117

Context Description in Synthesis Planning[†]

Rainer Moll

CASAF Gesellschaft für Computerchemie mbH, Beuchaer Strasse 3, D-04318 Leipzig, Germany

One of the main problems of computer-assisted synthesis planning is in which manner chemical knowledge can be used. The new program TRESOR is based on the synthon approach. For evaluation of synthesis steps a system of "context descriptions" in the form of "atom keys" is presented. They describe the influences of bonds and neighbor atoms to reaction centers and comprise information in three levels about hybridization, partial charges, free electron pairs, empty orbitals, etc.

INTRODUCTION

At the beginning of the seventies, the application of computer techniques in pesticide research was started at CHEMIE AG Bitterfeld/Wolfen (former VEB Chemiekombinat Bitterfeld). In a rather close relation to the chemistry information system SPRESI, the database WIFODATA was created and QSAR methods were introduced. In 1985, the first version of the synthesis planning program RDSS (reaction design by synthon substitution) was installed and all parts were combined in the project CARSA.[1] Since November 1992 this project is continued completely by "CASAF Gesellschaft für Computerchemie mbH", Bitterfeld/Leipzig.

At present CARSA consists of succeeding parts:

WIFODATA	research database; substance and finding documentation
QSAR	methods and programs for quantitative structure–activity relationships
RDSS/TRESOR	program packages for synthesis planning
MOLMOD	PC-program for molecular modeling

BACKGROUND

At present synthesis planning is the main point within CARSA, and some remarks about the history of our systems are necessary for better understanding of the following.

The origin of our systems is AHMOS (automatic and heuristic modeling of organic syntheses), developed by A. Weise in the seventies.[2] The basic pilosophy for this and all other programs is the so-called "synthon approach", derived from the definition of Corey, but in a more formal sense.[3] From AHMOS in the next step RDS (reaction design software) was developed. We came into touch with it in the middle of the eighties, and this had been the starting point for our own development of first RDSS (reaction design by synthon substitution) and now TRESOR.[4]

The most important problems in all programs for synthesis planning are how can I achieve reasonable suggestions for the synthesis of a given molecule or the synthetic possibilities for a special starting material, and how can I reduce the irrelevant material, without suppressing or missing interesting solutions? The crucial point is the using and application of "chemical intelligence" in a suitable form. At present there are some different systems for solving this problems. Following

Ugi,[5] synthesis planning programs can be divided into empirical, semiformal and formal (or logic) oriented programs. All systems have their "pros and cons" and their special philosophies. They work in connection with or without reaction databases and/or with or without special evaluations, rules, assumptions, etc.: but it cannot be discussed here in detail (cf. refs 5–7).

TRESOR

The principle of TRESOR (tracing and evaluation of synthons in organic reactions), the latest issue of our programs, is drawn schematically in Figure 1. To our opinion TRESOR can be classed with the semiformal programs, because on the one side we use a "synthon library" as knowledge base, which is generated from a reaction database. On the other side we have a special rule system in the form of—like we say—"context descriptions" for evaluation.

In the first step from the SPRESI reaction database reaction equations are extracted (only with a minimum of bibliographical and factual information) and from them, in the second, more abstracting step, synthon equations are derived by means of a special synthonization program. Besides the mentioned large database, another source for synthons is a fundamental reaction database created by our group in Bitterfeld, which contains "textbook chemistry".

At this point it has to be explained more in detail, what we mean, if we speak about synthons (Figure 2). They are defined as those parts of molecules or educts and products, which are involved directly in the reaction and not identical on both sides of reaction equations. To each reaction equation, a synthon equation can be formulated, but it has to be mentioned, that not each reaction equation gives raise to an individual synthon equation. In order to clarify that, one may only exchange the cyano groups in Figure 2 to other electron withdrawing groups (COOR, COR, etc.); also in these cases ketene-*S,N*-acetals are the reaction products, but the synthon equation is always the same.

According to our understanding the synthons represent the reaction centers, that means, the atoms or atom groups, on which bonds are broken or attached, additionly the α-sphere around them. This is an important point for the influence of other parts of a molecule to the reactivity. As it was mentioned above, synthons form a "synthon library" as knowledge base for the program. In some respect this library can be regarded as a "reaction equivalent database" or "database for synthesis building blocks".

The synthonization is—in contrast to the formulation of transforms, etc., in LHASA or other programs—an automatic

---

TRESOR

reaction database

↓

reaction equation database
fundamental reaction database

↓

SYNTHON LIBRARY

↓

context description

↓

synthesis simulation

reaction prediction   synthesis planning
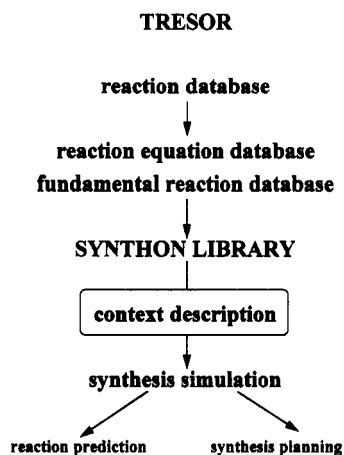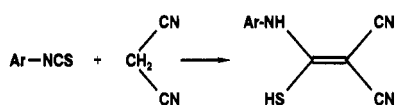
Figure 1. Structural scheme of TRESOR.

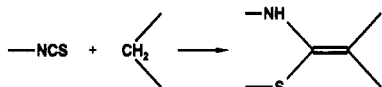- **Reaction Equation:**

- **Synthon Equation:**

Figure 2. Synthon principle.

process and the different steps are the following: calculation of charge distribution of educts and products; comparison with standard electronegativity values for estimating of inductive and other electronic effects; partial charge calculation by means of EPEE (empiric partial electronegativity equalization) method;[8] on basis of the results searching for "breaking (disconnective) points" in the molecules; generation of synthon connection tables. TRESOR is designed, like its predecessors, for synthesis planning and reaction prediction. It operates in such a manner that in the first step an inputted molecule is fragmented. This fragmentation is a formal process, running automatically by defaults, but the user has the possibility, to mark the starting atom in the molecule and to limit the size of fragments by giving the largest number of atoms. In the next step the system compares the fragments in a mapping process with the synthons in the synthon library, in order to find potential son structures and so on. This is not a "rigid" but more a "fuzzy" mapping, made by a special similarity analysis. One of the most serious problems is, how can I make these processes rational in chemical sense? It has already been mentioned that this is a general problem for all synthesis planning systems. The reason for this is that there is no general valid and consistent quantitative theory of chemical reactivity. There are rather numerous concepts and ideas, some very sophisticated, to master the problems.

In general, the knowledge based programs use the chemical experience by utilizing (large) reaction databases (e.g. in form of transform libraries), and the reaction proposals are evaluated by ratings,[9] by mechanistic schemes, and/or heuristic and selection rules.[10] Another possibility is the application of empirical data from thermochemistry,[6] the evaluation by estimating electronic effects, and so on.

Our approach is the conception of "context description".[11] We define as context—in an admittedly very generalizing
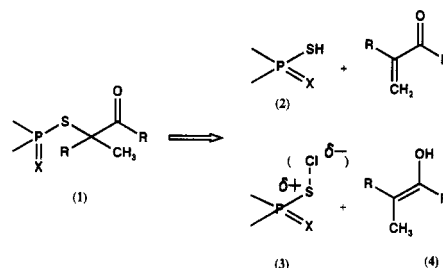

Figure 3. Retrosynthetic analysis of formation of thiolophosphorus esters.

manner—all intramolecular factors, which influence directly the reaction centers. It is clear, that also intermolecular factors and the reaction conditions have a great influence, but to formulate it in a quantitative manner and as a context is a task for the future.

We describe the context as so-called "atom keys". They consist of three levels:

(a) The basic level comprises the atomic number and the hybridization; the latter is used instead of connectivity, because it is more apparent.

(b) The second level consists of descriptors for partial charge, conjugation sharing, free electron pairs, empty orbitals, and radical character.

(c) The third level includes information like chirality marking, equivalence criteria, and ring sharing. The latter is characterized by ring size and stability.

At present the development and testing of context description is still in a state of flux and there are some intricate problems. Examples are the exact representation of donor–acceptor properties and of the aromaticity.

But in spite of all problems still existing we believe, that the idea of context description is promising and that it is a contribution to reactivity evaluation within synthesis planning.

Two examples of investigating synthetic problems with TRESOR shall be given. The first is the examination of methods for synthesis of thiolophosphorus esters (cf. **1** in Figure 3). They can be prepared by standard methods, like the addition of thiophosphorus esters to unsaturated compounds or their reaction with α-halogen–carbonyl derivatives and so on. We found a new and unusual way by reaction of chlorosulfane–phosphorus esters (**3**) with CH–acidic (enolizable) components.[12] In a retrosynthetic analysis some proposals were made by the program, e.g., the addition of **2** to vinylketones. In this example the S-atom acts as nucleophile. But the reaction of the other type of phosphorus compound **3** to ketones (formulated as enol (**4**))—as realized by us—were suggested too. In this case the S is an electrophilic center. That means the program reflected the amphiphilic character of thiophosphorus groups in a correct manner.

The next is an example for a true prediction. For some times we were dealt with syntheses and reactions of N-phosphorylated sulfinylamines.[13] Among others we investigated the formation of thiazin-S-oxides (**1**) from sulfinylamines. The retrosynthetic analysis in TRESOR (Figure 4) showed the formation of them by a Diels–Alder reaction of dienes with sulfinylamines. This proposal is not as trivial as it seems, because the reactivity of sulfinylamines is great influenced from the substituent R in R–NSO and cyclizations generally are not favored. Furthermore TRESOR was capable of giving correct disconnection of (**1**); this is not a simple matter in heterocycles with two heteroatoms in 1,2 position.
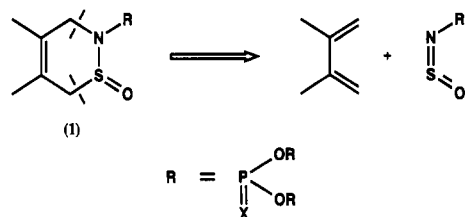
Context Description in Synthesis Planning

J. Chem. Inf. Comput. Sci., Vol. 34, No. 1, 1994 **119**

Figure 4. Retrosynthetic analysis of formation of thiazin-*S*-oxides from dienes and sulfinylamines.

FINAL REMARKS AND CONCLUSION

In the present state TRESOR exists in a test version on PC and consists of a structure editor, modules for file synthonization, canonization, fragmentation, etc., and the first development step of atom keys for evaluation (see above). Furthermore there are modules for automatic generation of tautomers and mesomers of a given structure, the calculation of topological indices, etc.; they are integrated in TRESOR.

In the following some aims for future developments are short outlined: enlargement and reorganization of synthon library on the basis of large reaction databases and improvement of synthonization algorithms (e.g., "separation" of synthons from their origin; it could be shown in some cases, that in such a way it is possible to come to unprecedented reactions); advancement of context description; flexible shell design; interfaces to different reaction databases; application of TRESOR in combination with reaction retrieval (not only for synthesis planning and reaction prediction, but also for method search); platform independence.

REFERENCES AND NOTES

(1) Moll, R. CARSA, a computer-assisted system for synthesis research in industry. *Anal. Chim. Acta* **1990**, *235*, 189–193.
(2) Weise, A. Derivation of organochemical reactions with the AHMOS [automated heuristic modeling of organochemical syntheses] simulation program. *Z. Chem.* **1975**, *15*, 333–340.
(3) Weise, A. Computerized analysis of chemical reaction equations with correction of stoichiometry. *J. Prakt. Chem.* **1980**, *322*, 761–768.
(4) Moll, R. CARSA – Computer Project for Synthesis Research in Industry. *Mitteilungsbl.—Ges. Dtsch. Chem., Fachgruppe Chem.-Inf.-Comput.* **1992**, *No. 22*, 4–10.
(5) Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz, A.; Fontain, E.; Gruber, B.; Herges, R.; Knauer, M.; Reitsam, K.; Stein, N. Computer-assisted solution of chemical problems: a new discipline in chemistry. *Angew. Chem.* **1993**, *105*, 210–239.
(6) Gasteiger, J.; Marsili, M.; Hutchings, M. G.; Saller, H.; Löw, P.; Röse, P.; Rafeiner, K. Models for the Representation of Knowledge about Chemical Reactions. *J. Chem. Int. Comput. Sci.* **1990**, *30*, 467–476.
(7) Moll, R. In *Strategies for Computer Chemistry*; Tosi, C., Ed.; Kluwer Academic: Dordrecht, The Netherlands, 1989; pp 213–226.
(8) Senf, I.; Schädler, H. D. Unpublished work.
(9) Corey, E. J.; Long, E. K.; Lott, G. I.; Rubenstein, S. D. Computer-assisted synthesis analysis. Quantitative assessment of transform utilities. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 304–309.
(10) Jorgensen, W. L.; Laird, E. R.; Gushurst, E. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. *Pure Appl. Chem.* **1990**, *62*, 1921.
(11) The conception was developed common with H. D. Schädler (Halle University) and his co-workers.
(12) Moll, R.; Meinhold, H. Synthesis of New Thiophosphates from Chlorosulfane Compounds. *Z. Chem.* **1989**, *29*, 60–61.
(13) Doerfelt, S.; Moll, R. Synthesis of N-phosphorylated sulphinyl compounds and reactions with various dienes. *Phosphorus, Sulfur, Silicon Relat. Elem.* **1993**, *75*, 119–122.