr, E. M. Frühauf, *Methoden Org. Chem. (Houben-Weyl) 4th ed. 1952–Band E19b*, **1989**, p. 775.
[217] W. von E. Doering, J. F. Coburn, Jr., *Tetrahedron Lett.* **1965**, 991.
[218] C. Wentrup, R. Harder, *Top. Curr. Chem.* **1976**, *62*, 173.
[219] G. B. Fisher, J. J. Juarez-Brambila, C. T. Goralski, W. T. Wipke, B. Singaram, *J. Am. Chem. Soc.*, submitted.
[220] G. Augelman, H. Fritz, G. Rihs, J. Streith, *J. Chem. Soc. Chem. Commun.* **1982**, 1112.
[221] A. Defoin, G. Augelman, H. Fritz, G. Geoffroy, C. Schmidlin, J. Streith, *Helv. Chim. Acta* **1985**, *68*, 1998.
[222] W. Ried, R. Dietrich, *Angew. Chem.* **1963**, *75*, 476; *Angew. Chem. Int. Ed. Engl.* **1963**, *2*, 495.
[223] J. Metzger in *Comprehensive Heterocyclic Chemistry* 6 (Eds.: A. R. Katritzky, C. W. Rees, K. T. Potts), Pergamon Press, Oxford, **1984**, p. 294.
[224] A. Baretta, B. Waegell, *React. Intermed. (Plenum)* **1982**, *3*, 527.
[225] K. Shishido, E. Shitara, K. Fukumoto, T. Kametani, *J. Am. Chem. Soc.* **1985**, *107*, 5810.
[226] J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1961**, *94*, 217; J. Oró, S. S. Amat, *Nature* **1961**, *190*, 442; R. E. Dickerson, *Sci. Am.* **1978**, *239*, (9) 70; *Spektrum Wiss.* **1979** (9) *98*; see also S. Drenkard, J. Ferris, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1373.