Planning of Synthetic Pathways on the Basis of Synthesis Strategies

JAKOB HERMANN WINTER

HOECHST AG, D-6230 Frankfurt am Main 80, West Germany

Received November 5, 1984

Methods for computer-assisted synthesis planning usually proceed retrosynthetically step by step from a target molecule to determine the reaction steps required in a synthetic pathway. In a previous article it was shown how synthetic pathways can be classified by the total generalization of the starting products, intermediates, and end products involved. This procedure reveals strategies of skeleton-transforming reaction sequences. These can in turn be used to plan new synthetic pathways. The concept described permits planning in the direction of synthesis, retrosynthetic planning, and planning that proceeds from intermediate products. Furthermore, the synthesis pathway to be planned can exhibit any amount of branching in accordance with the respective strategies used or proceed via intermediate products that are more complicated than the end product itself.

INTRODUCTION

General interest in the systematic planning of syntheses in organic chemistry has existed at least since the appearance in the 1960s and 1970s of the first publications of Corey, Wipke, Bersohn, Gelernter, Ugi, etc.[1] on computer-assisted synthesis planning. Principles were adopted and planning strategies were considered from existing theories of decision making and problem solving. Corey coined the term synthon for the substructure that must be introduced into a target molecule. Frequently, however, the term has been used to designate a particularly favorable starting material. The planning methods actually used are based essentially on the stepwise determination and evaluation of possible precursor molecules or of possible reaction products of substances for which one wants to find a use. However, some programs also consider long-range strategies whenever the target molecule has been found to contain structural elements that can be produced via certain reaction sequences that start from suitable types of starting materials.

A characteristic feature of most methods is that they apply known reactions to new synthesis problems in an analogous fashion.[2] Restrictions on the possible combinations of sequential synthesis steps, which could otherwise become boundless, can be imposed through the type of evaluation employed and by a comparison of possible precursor molecules with commercially available starting materials. An important criterion is that the synthesis of complex substances from less complex substances should consist of a steady progression from the less complex starting material to ever more complex intermediates, a principle which, however, is hard to define, and one which also places a limit on the planning possibilities.

Certainly the application of analogies with known and successful reactions is a very good method and one that already occupies a dominant position in chemistry. However, it is not necessary to restrict the use of analogy to individual reaction steps, and in fact, conventional planning often applies analogy to successful long-range synthetic pathways. Central to this approach is not the application of analogy to the sequence of individual reaction steps but to the overall strategy of the synthesis pathway. Consequently, what is needed are methods that enable synthesis strategies to be easily recognized and represented, especially when they also permit the large number of possible synthesis strategies to be systematized. Such methods will be of particularly great value if at the same time they can form the basis for the machine planning of new synthesis pathways.

RECOGNITION OF SYNTHESIS STRATEGIES

In an earlier article[3] a method was presented for recognizing and representing synthesis strategies on the basis of useful generalizations of known synthetic pathways. In this method the reaction conditions are not taken into account, and the auxiliary substances are not mentioned. Only the starting materials, intermediates, and end products are described, but in a fully generalized topological system of representation derived from the conventional structural formula. The basic idea is as follows (see Figure 1).

It is useful to distinguish between skeleton-transforming reactions on the one hand and the transformation of functional groups on the other—the latter as groups of heteroatoms with each group on one carbon atom, as long as they are not aromatically bonded, or as carbon–hydrogen groups in a multiple carbon–carbon bond, which is not aromatic, and in the limiting case as a single carbon atom in a multiple carbon–carbon bond. The total generalization of a structural formula can be achieved by individually symbolizing only skeleton-forming atoms and disregarding the distinctions among the functional groups. Further, only σ-bonds and aromatic bonds are indicated. At the first level of generalization the only distinction made among skeleton-forming atoms is that the carbon atoms in functional groups are given a symbol different from that for the other skeleton-forming carbon atoms. At the second level of generalization all skeleton-forming carbon atoms are given the same symbol. Finally, at the third level of generalization there is only one symbol for all skeleton-forming atoms, including the heteroatoms. In this way the formulas of many intermediates coalesce into one formula. Synthesis pathways become very comprehensible, which makes the strategies of skeleton transformation clear. Nevertheless, all intermediate products are retained and are storable and retrievable by means of appropriate topological processes.

For simplicity's sake topological, totally generalized structural formulas, as shown in Figure 1 in comparison with the usual specific structural formula, will hereafter be referred to as "generon(s)". This designation is to be applied to all total generalizations regardless of the level of generalization. Figure 2 reproduces a sequence of generons, which thus represents a generalization of the corresponding synthesis pathways and their synthesis strategies.[5] Consequently, this sequence is the strategic pattern of the corresponding synthetic pathway—referred to below as "pathway pattern", which has an inestimable number of possible specific realizations. The pathway pattern indicates nothing at all about the type and number of transformations of functional groups involved. It is very easy to work out pathway patterns with pencil and paper, but

0095-2338/85/1625-0389$01.50/0 © 1985 American Chemical Society

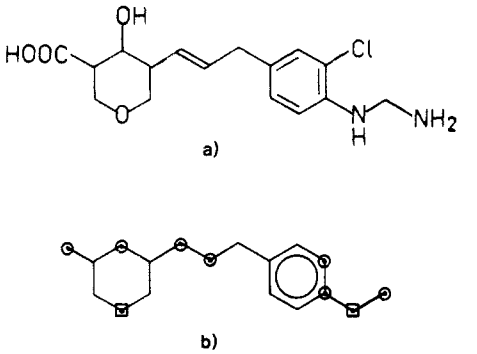

Figure 1. Structural formula in normal specific description (a) and in a first-level total generalization description (b) as "generon". (Additional rules and explanations, see reference 3.) (·) (Or as angle formed by two bond lines) C saturated or aromatic, otherwise bound to C or H; (⊙) functional groups with central C atom and also C joined to another C by a double or triple bond and maybe in addition to C or H; (□) skeleton-forming heteroatom and may also be linked with nonskeleton-forming heteroatom.

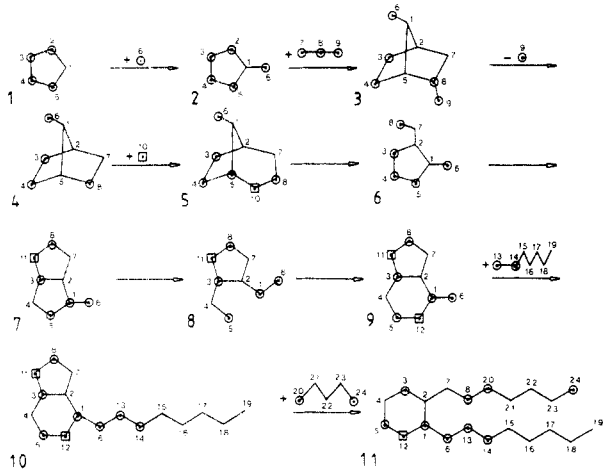

Figure 2. Strategy pattern of a synthesis as a sequence of generons. (The numbers of the skeleton atoms are identification numbers and therefore are the same for the same atoms.)

it is also possible to program a machine to perform this task once the specific structural formulas are stored in the machine in an unambiguous way.

PATHWAY PATTERNS FOR PLANNING SOLUTIONS TO SYNTHESIS PROBLEMS

As soon as a pathway pattern is available, a practicable synthetic pathway can be derived from it through specification of the generons, i.e., transformation of the generons to specific structural formulas that can be derived from one another via appropriate specific reactions. Of course, each one of these reactions can be in turn a multistep process. It is obvious that through the first specification of a generon by atoms and groups of atoms the other generons of the pathway pattern used are also largely determined because they must derive from each other. One can first decide on a starting material or on an intermediate or, what will be the most frequent case, on a target molecule to be synthetized. In other cases it may be necessary to specify several substances of a pathway pattern simultaneously, as for instance when it has been decided that a starting material must be used for the manufacture of a certain target molecule. The only question remaining is then which pathway pattern is most suitable for the purpose.

In order to clarify the concept further, let us assume that the target molecule A (Figure 3) is to be produced. A retrieval procedure has found that the pathway pattern of Figure 2 can be recommended because molecule A conforms to generon 6.

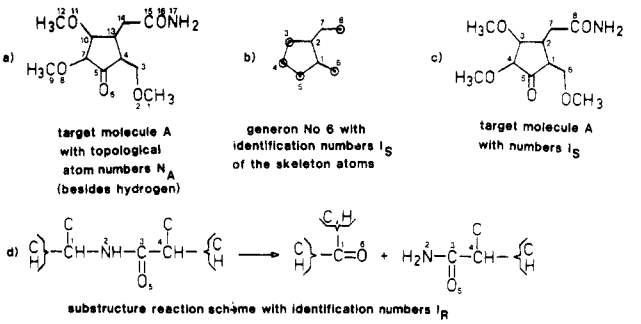

Figure 3. (a) Target molecule, (*b*) generon of the target molecule, (*c*) target molecule with the identification numbers of the generon, and (d) suitable reaction scheme.

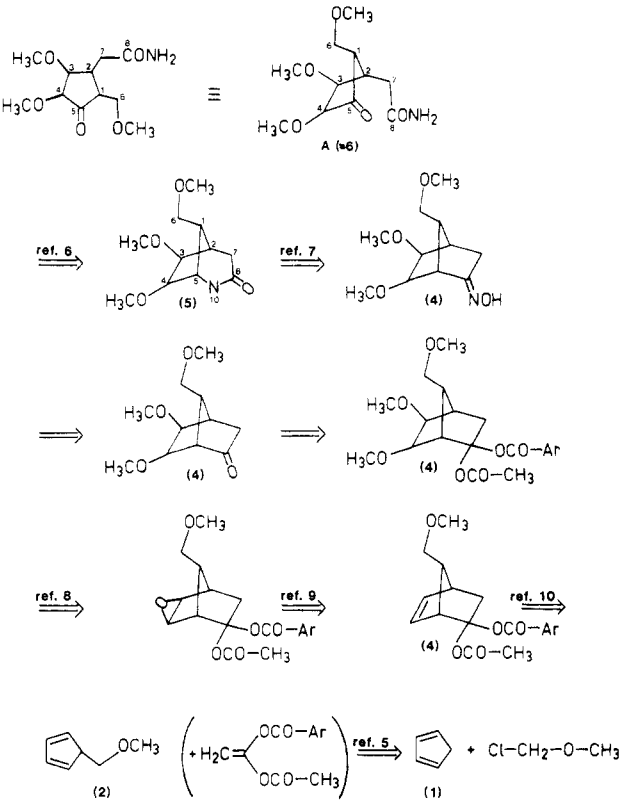

Figure 4. Synthesis pathway for target molecule A planned in retrosynthetic directions by using the way pattern of Figure 2. (Arrows are against synthesis direction. Numbers in parentheses are the numbers of the generons in the pattern of Figure 2 suitable for these special molecules.)

"To conform to" means that in A a number of skeleton-forming atoms 1–8 exist that satisfy the conditions of generon 6, for example, the fact that a ring carbon atom is linked to two other ring carbon atoms and with a nonring carbon atom that is a functional group and further is linked with a hydrogen atom. This can only be the atom with the identification number $I_S = 1$ in generon 6. In Figure 3 that atom in target molecule A has at first its topological atom number. By comparison operations, the identification numbers of generon 6 can be transferred to target molecule A. Thus, all skeleton atoms of the pathway pattern have been determined with respect to the synthesis of A, to the extent that identification numbers are provided that fit the molecule. Thereupon, the reaction steps stored in a corresponding, essential file of substructural schemata (see Figure 3d) can be examined for conformity with a possible precursor molecule and at the same time with its product, thus in this case with generon 5 as the precursor molecule and target molecule A as the product (last step of the pathway in synthesis direction, Figure 4). The substructures of a reaction scheme (like Figure 3d) must agree

specifically with the target molecule and specifically with generon 5 to the extent that the skeleton atoms have meanwhile been specified, as described. As Figure 4 shows, this is possible with the reaction of Figure 3d according to reference 6. (References are only given as far as the reaction steps are not trivial.) The unspecified parts of the precursor generon are then specified for the reacting structural parts by means of the starting material substructure of the reaction scheme and for the nonreacting parts by way of the target molecule through transfer of the atoms and atom groups identified there.

Of course, a number of evaluative measures must be applied that confirm or deny the feasibility of the reaction. A refinement of the method consists in omitting generons from the pathway pattern sequence. Furthermore, there will be cases in which additional intermediate steps are required between two generons. In such a case the substructure of the precursor molecule in the reaction scheme is not transferred to the next generon in the retrosynthetic direction, but with the help of this substructure from the previously specified molecule, a specified precursor molecule is created (see Figure 4, epoxide precursor), and this is used for further reaction planning back to the next following generon (provided that no further intermediate steps are involved).

Figure 4 shows a specific synthesis pathway such as can be obtained from the pathway pattern of Figure 2 and a lot of reaction schemata like Figure 3d in the manner described. As just mentioned, the process of epoxidation is a reaction step that has been inserted into the generon sequence. Pure transformations of functional groups, which do not change the diagram of a generon, use the same generon to plan several reaction steps, as illustrated by the conversion of an acetal diester to a ketone and of the latter to an oxime, always by means of fitting reaction schemata.

CONCLUSIONS

Implementation of the concept described requires that adequate pathway patterns are available and that a reaction library exists for specifying the pathway patterns under consideration. The pathway patterns themselves can be derived from known synthesis pathways, or they can be developed systematically. When these conditions have been satisfied, the planning of synthesis pathways of any form whatever is possible, which eliminates a bottleneck that hampers other procedures.

ACKNOWLEDGMENT

I express my sincere thanks to Dr. Robert Fugmann, HOECHST AG, and to Dr. Douglas Maass, Beilstein-Institut, for their interest in this paper and for their suggestions.

REFERENCES AND NOTES

(1) Corey, E. J.; Wipke, W. T. "Computer-Assisted Design of Complex Organic Syntheses". *Science (Washington, D.C.)* **1969**, *166*, 178–192. Bersohn, M. "Automatic Problem Solving Applied to Synthetic Chemistry". *Bull. Chem. Soc. Jpn.* **1972**, *45*, 1897–1903. Gelernter, H.; Shridharan, N. S.; Hart, A. J.; Yen, S.-C.; Fowler, F. W.; Shue, H.-J. "The Discovery of Organic Synthetic Routes by Computer". *Top. Curr. Chem.* **1973**, *41*, 113–150. Dugundji, J.; Ugi, I. "An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs". *Top. Curr. Chem.* **1973**, *39*, 19–64. Salatin, T. D.; Jorgensen, W. L. "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 1. Overview". *J. Org. Chem.* **1980**, *45*, 2043–2051.

(2) See Winter, J. H. "Chemische Syntheseplanung in Forschung und Industrie" ("Chemical Synthesis Planning in Research and Industry"); Springer-Verlag: Berlin, 1982.

(3) Winter, J. H. "Classification of Synthetic Pathways in Organic Chemistry". *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 263–265.

(4) For the complete list of rules used, see Figure 1 of reference 3.

(5) Based on a synthesis pathway in Nelson, N. A.; Jackson, R. W. "Total Synthesis of Thromboxane B_2". *Tetrahedron Lett.* **1976**, 3275–3278. See also, *Synform* **1983**, *1*, 194–199.

(6) See Lohmar, R.; Steglich, W. "α-Amino Acids as Nucleophilic Acyl Equivalents: Synthesis of Ketones and Aldehydes by Means of Oxazolin-5-ones". *Angew. Chem., Int. Ed. Engl.* **1978**, *17*, 450–451.

(7) Approximately in accordance with Sakai, I.; Kawabe, N.; Ohno, M. "Reactions of 1-Halo-1-nitroso- and 1-Halo-1-nitrocycloalkanes with Triphenylphosphine. A New Synthesis of Lactam". *Bull. Chem. Soc. Jpn.* **1979**, *52*, 3381–3383.

(8) Possibly via two steps, first ring opening and simple etherification in accordance with Posner, G. H.; Rogers, D. Z. "Organic Reactions at Alumina Surfaces. Mild and Selective Opening of Epoxides by Alcohols, Thiols, Amines, and Acetic Acid". *J. Am. Chem. Soc.* **1977**, *99*, 8208–8214 and then further etherification.

(9) Confer epoxidation approximately in accordance with Arias, L. A.; Adkins, St.; Nagel, Ch. J.; Bach, R. D. "Epoxidation of Alkenes with Trichloracetonitrile/Hydrogen Peroxide in a Neutral Biphasic Solvent System". *J. Org. Chem.* **1983**, *48*, 888–890.

(10) See Tamariz, J.; Vogel, P. "New Dienophiles: 1-Acetylvinyl Arenecarboxylates. Reactivity toward Cyclopentadiene and Exocyclic Dienes". *Helv. Chim. Acta* **1981**, *64*, 188–197.