Synthesis Simulation by Synthon Substitution

A. WEISE

Logic Design Team, Computer-Systeme GmbH, Chiemgaustrasse 38, 8000 München 90, FRG

A computer program which can recognize the transforms in reaction equations has been developed. The program identifies the transforms, supplies corrected stoichiometry, and uses the transform data in synthesis prediction.

INTRODUCTION

There are in principal three ways to computer-assisted synthesis design in organic chemistry:

1. standard expert system technology
   Rules about synthetic transformations are derived from the literature by experts and transferred to the expert system.[1-4]
2. programming a formalism derived from organic mechanism theory
   Naturally such systems are best suited for the simulation of mechanistic reaction steps.[5-9]
3. automatic analysis and evaluation of a reaction data base

The third approach requires a module for the recognition of transforms in reaction equations.[10-12] A third version of such a module has been recently completed. This program is also capable of treating nonbalanced reaction equations and doing stoichiometric corrections.

The *X*-expanded substructures in Figure 1 represent supplements derived by the program SYNTHON from the information inherent on the other side of the reaction equations.

The automated transform recognition delivers very valuable information. Simply by counting the number of atoms, the transform pool may be divided into simple transforms, which are predominantly functional group transformations, and into transforms with higher synthetic potential. The latter are less likely to be memorized and hence less likely to be put forward as suggestions by chemists.

It is possible to concatenate single transforms to strategic transforms. Such strategic transforms as well as the single transforms form a good basis for a reaction-type-oriented retrieval system. But somewhat more interesting is the use of the automatically generated transforms for synthesis design.

THE PRINCIPLE OF SYNTHON SUBSTITUTION

The transforms have to be organized in a library, called the transform library (TLIB), which allows the direct access to the single synthons from both sides of the equations. Then we are able to simulate syntheses in both directions, in the forward as well as in the backward direction.

The operation which has to take place is **synthon substitution**. The principle is illustrated in Figure 2. In the structure a fragment designated as "target" is marked by a dotted line. This fragment matches with an end synthon in the transform library. Replacement by the start synthon of this transform simulates a synthesis step in the retrosynthetic direction, and vice versa, start synthons deliver successors by synthon substitution if they can be matched with fragments in start structures. In this way, a simulation system incorporating the principle of synthon substitution can make synthesis predictions if the targets are the start synthons of the TLIB. At the same time it can make synthesis planning if the targets are the end synthons of the TLIB.
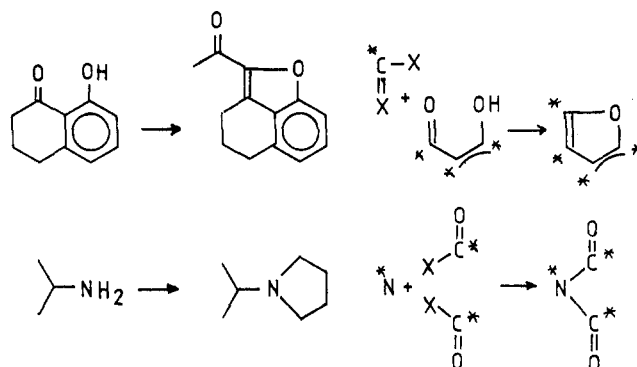

Figure 1. Stoichiometry-corrected transforms generated from non-balanced reaction equations by the program SYNTHON. The asterisks mark the earlier connections to the context substructures.
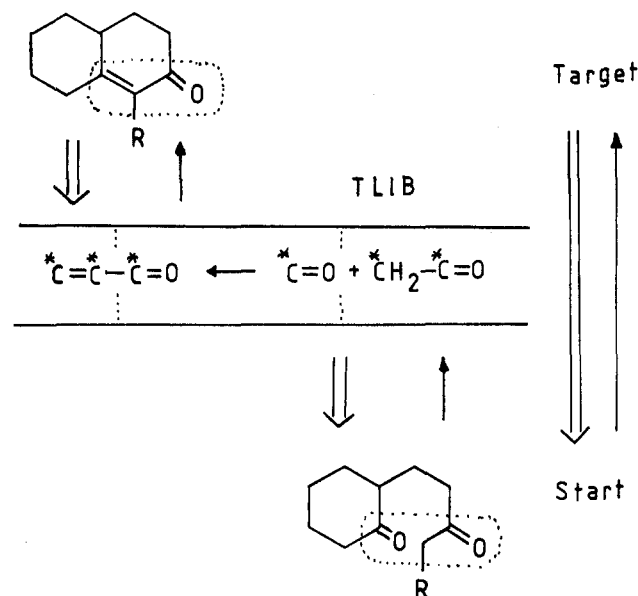

Figure 2. Synthon substitution scheme.

In Figure 3 a synopsis is given of the simulation system based on automatic transform recognition and synthon substitution.

SYNTHESIS STEP EVALUATION

Evaluation of the simulated synthesis steps is the crucial component of every synthesis simulation program.

In our case transform recognition is also the source of the data for evaluation. These data are context descriptors which are associated with each transform. After a lot of experiments with different context group codes, the principle of "full context registration" has now been applied. When the program SYNTHON separates the transform substructures from the reaction
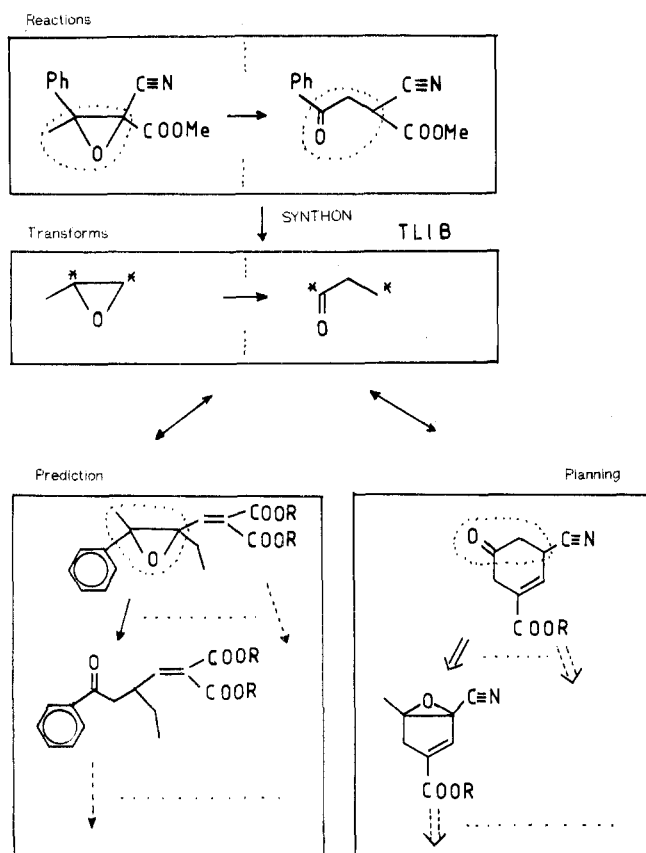
Reactions

Transforms

SYNTHON

TLIB

Prediction

Planning

Figure 3. Synthesis prediction and planning with automatically derived transforms.

Figure 4. Context groups of a special reaction, marked by dotted lines.

equations, it also generates the context substructures in the form of canonized and condensed connection tables.

A context group is limited by the dissected bond to the synthon on the one side, and on the other side, the group will be cut after the first saturated carbon–carbon bond. A context group may contain 12 atoms at maximum. Figure 4 shows the context groups of a single reaction belonging to the enamine–chloral trimerization transform.

In the simulation part of the system, the context groups associated with each transform are exploited for evaluation by a mapping procedure. The context groups of the fragments,
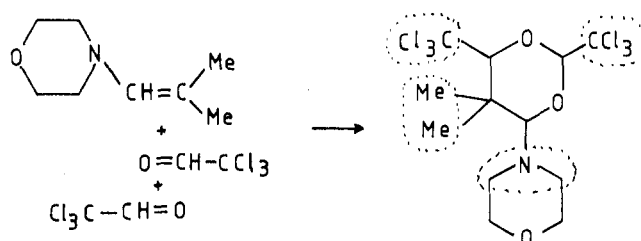
generated from the target or start structures, are compared with the context groups of the matching transforms. The outcome of this comparison is a set of similarity values. The simulations can be very effectively controlled by setting variable threshold values.

At high threshold values the simulation runs like a reaction retrieval system. By lowering the context evaluation threshold, we come more and more into the sphere of analogous reasoning or programmed presumption. This flexibility demonstrates one of the advantages of "full context registration". Another advantage is that the system remains open to the application of any other evaluation method.

In summary, the model of synthesis simulation by synthon substitution in connection with automated transform recognition has a role to play in the future development of computer chemistry.

REFERENCES AND NOTES

(1) Corey, E.; Wipke, W. T. *Science* **1969**, *166*, 178.
(2) Wipke, W. T.; Ouchi, G. I.; Krishnan, S. *Artif. Intell.* **1978**, *9*, 173–193.
(3) Gelernter, H. L.; Sridharan, N. S.; Hart, H. J.; Yen, S. C.; Fowler, F. W.; Shue, H. J. *Top. Curr. Chem.* **1973**, *41*, 113.
(4) Bersohn, M.; Esack, A. *Chem. Rev.* **1976**, *76*, 269.
(5) Blair, J.; Gasteiger, J.; Gillespie, C.; Gillespie, P. D.; Ugi, I. In *Computer Representation and Manipulation of Chemical Information*; Wipke, W. T., Ed.; John Wiley: New York, 1974; p 129.
(6) Ugi, I.; Gillespie, P. *Angew. Chem.* **1971**, *83*, 982.
(7) Weise, A. *Z. Chem.* **1973**, *13*, 155. Weise, A. *Z. Chem.* **1975**, *15*, 333.
(8) Herges, R. *Tetrahedron Comput. Methodol.* **1988**, *1* (1), 15.
(9) Barone, R.; Chanon, M. In *Computer Aids to Chemistry*; Vernin, G., Chanon, M., Eds.; John Wiley: Chichester, 1986; p 19.
(10) Weise, A. *J. Prakt. Chem.* **1980**, *322*, 761–768.
(11) Fanatsu, K.; Del Carpio, C. A.; Sasaki, S.-I. *Tetrahedron Comput. Methodol.* **1988**, *1* (1), 39.
(12) Wipke, W. T.; Rogers, D. *Tetrahedron Comput. Methodol.* **1989**, *2* (3), 177.