resonant sextet benzenoids, and further work is underway to understand fully the congruent relationships that exist between these two divergent and important subsets of benzenoids.

More than 16 total resonant sextet benzenoids and 16 isoskeletal analogues have been synthesized and characterized in the literature.[1,4] Among these more stable compounds, some as large as $C_{48}H_{24}$, $C_{48}H_{22}$, and $C_{60}H_{30}$ ($C_{56}H_{26}N_4$) are represented. Should an analytical chemist determine, by GC-MS, for example, that an environmental sample contained a benzenoid with a molecular formula of $C_{54}H_{22}$, he would still be faced with determining the correct structure from a group of about 10 000 nonradical isomers. The four most stable structures from this group have in fact been identified,[4] and targeted synthesis of these compounds may prove to be worthwhile. If the same sample contained a compound with the molecular formula $C_{54}H_{18}$, its structure would be immediately apparent because this molecular formula corresponds to the third generation of the strictly peri-condensed one-isomer series.[3] Thus Table PAH6, together with these enumeration results, defines the epistemology of the field of benzenoid hydrocarbons and assists in an evaluation of how much is known and can be known about these compounds. The largest characterizable benzenoids will no doubt belong to those constant-isomer series that have just a few members.

Recently Kirby[16] has corroborated our enumeration[1,4] of total resonant sextet benzenoid isomers and this work greatly extends these earlier results. For examples of work in which

benzenoid resonant sextets are treated mainly as mathematical entities, the reader is referred to the work of Hosoya.[6]

REFERENCES

(1) Dias, J. R. *Thermochim. Acta* **1987**, *122*, 313; *J. Mol. Struct.: THEOCHEM.* **1989**, *185*, 57; **1990**, *207*, 141. Clar, E. *The Aromatic Sextet*; Wiley: New York, 1972.
(2) Dias, J. R. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 61.
(3) Dias, J. R. *Theor. Chim. Acta* **1990**, *77*, 143; *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 251. Hall, G. G. *Theor. Chim. Acta* **1988**, *73*, 425.
(4) Dias, J. R. *Handbook of Polycyclic Hydrocarbons*; Elsevier: Amsterdam, 1987, 1988; Parts A and B.
(5) Knop, J.; Müller, W.; Szymanski, K.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; SKTH/Kemija u Industriji: Zagreb, 1985. Stojmenovic, I.; Tosic, R.; Doroslovacki, R. In *Graph Theory*, Proceedings of the Sixth Yugloslav Seminar on Graph Theory, Dubrovnik, 1985; Tosik, R., Aceta, D., Petrovic, V., Eds.; University of Novi Sad; Novi Sad, 1986. Brunvoll, J.; Cyvin, S. J. *Z. Naturforsch.* **1990**, *45A*, 69.
(6) Ohkami, N.; Hosoya, H. *Theor. Chim. Acta* **1983**, *64*, 153. Hosoya, H. *Top. Curr. Chem.* **1990**, *153*, 255.
(7) Stein, S. E. *J. Phys. Chem.* **1978**, *82*, 566.
(8) Sullivan, R. F.; Boduszynski, M.; Fetzer, J. C. *Energy Fuels* **1989**, *3*, 603.
(9) Qain, S.; Xiao, Y.; Gu, F. *Fuel* **1987**, *66*, 242.
(10) Lewis, I. C.; Petro, B. *J. Polym. Sci.* **1976**, *14*, 1975.
(11) Gerhardt, P.; Homann, K. H. *J. Phys. Chem.* **1990**, *94*, 5381.
(12) Pikus, A.; Feigelman, V.; Mezheritskii, V. *Zh. Org. Khim.* **1989**, *25*, 2603.
(13) Dias, J. R. *Nouv. J. Chim.* **1985**, *9*, 107; *Acc. Chem. Res.* **1985**, *18*, 241.
(14) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. *J. Comput. Chem.* **1986**, *7*, 547.
(15) Dias, J. R.; Cyvin, S. J.; Brunvoll, J. Manuscript in preparation.
(16) Kirby, E. C. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 447.

The Generation of Reaction Networks with RAIN. 1. The Reaction Generator[†]

ERIC FONTAIN* and KLAUS REITSAM

Technical University of Munich, Institute for Organic Chemistry, Lichtenbergstrasse 4, D-8046 Garching, FRG

The formal program system RAIN is described. It generates the conceivable reaction paths that join known educts with known products of a chemical reaction. The reaction generator of this system is not based on known reactions but operates in a deductive manner. It is guided by a small number of constraints that describe the maximum complexity of a reaction step and introduce rules for the constitutional features of the generated intermediates. The chemical behavior of the elements considered is taken into account by means of transition tables representing the valence schemes and their allowed interconversions.

INTRODUCTION

This is the first of a series of articles concerning the program system RAIN. The acronym RAIN stands for **R**eactions **A**nd **I**ntermediates **N**etworks. RAIN[1] is a general system for the formal generation of reaction networks. The main task of RAIN is the production of *all formally possible reaction paths* that connect given starting materials and products of a chemical reaction, and comply with user-defined constraints. Depending on the chosen boundary conditions, the reaction paths can either form a network of mechanistic elementary reaction steps or a network of more complex "synthetic" reactions. For the elaboration of the reactions and their intermediates, RAIN uses neither a knowledge base (e.g., transforms[2]) nor a fixed set of formal transformations (e.g., a fixed set of R matrices,[3] base transformations,[4] or half reactions[5]). On the contrary, RAIN supplies the user with an extensive set of optional rules for the description of allowed

reactions and intermediates and generates within the user-defined boundary conditions a combinatorially complete set of solutions. The values of these constraints change according to the problem to be solved. They determine, e.g.,

- the maximum complexity (number of reacting atoms and bonds) of a reaction step
- the atomic valence schemes of the intermediates (and their interconvertibility)
- constitutional properties of the intermediates (rings, etc.)

In essence, RAIN consists of a formal reaction generator and a network management system. The reaction generator produces for a given molecule those structures that may be obtained by a chemical reaction from the molecule considered under the given boundary conditions. All generated reactions and intermediates are stored in a network file. Analogously, the reaction generator of RAIN can produce all conceivable precursors of a molecule. The successive application of this procedure to "both ends" of a chemical reaction or sequence of reactions leads to two growing "trees" of reaction pathways,

RAIN REACTION GENERATOR

J. Chem. Inf. Comput. Sci., Vol. 31, No. 1, 1991 **97**

E : starting materials   ● : intermediates
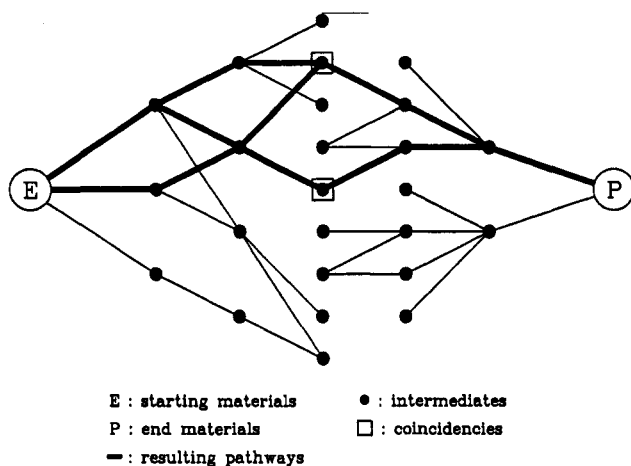P : end materials   □ : coincidencies
— : resulting pathways

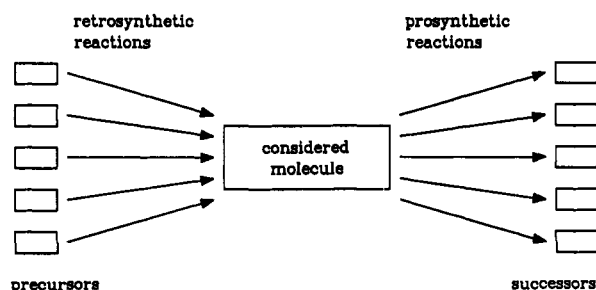**Figure 1.** Bilateral generation of reaction pathways by RAIN.

Figure 2. Applications of reaction generators.

one evolving from the starting materials, the other emerging from the products of the investigated reaction (see Figure 1). If these trees hit the same molecules, these molecules join the trees, and thus reaction pathways are found that connect the beginning and the end of the reaction. Such a bilateral procedure reduces the number of otherwise generated intermediates very effectively. This is importance because, as encountered in neary all synthesis planning or reaction simulation programs, the trees of possible pathways grow very fast with increasing depth.

RAIN can help to solve mechanistic as well as synthetic problems[6,7] by bilaterally generating reaction paths, but it also can be used in the field of monolateral reaction prediction and retrosynthetic synthesis design.

The following sections describe the system's components, beginning with the reaction generator, which represents the "engine" of the program.

WHAT IS A REACTION GENERATOR?

The main operating principle of each system for reaction simulation or synthesis planning is a reaction generator. In general terms, this is a device that generates all molecules (or ensemble of molecules[8]) which can emerge from a given molecule (or ensemble of molecules) in a chemical reaction (prosynthesis) or can be immediate precursors thereof (retrosynthesis) (see Figure 2).

There are reaction generators that are capable only of working in one single direction (prosynthetically or retrosynthetically). However, there are some reaction generators, like the one shown in Figure 2, that can perform both tasks. We call procedures that generate reactions (and thus produce molecules) from a considered molecule (either prosynthetically, retrosynthetically, or both) "reaction generators of type I".[8] An alternative type, called "reaction generator of type II",[9] produces reactions according to an hierarchic classification of reactions independent of starting or target materials (e.g., reactions in a given R matrix: IGOR[10]). The SYMBEQ[11] system contributes to this class in a very similar manner.

Figure 3. Transitions table for carbon (example).

Reaction generators of type II are less important for reaction simulating or synthesis planning programs.

The reaction generators of type I can further be subdivided in two classes:

·knowledge-based systems
·formal systems

Knowledge-based systems depend mostly on a collection of generalized reactions (so-called transforms) that have to be extracted from literature by hand or that are elaborated automatically from reaction data bases according to specific rules.

Formal systems solve problems by applying general rules in a deductive manner, and thus generate all possible solutions that observe these rules. The formal approach is advantageous if unconventional problems are to be solved, because it is more creative and is not confined to combining already known reactions. The potential weakness of those systems is often the combinatorial wealth of the solutions obtained. This can be overcome by a skilful use of constraints and by a high degree of interactivity, which makes the process of problem solving transparent to the user.

The boundary conditions that control the reaction generator of RAIN are described in the following sections.

VALENCE SCHEMES AND TRANSITION TABLES

In the course of a chemical reaction, the behavior of an atom is strongly influenced by its chemical nature (type of element) and its embedding within the molecule (number and type of bonds and bonded neighbors). The valence scheme of an atom describes the distribution of its valence electrons in terms of bonds of distinct orders and of lone electrons. Note that the valence scheme of an atom participating in a chemical reaction may change or may be invariant.

The interconvertibility of one valence scheme of an atom into another valence scheme is one of the main guiding mechanisms of the reaction generator in RAIN. The user describes each element that occurs in the problem investigated by a list of allowable valence schemes. Furthermore, this list indicates the allowable transitions that these valence schemes can undergo in the course of a reaction step. The allowed valence schemes are related to the rows i and the columns j of a transition table TT. A table entry tt_{ij} determines whether the reaction generator may transform the valence scheme i into the valence scheme j. This applies when the reaction generator produces reactions in a forward direction (prosynthesis). In order to generate reactions in a reverse direction (retrosynthesis), the reaction generator reads the transition tables from the columns to the rows. The exemplary transition table for a carbon atom shown in Figure 3 may serve to illustrate the concept.

Figure 3 is a typical example for a transition table as it could be used in a mechanistic study. The transition of the carbocation into itself (3rd row, 3rd column) is not allowed, which expresses the instability of this species. According to this transition table, a carbocation may be formed from other valence schemes, but must be transformed into an uncharged
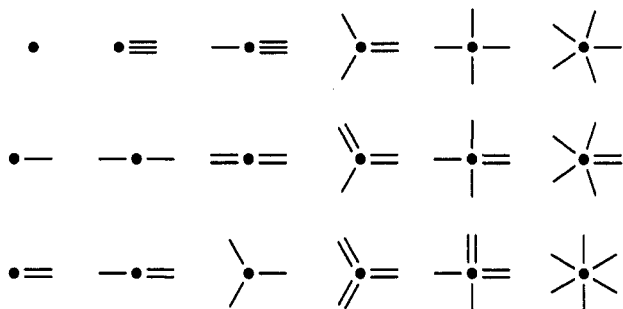
98 *J. Chem. Inf. Comput. Sci., Vol. 31, No. 1, 1991*

FONTAIN AND REITSAM

Figure 4. Bond schemes used within RAIN.

carbon atom immediately in the following reaction step.

The transition tables play an important role as a means for describing the "horizon of chemical events", as well as an operational basis for the reaction generator, as described in the following sections. A description of the algebraic properties of valence schemes was given by Koca et al.[12]

BOND SCHEMES

A bond scheme describes the way an atom is imbedded into a molecule, neglecting the information about the lone electrons, the formal charge, and the type of the chemical element. It represents a higher degree of abstraction than the valence schemes because one bond scheme describes the bonding conditions for a series of elements respective valence schemes. Thus the bond scheme


is found in the following valence schemes:
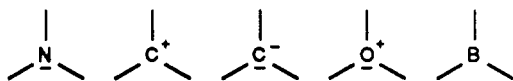
The maximum number of different bond schemes occurring in organic chemistry is limited. It is only dependent on the maximum coordination number and valence of all considered elements. Figure 4 shows the complete list of all bond schemes that can be handled by RAIN. These are all combinations of single, double, and triple bonds with a maximum connectivity of 6. Combinations with more than one triple bond, as well as combinations with a triple bond and a double bond or a triple bond with more than one single bond, are not considered. Note that the reaction-generating process is by no means limited to some preselected bond schemes. The above selection was only made for convenience and could easily be expanded to handle, e.g., special bonding conditions in inorganic chemistry.

BR TRANSFORMERS

The transition between two bonds schemes of an atom during a chemical reaction can generally be achieved in several ways. Thus the change from
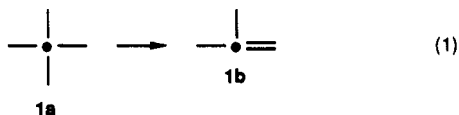
could be achieved by, e.g., the following transition rule:
·*break* one single bond
·*transform* one single bond into a double bond
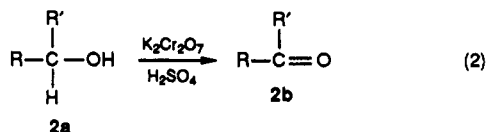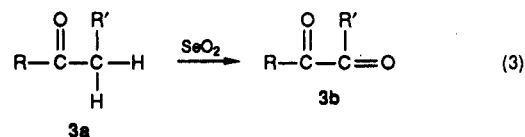as realized, e.g., at the carbon atom in

Table I. BR Transformers for **1a** → **1b**

b_1	1	1					
r_1	-1	+1					
b_2	0	1	1				
r_2	+2	-1	-1				
b_3	0	1	1	1			
r_3	+1	-1	-1	+1			
b_4	0	0	1	1	1		
r_4	+1	+2	-1	-1	-1		
b_5	0	0	1	1	1	1	
r_5	+1	+1	-1	-1	-1	+1	
b_6	0	0	0	1	1	1	1
r_6	+1	+1	+2	-1	-1	-1	-1

But also the transition rule:
·*break* two single bonds
·*make* a new double bond
as realized, e.g., at the α-carbon atom in:

transforms bond scheme **1a** into **1b**.

The different possibilities to achieve one bond scheme transition are formally distinguished by their transition rules. These describe how many bonds and which bond orders are affected. We call such a transition rule a BR transformer. It consists of two components:
·a list of bond orders b
·a list of changes r to these bond orders

The bond scheme transition **1a** → **1b**, e.g., can be obtained by the following BR transformers:

Table I shows the complete list of all formally possible BR transformers that produce the bond scheme transition **1a** → **1b**.

The complete list of *all* BR transformers of *all* 324 transitions for the 18 bond schemes in Figure 4 serves as a basis for the reaction generator in RAIN. Note that this list depends only on the fundamental choice of allowed bond schemes, and thus has to be generated only once, independently from the problem to be investigated.

BOUNDARY CONDITIONS

The fundamental operating principle of the reaction generator is the following:

> First determine the bond schemes of the considered molecule.
> Then generate for each atom a list of BR transformers that comply with the user-defined transition tables and apply these in all combinatorally possible ways.

This basic procedure only leads to a reasonable set of reactions if it is limited numerically by boundary conditions. These constraints must be set by the user and thus can be accurately tuned according to the investigated problem. Table II shows a list of the main boundary conditions with some typical numerical values.

The combinatorial generating procedure is conducted by the boundary conditions in such a manner that each limiting value takes effect as early as possible in the backtracking algorithms.

Table II. Boundary Conditions for Reaction Generator in RAIN

no.	boundary condition	typical value
1	max no. of reacting atoms	2–6
2	max no. of atoms changing adjacency	2–6
3	max no. of changing adjacencies at one atom	1–2
4	max no. of atoms changing lone electrons	1–2
5	max change of lone electrons	2
6	max no. of reacting bonds	2–6
7	max no. of changing adjacencies	2–6
8	max no. of changing bonds at one atom	1–2
9	max change of bond order	1–2
10	max no. of charged atoms	0–4
11	max no. of charged atoms in one molecule	0–2
12	max net charge (absolute value)	0–1
13	max net charge of one molecule (absolute value)	0–1
14	max no. of rings	0–6
15	max no. of n-membered rings (n = 3–9)	0–6
16	min ring size for triple bonds within rings	10
17	min ring size for cumulated bonds within rings	10
18	max no. of separate molecules	1–4

This avoids unnecessary combinations and permutation steps, and thus considerably enhances the performance of the generator.

THE REACTION GENERATOR

The following procedure generates all molecules that can be obtained from a given molecule under the user-defined constraints.

step 1   Read starting molecule and determine the *number of atoms* (→NATOM).

step 2   Determine the *valence schemes* of all atoms.

step 3   Determine for all atoms the valence schemes into which the actual valence schemes may be transformed according to the *transition tables*. Check for *boundary conditions nos. 4 and 5* (see Table II).

step 4   Reduce the valence schemes to *bond schemes* by discarding the formal charges, element types, and lone electrons.

step 5   Create for each atom a list of all *BR transformers* that can induce the allowed bond scheme transitions. Check for *boundary conditions nos. 3, 8, and 9* (see Table II).

step 6   Set IATOM = 0.

step 7   Set IATOM = IATOM + 1.

step 8   Are there any *unprocessed BR transformers* for atom IATOM?
Yes → continue with step 11

step 9   Set IATOM = IATOM − 1.
Is IATOM = 0?
No → continue with step 15

step 10   *All* molecules (reactions) have been *generated* → STOP!

step 11   Does the application of the next unprocessed BR transformer violate any of the *boundary conditions nos. 1, 2, 4, 6, or 7* (see Table II)?
Yes → continue with step 8

step 12   Try to *remove* all elements from the new BR transformers that are used for bond changes to atoms <IATOM.
Does the BR transformer contain *enough elements* for this reduction?
No → continue with step 8

step 13   Are there any *remaining BR elements* after reducing the BR transformer?
No → continue with step 7

step 14   Generate *all possibilities of application* of the BR transformer onto atoms > IATOM.

Application means: relate all remaining BR elements to atoms > IATOM, whose bonding conditions to atom IATOM correspond to the required bond order of the BR elements.

step 15   Are there any *unprocessed applications* of the actual BR transformer for atom IATOM?
No → continue with step 8

step 16   Generate *new bonding conditions* for atom IATOM to atoms > IATOM by applying the BR transformer

step 17   Is IATOM = NATOM?
No → continue with step 7

step 18   The new *molecule is complete*, now check for constitutional *boundary conditions nos. 10–18* and store the molecule in the file.
Redundancies are discarded by applying a procedure for the unique description of molecules.[13]

step 19   Continue with step 15.

EXAMPLES

The following examples should illustrate the capabilities of the reaction generator and the effects of varying constraints on the results. Note that in many cases the generation of improbable structures and reactions can be suppressed by the application of further rules that are the subject of some of the following articles in this series (e.g., Bredt's rule, "forbidden" substructures, etc.).

Example 1 (Initial Step of the Streith–Defoin Reaction[1b,14]).
Boundary conditions (see Table II):

no. 1: 6	no. 2: 4
no. 3: 2	no. 4: ∞
no. 5: ∞	no. 6: 6
no. 7: 2	no. 8: 2
no. 9: 1	no. 10: ∞
no. 11: ∞	no. 12: ∞
no. 13: ∞	no. 14: 6
no. 16: ∞	no. 17: ∞
no. 18: ∞	

no. 15: 3- and 4- membered rings are not allowed

Valence schemes:

Initial molecule(s):

Directly resulting molecules:

Example 2 (Carbene Insertion with up to Three Changing Adjacencies).

Boundary conditions (see Table II):

no. 1: 6	no. 2: 6
no. 3: 2	no. 4: ∞
no. 5: ∞	no. 6: 6
no. 7: 3	no. 8: 3
no. 9: 2	no. 10: ∞
no. 11: ∞	no. 12: ∞
no. 13: ∞	no. 14: 6
no. 16: 0	no. 17: 0
no. 18: 2	

no. 15: no restrictions on rings

Valence schemes:
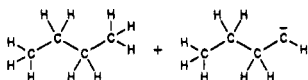
Initial molecule(s):

Directly resulting molecules:
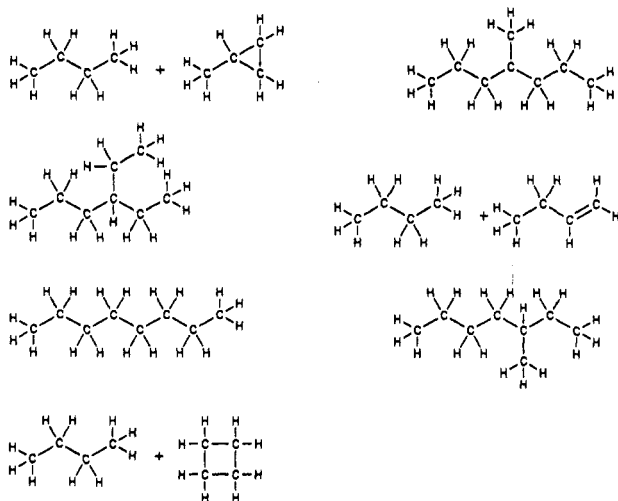
Example 3 (Carbene Insertion with up to Four Changing Adjacencies).

Boundary conditions (see Table II):
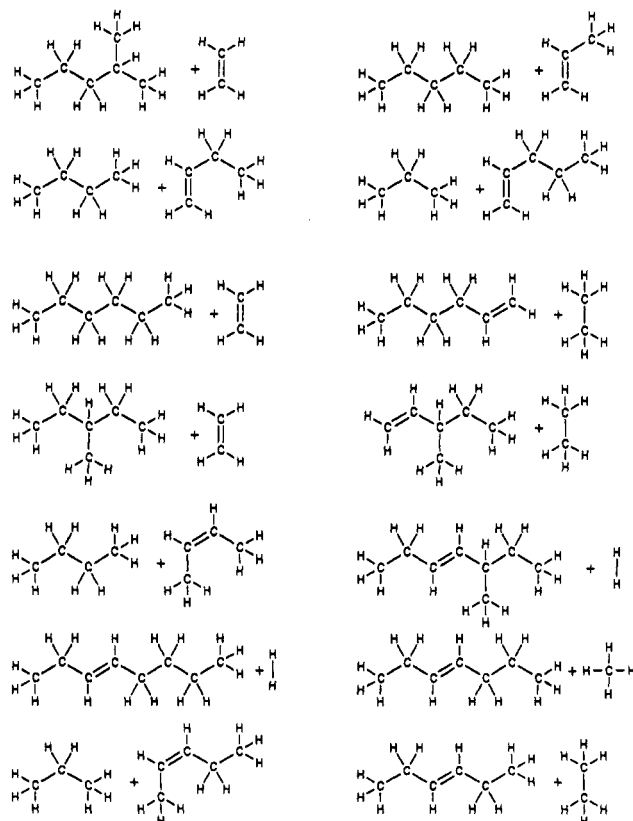
same as in example 2 except no. 7: 4

Valence schemes:

same as in example 2

Initial molecule(s):

same as in example 2

Directly resulting molecules:
(in addition to those of example 2)

CONCLUSION

The new reaction generator implemented in RAIN serves as a versatile tool for the production of reaction pathways. Its capability of generating any thinkable chemical reaction makes it indispensable when searching for the solution of unconventional problems.

The program system RAIN is implemented on MS-DOS PC. It is supplied with a mouse-driven graphic-user interface and is interactive to a very high degree. This type of implementation guarantees good transparency and enhances the user impact on the strategy of problem solving. RAIN is no expert system that intends to replace an expert, but it supplies the expert user with powerful tools that support him, extend his range of competence, and widen his horizon in many ways.

REFERENCES AND NOTES

(1) (a) Fontain, E.; Bauer, J.; Ugi, I. Computer-Assisted Bilateral Generation of Reaction Networks from Educts and Products. *Chem. Lett.* **1987**, 37–40. (b) Fontain, E.; Bauer, J.; Ugi, I. Computerunterstützte mechanistische Analyse de Streigh-Reaktion mit dem Programm RAIN. *Z. Naturforsch.* **1987**, *42B*, 889–891. (c) Forstmeyer, D.; Bauer, J.; Fontain, E.; Herges, R.; Herrmann, R.; Ugi, I. Die Umsetzung von Tropon mit einem Homopyrrol als Ergebnis einer computergestützten Suche nach präzedenzlosen chemischen Reaktionen. *Angew Chem.* **1988**, *100*, 1618–1619; *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1558–1559. (d) Bauer, J.; Fontain, E.; Ugi, I. Computer-Assisted Bilateral Solution of Chemical Problems And Generation of Reaction Networks. *Anal. Chem. Acta* **1988**, *210*, 123–134. (e) Ugi, I.; Bauer, J.; Baumgartner, R.; Fontain, E.; Forstmeyer, D.; Lohberger, S. Computer Assistance in the Design of Syntheses and a New Generation of Computer Programs for the Solution of Chemical Problems by Molecular Logic. *Pure Appl. Chem.* **1988**, *60*, 1573–1586. (f) Ugi, I.; Bauer, J.; Fontain, E. Molecular Logic and the Deductive Solution of Chemical Problems. *Wiss. Z. Tech. Hochsch. Leuna*-Merseburg **1989**, *31*, 9–17. (g) Fontain, E.; Bauer, J.; Ugi, I. Reaction Pathways on a PC. In *PCs for Chemists*; Zupan, J., Ed.; Elsevier: Amsterdam, 1990; Vol. 5, pp 135–154. (h) Ugi, I.; Wochner, M.; Fontain, E.; Bauer, J.; Gruber, B.; Karl, R. Chemical Similarity, Chemical Distance, and Computer Assisted Formalized Reasoning by Analogy. In *Concepts and Applications of Chemical Similarity*; Johnson, M., Maggiora, G., Eds.; Wiley: New York, 1990; in press. (i) Ugi, I.; Fontain, E.; Bauer,

J. Transparent Formal Methods for Reducing the Combinatorial Wealth of Conceivable Solutions to a Chemical Problem—Computer-Assisted Elucidation of Complex Reaction Mechanisms. *Anal. Chim. Acta* **1990**, *235*, 155–161.

(2) (a) Corey, E. J.; Long, A. K.; Rubenstein, S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408–418. (b) Gund, P.; Grabowski, E. J. J.; Hoff, D. R.; Smith, G. M.; Andose, J. D.; Rhodes, J. B.; Wipke, W. T. Computer-Assisted Synthetic Analysis at Merck. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 88–93.

(3) Gasteiger, J.; Hutchings, M. G.; Christoph, B.; Gann, L.; Hiller, C.; Löw, P.; Marsili, M.; Saller, H.; Yuki, K. A New Treatment of Chemical Reactivity: Development of EROS, an Expert System for Reaction Prediction and Synthesis Design. *Top. Curr. Chem.* **1987**, *137*, 19–73.

(4) Schubert, W. ASSOR—Allgemeines Simulationssystem Organischer Reaktionen. *Informal Commun. Math. Chem. (Match)* **1979**, *6*, 213–255.

(5) (a) Hendrickson, J. B. A General Protocol for Systematic Synthesis Design. *Top. Curr. Chem.* **1976**, *62*, 49–172. (b) Hendrickson, J. B.; Grier, D. L.; Toczko, A. G. A Logic-Based Program for Synthesis Design. *J. Am. Chem. Soc.* **1985**, *107*, 5228–5238. (c) Hendrickson, J. B. Approaching the Logic of Synthesis Design. *Acc. Chem. Res.* **1986**, *19*, 274–281. (d) Hendrickson, J. B.; Huang, P. Multiple Construction in Synthesis Design. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 145–151. (e) Hendrickson, J. B.; Toczko, A. G. SYNGEN Program for Synthesis Design: Basic Computing Techniques. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 137–145.

(6) Ugi, I.; Bauer, J.; Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W.: Dugundji, J. Computer Programs for the Deductive Solution of Chemical Problems on the Basis of a Mathematical Model—A Systematic Bilateral Approach to Reaction Pathways. In *Computational Methods in Chemistry*; Bargon, J., Ed.; Plenum: New York, 1980; pp 275–300.

(7) Wipke, W. T.; Rogers, D. Artificial Intelligence in Organic Synthesis. SST: Starting Material Selection Strategies. An Application of Superstructure Search. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 71–81.

(8) (a) Ugi, I.; Gillespie, P. D. Stoffbilanz-erhaltende Synthesewege und semi-empirische Syntheseplanung mittels elektronischer Datenverarbeitung. *Angew. Chem.* **1971**, *83*, 982–985; *Angew. Chem., Int. Ed. Engl.* **1971**, *10*, 915–918. (b) Ugi, I.; Gillespie, P. D. Beschreibung chemischer Systeme und ihrer Umwandlungen durch BE-Matrizen und ihre Transformationen. *Angew. Chem.* **1971**, *83*, 980–981; *Angew. Chem., Int. Ed. Engl.* **1971**, *10*, 914–915. (c) Dugundji, J.; Ugi, I. An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs. *Top. Curr. Chem.* **1973**, *39*, 19–64.

(9) Bauer, J.; Fontain, E.; Forstmeyer, D.; Ugi, I. Interactive Generation of Organic Reactions by IGOR2 and the PC-assisted Discovery of a New Reaction. *Tetrahedron Comput. Methodol.* **1988**, *1*, 129–132.

(10) Bauer, J. IGOR2: A PC-Program for Generating New Reactions and Molecular Structures. *Tetrahedron Comput. Methodol.* **1990**, in press.

(11) Zefirov, N. S. An Approach to Systemization and Design of Organic Reactions. *Acc. Chem. Res.* **1987**, *20*, 237–243.

(12) (a) Koca, J.; Kratochvil, M.; Kunz, M.; Kvasnicka, V. Mathematical Model of Organic Chemistry. VI. Valence States of Atoms and Their Conversions. *Collect. Czech. Chem. Commun.* **1984**, *49*, 1247–1261. (b) Koca, J.; Kratochvil, M.; Kvasnicka, V.; Matyska, L.; Pospichal, J. Synthon Model of Organic Chemistry and Synthesis Design. *Lect. Notes Chem.* **1989**, *51*, 78–84.

(13) Schubert, W.; Ugi, I. Constitutional Symmetry and Unique Descriptors of Molecules. *J. Am. Chem. Soc.* **1978**, *100*, 37–41.

(14) (a) Augelmann, G.; Fritz, H.; Rihs, G.; Streith, J. An Unexpected Addition Product of Nitrosobenzene with Pyran-2-thione. *J. Chem. Soc., Chem. Commun.* **1982**, 1112–1113. (b) Wochner, M.; Brandt, J.; v. Scholley, A.; Ugi, I. Chemical Similarity, Chemical Distance, and its Exact Determination. *Chimia* **1988**, *42*, 217–225. (c) Defoin, A.; Augelmann, G.; Fritz, H.; Geffroy, G.; Schmidlin, C.; Streith, J. Cycloaddition of 2*H*-Pyran-2-thiones with Nitroso Derivatives. An Unexpected Cycloaddition-Rearrangement Reaction. *Helv. Chim. Acta* **1985**, *68*, 1998–2014. (d) Defoin, A.; Geffroy, G.; Le Nouen, D.; Spileers, D.; Streith, J. Cascade Reactions. A Simple One-Pot Synthesis of the Mitomycin Skeleton. *Helv. Chim. Acta* **1989**, *72*, 1199–1215.

Generation and Enumeration of Carbon Skeletons

JAMES B. HENDRICKSON* and CAMDEN A. PARKS

Department of Chemistry, Brandeis University, Waltham, Massachusetts 02254-9110

Two different programs were written to enumerate all possible carbon skeletons, i.e., saturated hydrocarbons. These enumerations were carried out fully to 11 carbon atoms and for limited numbers of rings to 16 carbon atoms. The results of the two programs were consistent with each other as well as with such previous enumerations as were available. A program to show graphically the stored skeletons was also prepared. The skeletons of commercially available starting materials are a minute subset of the total. Various ways of classifying the data into subsets are illustrated as are superstructure searches from seed skeletons.

INTRODUCTION

Enumeration of structural isomers of organic compounds is a problem which has fascinated chemists and mathematicians for over a century. The first attempts at solutions concentrated on the development of mathematical formulas which would give the numbers of isomers. This line of work began with Cayley[1] in 1874 and was later taken up by Henze and Blair,[2] Pólya,[3] and Read.[4] The development of the computer allowed a different approach, which involved the generation and subsequent counting of graphs representing the isomers. Knop[5] used this method to enumerate the alkanes as well as the substituted alkanes. Masinter's program[6a] generates all compounds for a given empirical formula, and Read[7] developed a program which generated all graphs for up to 10 vertices.

Our interest in the problem arose from the construction of the SYNGEN program for synthesis design.[8] This logic-centered program aims to find all shortest, most efficient synthetic paths to a given target structure. In order to simplify the vast synthesis tree of possibilities, SYNGEN first determines all convergent assemblies for the carbon skeleton of the target

from four starting skeletons, and these must be contained in a catalogue of actual starting materials. Having established a base catalogue of about 6000 commercial compounds, we questioned what proportion of all possible carbon skeletons were actually represented in this catalogue. Since we had developed an efficient canonical numbering system for rapid computer comparison of skeletons,[9] we undertook to create a program to generate all possible carbon skeletons, to act as an "ideal catalogue" of all starting skeletons. When our work was essentially complete, a paper appeared from Kvasnička and Pospichal,[10] who generated all connected graphs of up to eight vertices.

Our interest is to find all carbon skeletons. A carbon skeleton is the connected framework of a compound which has been stripped of all π-bonds and all non-carbon atoms; any chirality which may be present in this framework is ignored. Defined in this manner, skeletons are equivalent to connected topological graphs of degree no greater than four. This will be a subset of the enumerations of Read[7] and Kvasnička and Pospichal,[10] since these include graphs of higher degree. It will also be a subset of the Masinter enumeration,[6a] which