Reaction Planning: Prediction of New Organic Reactions

RAINER HERGES

Institut für Organische Chemie der Universität Erlangen–Nürnberg, Henkestrasse 42, 8520 Erlangen, Germany

A new algorithm for the generation of the complete set of non-isomorphic spanning subgraphs of a given graph or pseudograph is presented. With this procedure unprecedented organic reactions can be predicted. A predicted new carbene reaction, which has been subsequently verified in the laboratory, serves as an example.

INTRODUCTION

Synthesis planning has been developed in the past 30 years from an almost purely inductive procedure to a highly sophisticated science with general problem-solving strategies that can be applied to the analysis of any synthetic problem.[1] Moreover, algorithms have been derived from these principles of synthesis and guiding strategies, which have been programmed on computers.[2] Logic in synthesis[3] culminated in the preparation of highly complex structures such as vitamin B_{12} or erythromycin.[4]

The other important field in synthetic organic chemistry, "reaction planning",[5,6] the search and development of new reactions and methods, is still at a stage where synthesis planning was 30 years ago. Most of the known name reactions were discovered serendipitously or by "chemical intuition or ingenuity". There are frequently occurring problems in synthetic organic chemistry, however, where a deductive, general problem-solving strategy for the "design" of a new reaction would be highly desirable. For example, if a reaction in the late stage of a multistep synthesis fails, a new method has to be developed to overcome the difficulties, otherwise the whole synthetic plan has to be revised and the synthesis has to be started from scratch. The development of a number of new methods was triggered this way. Another problem suitable to be solved by reaction planning is the synthesis of a new class of compounds, which cannot be prepared by using known methods. The preparation of isonitriles with strong electron-withdrawing substituents (Figure 1, R = tosyl, SO_2Cl, and $COCCl_3$) may serve as an example. All attempts in the past, using conventional methods, have failed. Major implications are thermal isomerization to the corresponding nitrile or reaction with nucleophilic starting material or solvent. In order to solve this problem, one has to search for a new reaction which overcomes the known difficulties. Indeed the problem was solved by using our reaction-planning strategies in a deductive and straightforward manner.[6]

In view of the general importance and a lack of deductive principles, we proposed a general problem-solving methodology for the computer-aided discovery and design of new chemical reactions.[5,6] Our approach has proven to be useful in practice: Using the program IGOR[7,8] we did not only predict new chemical reactions but also have been able to verify five of our predictions in the laboratory. Some of these reactions are particularly useful in synthesis, e.g., two cycloadditions for the preparation of 7-[9,10] and 11-membered rings[11] in very high yields, a new retro ene reaction, and a photochemical reaction for the synthesis of strong electrophilic isonitriles.[6]

GENERATION OF BASIC REACTIONS

A crucial point of the reaction prediction procedure is the generation of the complete set of basic reactions[12] to a given reaction matrix. Figure 2 explains the basic principle with a simplified example. Reactions are separated into two graphs. One graph, the "reaction matrix", indicates the changes in bond order during the reaction (dotted lines in Figure 2) and is equivalent to the topology of the transition state in concerted reactions; the other graph, the "σ-frame" (bold lines in Figure 2), indicates the bonds remaining constant during the reaction. Basic reactions are generated by the systematic variation of the σ-frame, without changing the reaction matrix. In Figure 2 the reaction matrix corresponds to the 6-center 6-electron pericyclic reactions, and the "generated" basic reactions are equivalent to the complete set of possible reaction schemes belonging to this reaction category.[13] There are, however, implications since the variation of symmetry equivalent bonds leads to the same basic reaction. As exemplified in Figure 2, the σ-frame of the electrocyclic reaction contains three classes of symmetry equivalent bonds (α, β, and γ). Successive removal of one bond from each class generates the next level of basic reactions. Without considering symmetry, Diels–Alder and 1,5-sigmatropic shift would each be generated twice. In order to avoid redundancy, the bonds of the σ-frame have to be divided into classes of symmetry equivalent bonds that are treated jointly.

At first sight it seems obvious that the bonds are equivalent, if the corresponding centers are equivalent. This is not the case as shown in Figure 3. Prismane has one class of carbon atoms (the ^{13}C shows one signal) but two classes of C–C bonds, a and b. Canonization of bonds can be achieved by converting the molecular graph into the corresponding vertex graph,[14,15] transforming bonds (edges) into centers (vertices) and centers into bonds. The resulting vertex graph now has two classes of symmetry equivalent centers and one type of C–C bonds, and the well-known canonization procedures[16,17] for assigning symmetry equivalence can be applied. Adamantane (Figure 3) is an example of a molecule with two types of centers (tertiary and secondary C atoms), but one type of bond. Conversion to the vertex graph leads to a polycyclic structure[18] with one type of center and two types of bonds.

Redundancy is strongly diminished if, at each level of generation, only one bond of each class of equivalent bonds is removed (or added). Redundancy, however, is not completely avoided since different precursors can lead to the same structure. The 1,5 sigmatropic shift and Cope rearrangement from Figure 1 both can give the retro ene reaction upon re-
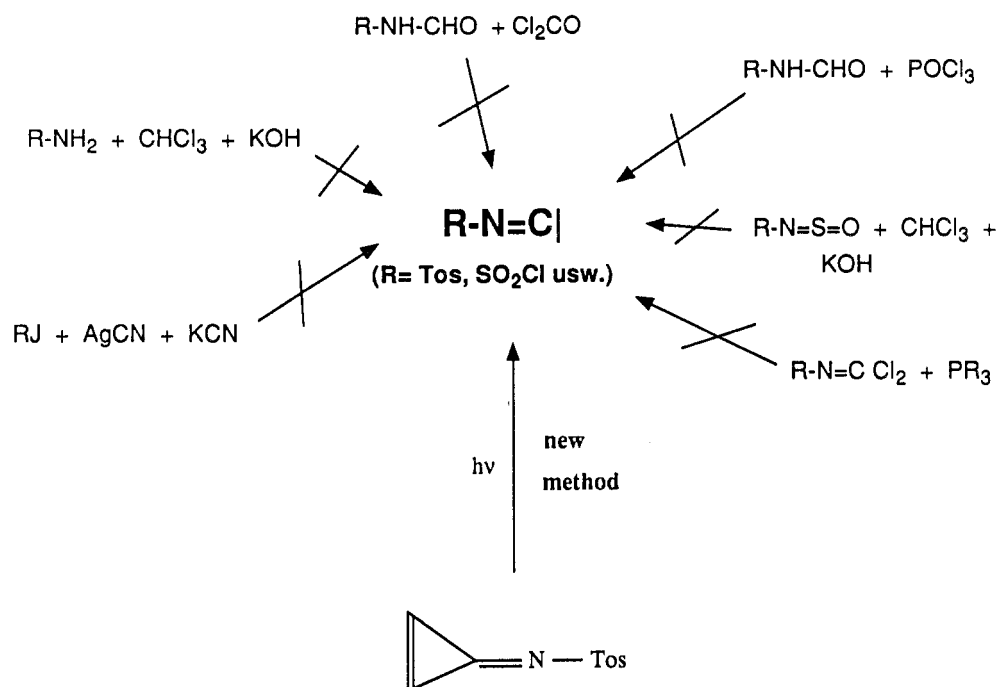
R-NH-CHO + Cl$_2$CO

R-NH-CHO + POCl$_3$

R-NH$_2$ + CHCl$_3$ + KOH

R-N=C|

(R= Tos, SO$_2$Cl usw.)

R-N=S=O + CHCl$_3$ +
KOH

RJ + AgCN + KCN

R-N=C Cl$_2$ + PR$_3$

hv   **new
method**

=N — Tos

Figure 1. Synthesis of strong electrophilic isonitriles.

resonance structures of benzene

"complete graph" G$_c$

β   α
γ
β   α

electrocyclic

1,5 sigmatropic   Cope-rearrangement   Diels Alder

"minimum graph" G$_1$

Figure 2. Generation of basic reactions. The reaction category corresponds to the 6-electron 6-center pericyclic reactions and is indicated with dotted lines. Bonds remaining constant (σ-frame) are boldfaced.

with

$$Z(S_p{}^2) = \frac{1}{p!}\sum_j \frac{p!}{\prod_{k=1}^p j_k!\, k^{jk}} \prod (a_k a_{2k}^{k-1})^{j2k} \prod_{k=0}^{(p-1)/2} a_{2k+1}^{kj_{2k+1}} \prod_{k=1}^{p/2} a_k{}^{kz^{jk}} \times$$

$$\prod_{1\le r<S\le p-1} a_{m(r,s)}^{d(r,s)j_rj_s}$$

was also given by Pólya[19,20] ($g_p = 154$ for a graph with $p = 6$ vertices; 1044 for $p = 7$; and 12 346 for $p = 8$). But to our knowledge no algorithm so far has been given to *generate* all non-isomorphic spanning subgraphs for *any* given graph without relying on *subsequent* removal of redundancy.

The reaction matrix is represented by a pseudograph R with p vertices and q edges. Let G_1 be the minimum spanning subgraph of R containing no constant bonds, with p vertices (Figure 2, bottom) and G_c the complete graph containing the maximum possible number of constant bonds (which is restricted by the maximum valence of the corresponding chemical element) (Figure 2, top). Reactions are generated by subsequently adding one bond to each class of equivalent bonds within G_1, forming the next level of basic reactions, $G_{1.1}-G_{1.i}$. Repeating the procedure with each structure generates a tree of structures $G_{(1...i_n)}$, with n defining the level (number of edges + 1) and i_n the structure i in level n. (See Chart I.)

Each structure $G_{(1...i_n)}$ contains p classes of equivalent bonds, $m_{(1...i_n)}$. In order to avoid multiply generated reactions within one level n, those classes of equivalent bonds are not allowed to be varied that do have bonds in common with a class of equivalent bonds that has been varied in one of the already generated sister structures. These classes are also not allowed to be varied in their direct descendant structures:

If structures $G_{(1...i_n i_{n+1})}$ are generated by a variation of classes of equivalent bonds $m_{(1...i_n)in+1}$ in $G_{(1...i_n)}$ from i_{n+1} to p, the following rules apply for a variation of a class of equivalent bonds $b = m_{(1...i_n i_{n+1})in+2}$ in $G_{(1...i_n i_{n+1})}$ for the generation of $G_{(1...i_n i_{n+1} i_{n+2})}$:

$$b \notin m_{(1...k_n k_{n+1})kn+2} \text{ for } k_{n+1} \ge i_{n+1} \text{ (at the same level)} \quad (1)$$

$$b \notin m_{(1...k_q)} \text{ for } k_q \ge i_q;\ q \in \{1...n\} \text{ (at preceding levels)} \quad (2)$$

This algorithm has been exhaustively tested by comparing the results with those obtained with the IGOR program.[8] In all cases, the results have been identical. Our algorithm seems to be complete and to work without redundancy, even though a mathematical proof cannot be given. Figure 4 shows the generation of the complete set of the 30 spanning subgraphs of a cyclic graph with 8 edges and 8 vertices. Chemically, the list can be interpreted as the set of σ-frames corresponding to the basic reactions proceeding via an 8-center 8-electron pericyclic transition state. Many of these reactions are known from examples in organic chemistry; e.g., three of the four
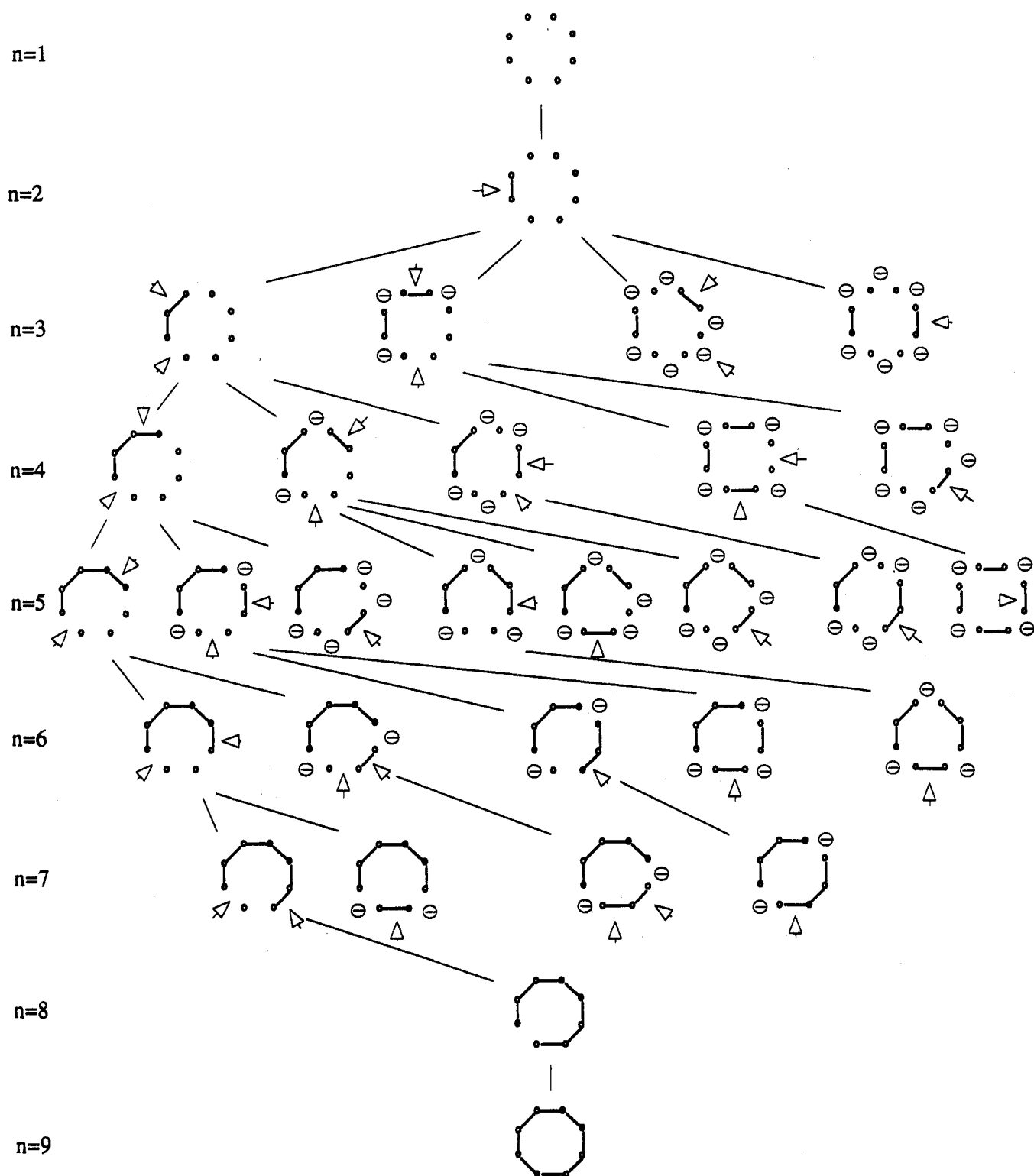

$$\begin{bmatrix} 0\,b\,b\,a\,0\,0 \\ b\,0\,b\,0\,a\,0 \\ a\,0\,0\,0\,b\,0 \\ 0\,a\,0\,a\,0\,b \\ 0\,0\,0\,b\,0\,0 \end{bmatrix} \qquad \begin{bmatrix} 0\,1\,1\,1\,1\,0\,0\,0\,0 \\ 1\,0\,1\,0\,1\,1\,0\,0\,0 \\ 1\,1\,0\,1\,0\,1\,0\,0\,0 \\ 1\,1\,0\,0\,0\,0\,1\,0\,1 \\ 1\,1\,0\,0\,0\,0\,1\,1\,0 \\ 0\,1\,1\,0\,0\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1\,0\,0\,1\,1 \\ 0\,0\,0\,0\,1\,1\,1\,0\,1 \\ 0\,0\,0\,1\,0\,1\,1\,1\,0 \end{bmatrix}$$

Figure 3. Classes of symmetry equivalent bonds and transformation of a molecular graph to the corresponding vertex graph.

moval of a bond from the σ-frame:

1,5 sigmatropic   Cope-rearrangement

retro-ene reaction   retro-ene reaction

In order to avoid redundancy, we propose a new algorithm. The problem of generating the set of basic reactions to a given reaction matrix can be reduced to the graph theoretical problem of finding all non-isomorphic spanning subgraphs (σ-frames) for a given graph. For cyclic graphs, the problem has been solved consistently by Pólya.[19] A polynomial $g_p(x)$ for counting all graphs with a fixed number, p, of vertices

$$g_p(x) = \sum g_{pq} x^q = Z(S_p{}^2, 1 + x)$$

Chart I

Figure 4. Generation of all spanning subgraphs of a cyclic graph with 8 edges and 8 vertices. An arrow denotes the class of bonds, which has been varied to generate the current structure. A circle with a line through the center of it denotes the bonds, which are not allowed to be varied (see eqs 1 and 2).

Reaction Category

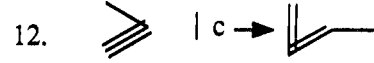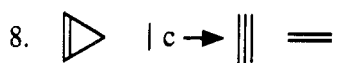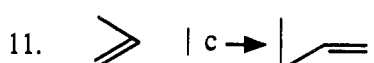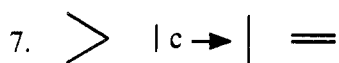### Basic Reactions:

Examples

1.   e.g.

2.   known

3.   e.g.

4.   e.g.

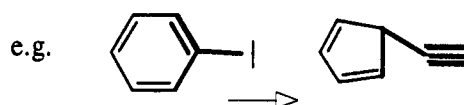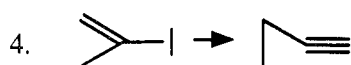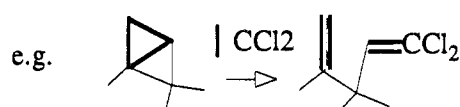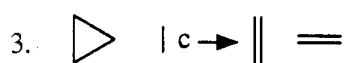5.   9.

6.   10.

7.   11.

8.   12.
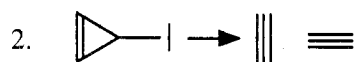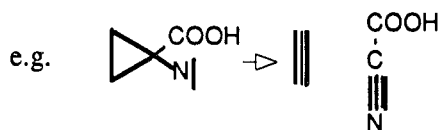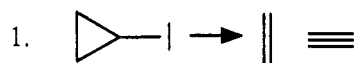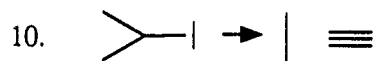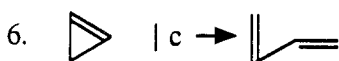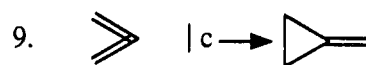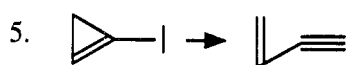
Figure 5. Basic reactions of a complex reaction category.

reactions corresponding to the σ-frames in level $n = 7$ (Figure 4) are known:

1,7 sigmatropic

[2+6] addition

homo-electrocyclic

[4+4] addition

not (yet) known

An example for a homoelectrocyclic reaction is not yet known. (The reaction could also be viewed as a higher homologue of the Cope rearrangement.)

SEARCH FOR A NEW CARBENE REACTION

So far our "designed" reactions have been exclusively pericyclic reactions. Reaction planning, however, is not restricted to any category of reactions. Most of the known name reactions in organic chemistry are covered by two topologies of reaction categories:

linear reactions

   e.g.

cyclic reactions

   e.g.
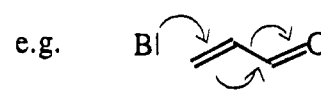
Figure 6. Preparative verification of the predicted carbene reaction.

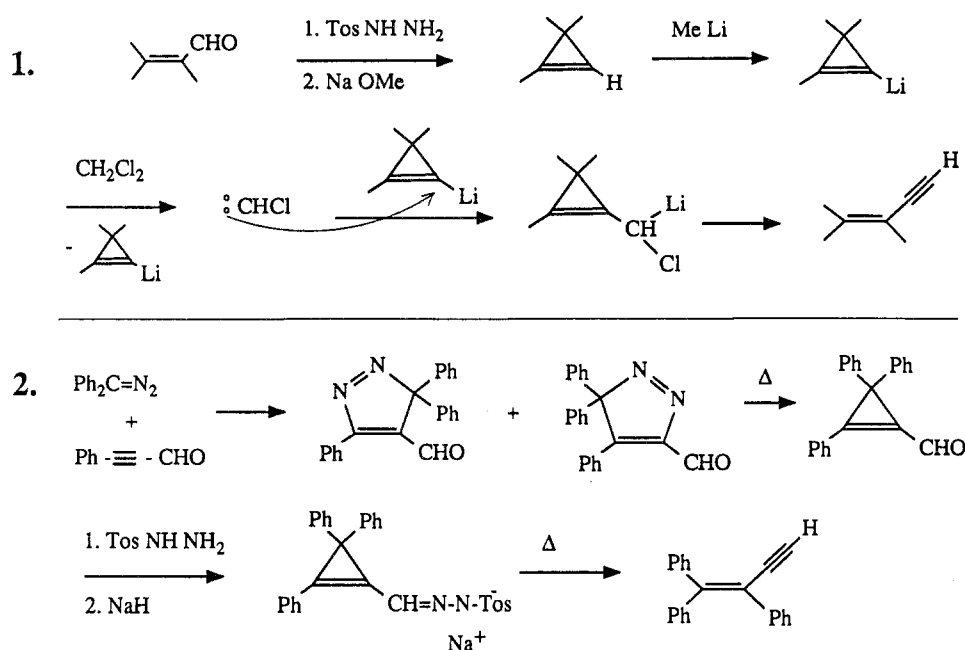The classification of several thousand reactions into reaction categories[21] suggested that there is a third important, but less populated, class of reactions worth investigating. These "complex reactions" are characterized by a set of two bonds between two centers, which are broken simultaneously. Both centers are terminated by a lone pair or a two, four, or higher membered terminator T:

complex reactions
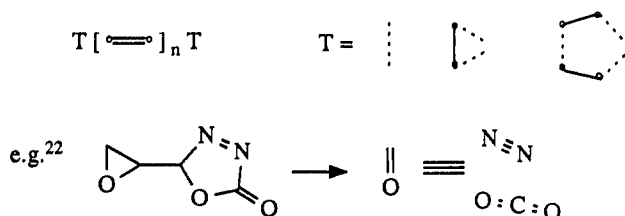
Using our algorithm for the generation of subgraphs, we generated the complete set of conceivable basic reactions for the complex reaction category with a lone pair and a two-center terminator (Figure 5).

Four of the 12 conceivable basic reactions are known. Fragmentation of oxidized cyclopropyl amino acid[23,24] (reaction 1[25]) even is known as a biological process (elimination of ethylene induces the ripening of fruit). Reactions of cyclopropenyl carbene and derivatives[26-30] (reaction 2) have been studied intensively, mainly in the vain hope to generate tetrahedranes. The main products are acetylenes. Dichlorocarbene and methylene add to extremely strained cyclopropanes, inducing the simultaneous cleavage of two cyclopropane bonds[31-35] (reaction 3). Phenylcarbenes and hetero analogues rearrange at very high temperatures to cyclopentadienes[36] (reaction 4). Reactions 5–12 are (so far) unknown. Reactions 6, 8, 9, 11, and 12 are bimolecular, and the addition of carbene to the π-system would probably be a competing or preferred pathway. In reactions 7 and 10, two sp^3 hybridized centers have to react with each other, which is unfavorable on steric grounds. Reaction 5

remains as a reasonable prediction and was chosen for further investigation.

Ab initio calculations at the quite reliable MP4SDTQ/6-31G*//MP2/6-31G* + ZPE level[37] indicate that the reaction proceeds without any barrier of activation. According to our calculation, the 1-cyclopropyl carbene is not a minimum on the energy hypersurface, but rearranges as predicted to vinyl acetylene.

EXPERIMENTAL VERIFICATION

In order to realize our predicted carbene reaction in the laboratory, we generated the 1-cyclopropyl carbene by two different methods. Both methods give the predicted vinyl acetylene in almost quantitative yields (Figure 6).[38] Even though 1-cyclopropyl carbene rearrangements are not useful from a synthetic point of view, the example shows that our reaction-planning algorithms are not restricted to pericyclic reactions, but can be extended to other areas of chemistry.

CONCLUSION

We proposed a new algorithm for the generation of the complete set of non-isomorphic spanning subgraphs of any given graph or pseudograph. The algorithm was used for predicting all basic reactions of a given reaction scheme. A new carbene reaction was discovered this way, which could be subsequently verified in the laboratory. While our "computer invented" reactions[5-7,9-11] so far exclusively were pericyclic reactions, this is the first example from another category of reactions. The most interesting and less exhaustively explored reaction categories in organometallic chemistry will be subject to our future research. Our goal is to develop a general problem-solving methodology for the deductive invention of new chemical reactions. Similar to the development of synthesis planning in the past 30 years, this methodology could not only have impact in synthesis but also in teaching synthetic chemistry.[39]

REFERENCES AND NOTES

(1) Corey, E. J.; Long, A. K.; Rubenstein, S. D. *Science* **1985**, *228*, 408.
(2) For review, see: Winter, H. *Chemische Syntheseplanung*; Springer: Berlin, 1982.

REACTION PLANNING

J. Chem. Inf. Comput. Sci., Vol. 30, No. 4, 1990 **383**

(3) Corey, E. J.; Cheng, X.-M. *The logic of chemical synthesis*; John Wiley & Sons: New York, 1989.

(4) Anand, N.; Bindra, J. S.; Ranganathan, S. *Art in Organic Synthesis*, 2nd ed.; John Wiley & Sons: New York, 1989.

(5) Herges, R. *Chemical Structures*; Warr, W., Ed.; Springer: Berlin, 1988.

(6) Herges, R. *Tetrahedron Comput. Methodol.* **1988**, *1*, 15.

(7) Bauer, J.; Herges, R.; Fontain, E.; Ugi, I. *Chimia* **1985**, *39*, 43.

(8) Bauer, J. *Tetrahedron Comput. Methodol.*, in press.

(9) Herges, R.; Ugi, I. *Angew. Chem.* **1985**, *97*, 596.

(10) Herges, R.; Ugi, I. *Chem. Ber.* **1986**, *119*, 829.

(11) Forstmeyer, D.; Bauer, J.; Fontain, E.; Herges, R.; Hermann, R.; Ugi, I. *Angew. Chem.* **1989**, *100*, 1618.

(12) For a graph theoretical definition of basic reactions see refs 5–7.

(13) A similar approach has already been made: (a) Balaban, A. T. *Rev. Roum. Chim.* **1967**, *12*, 875. See also: (b) Hendrickson, J. B. *Angew. Chem.* **1974**, *86*, 71. (c) Brownscombe, T. F. *Diss. Abstr. Int. B* **1973**, *34* (3), 1035. (d) Arens, J. F. *Recl. Trav. Chim. Pays-Bas* **1979**, *98*, 471; Zefirov, N. S.; Trach, S. S. *Chemica Scripta* **1980**, *15*, 4. (e) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 212.

(14) Withney, H. *Am. J. Math.* **1932**, *33*, 150–168.

(15) Behzad, M. *Proc. Cambridge Philos. Soc.* **1967**, *63*, 679.

(16) Schubert, W.; Ugi, I. *J. Am. Chem. Soc.* **1978**, *100*, 37.

(17) Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107.

(18) Decacyclo[4.3.0.0^{1,3}.0^{2,4}.0^{2,5}.0^{3,8}.0^{4,8}.0^{5,7}.0^{6,9}.0^{7,9}]nonane, numerous attempts not withstanding, could not yet have been synthesized; adamantane converted to antimatter would immediately give access.

(19) Pólya, G. *Acta Math.* **1937**, *68*, 145.

(20) Harary, F. *Trans. Am. Math. Soc.* **1955**, *78*, 445.

(21) Herges, R., unpublished data.

(22) McAlpine, G. A.; Warkentin, J. *Can. J. Chem.* **1978**, *56*, 308.

(23) Galluci, R. R.; Jones, M., Jr. *J. Am. Chem. Soc.* **1976**, *98*, 7704.

(24) For review, see: *Nachr. Chem. Techn. Lab.* **1984**, *32*, 697.

(25) Friedman, L.; Shechter, H. *J. Am. Chem. Soc.* **1960**, *82*, 1002; further examples are cited in: Shelvin, P. B.; McKee, M. L. *J. Am. Chem. Soc.* **1989**, *111*, 519.

(26) Closs, G. L.; Harrison, A. M. *J. Org. Chem.* **1972**, *37*, 1051.

(27) Masamune, S.; Kato, M. *J. Am. Chem. Soc.* **1965**, *87*, 4190.

(28) White, E. H.; Winter, R. E. K.; Graeve, R.; Zirngibl, U.; Friend, E. W.; Maskill, H.; Mende, U.; Kreiling, G.; Reisenauer, H. P.; Maier, G. *Chem. Ber.* **1981**, *114*, 3906.

(29) Rodewald, L. B.; Lee, H.-K. *J. Am. Chem. Soc.* **1973**, *95*, 623, 3038.

(30) Shelvin, P. B.; Wolf, A. P. *J. Am. Chem. Soc.* **1978**, *56*, 308.

(31) Doering, W. v. E.; Coburn, J. F., Jr. *Tetrahedron Lett.* **1965**, 991.

(32) Wiberg, K. B.; Lampman, G. M.; Ciula, R. P.; Cannor, D. S.; Schertler, P.; Lavanish, J. *Tetrahedron* **1965**, *21*, 2749.

(33) Applequist, J. E.; Wheeler, J. W. *Tetrahedron Lett.* **1977**, 3411.

(34) Frey, H. M.; Jackson, G. E.; Smith, R. A.; Walsh, R. *J. Chem. Soc., Faraday Trans I* **1975**, *71*, 1991.

(35) Jackson, J. E.; Mock, G. B.; Tetlef, M. L.; Zheng, G.-X.; Jones, M., Jr. *Tetrahedron* **1985**, *41*, 1453.

(36) For review, see: Wentrup, C.; Harder, R. *Top. Curr. Chem.* **1976**, *62*, 173.

(37) Herges, R., unpublished data.

(38) Experimental details will be published in *Tetrahedron Lett.*

(39) Barton, D. H. R. *Aldrichimica Acta* **1990**, *23*, 3.