Reaction Databases and Synthesis Planning—Combined Application and Synergetic Effects†

Rainer Moll

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

Received August 30, 1996⊗

Reaction databases are recognized as very useful tools for the organic chemist, whereas computer-aided synthesis planning does not play the same role up to now. The reason for this, among others, is that synthesis is seen as a creative process, and the assistance of it by computer programs is still considered as suspect. On the other hand, it could be shown that information extracted from reaction databases can be improved and completed by using synthesis planning programs. Interesting possibilities for the development of new synthesis schemes by combined utilization of both tools will be shown by some examples.

INTRODUCTION

The usual situation for a synthesis chemist is to synthesize new structures or to find a better synthesis for a known compound. To solve the problem, the chemist will try to get as much information as possible. In general the information sources are the printed literature, not forgetting the chemist's own knowledge and experience. But since the 1960s—to a steadily increasing extent—computer-aided sources have become available: databases, calculation-, simulation-, modeling-programs, etc. For the synthetic chemist, reaction databases and synthesis planning programs are of greatest interest; both are used, but in a very different manner and degree and are almost always isolated from each other.

In Table 1 we try to compare reaction databases and synthesis planning programs with their different possibilities and functions. In this paper it is not necessary to give examples of reaction databases and synthesis planning programs and to discuss the differences between the individual systems, because there are some profound reviews with detailed information, e.g.,[1−5] and literature cited therein.

DISCUSSION

As a simplification one could say that both databases and synthesis planning programs represent "chemical knowledge" but in a different manner. Databases are a more or less complete pool of collected knowledge including tools for searching reactions. Synthesis planning tries to present or simulate possible reaction steps on the basis of special formalisms, facts, and rules. "Formalisms" can be formalized reactions, like transforms[6] or synthons[7] or mathematical models,[4] "facts" are thermochemical or similar data,[8] and finally "rules" can be mechanisms,[9] heuristic algorithms,[10] context descriptors,[11] and so on. Synthesis planning makes suggestions about structural changes or transformations of educts into products, that is "what happens" and to some extent also "why it happens" but not about "how to do it". The source for the latter information is—or should be—a reaction database.

There are a number of databases and synthesis planning programs, but acceptance by chemists is quite different. The application of reaction databases is much stronger, while usage of synthesis planning is still an exception. This statement is corroborated by statistics about the frequency of application given in a recent review article.[5] It is interesting to read that synthesis planning programs are used in most cases only out of curiosity. Our experience is that in many cases chemists make queries so simple or so overcomplicated that they can never expect to get a proper answer.

What are the reasons for these prejudices and poor acceptance? The reasons, among others, are that synthesis is seen as a creative process, influenced mainly by the intuition, experience, and preparative skill of chemists. The assistance of such a process by computer programs is still considered as suspect. As natural scientists, chemists have confidence in "hard" and strong reproducible data or facts, as they are represented in factual databases. The facts in synthesis planning programs are—from the chemist's point of view—"soft" facts (although the basis for them is "hard"!). But it has to be recognized that synthesis planning programs are not designed to offer complete solutions; they are designed as intellectual stimulating tools for assisting problem solving and decision making processes.

One may not forget that the development of computer programs and dealing with problems of synthesis planning have led to an increasing understanding of problems of chemical reactivity, identification of reaction centers, reaction types, and reaction classification. It has promoted thinking about strategy and tactics and the systematization and categorization of syntheses. Of course synthesis planning has gotten positive results in its own field: designing of new syntheses for target structures and successful reaction predictions (examples s[4,6,9,12]).

The situation we have described could give the impression that databases and synthesis planning are diametrically opposed subjects. But on closer examination it can be realized, that they complement each other and an adequate combined application leads to synergetic effects. In order to support this thesis, in the following some examples from our own work are given.

EXAMPLES

For the investigations we used our synthesis planning program TRESOR (Tracing and Evaluation of Synthons in

† Presented at 4th International Conference on Chemical Structures, Noordwijkerhout (NL), June 1996.

⊗ Abstract published in *Advance ACS Abstracts,* December 15, 1996.

S0095-2338(96)00467-2 CCC: $14.00

Table 1

	definition	objective
reaction databases	computer-controlled reaction data poolings and programs for their organization and retrieval	information about single reactions incl. preparative methods, reaction conditions, yields, etc.
synthesis planning programs	computer-aided analysis of synthesis problems	simulation or modeling of reactions or reaction steps for synthesis design (retrosynthetic) or reaction prediction

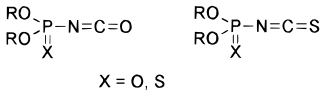

Figure 1. Phosphorus pseudohalogens.

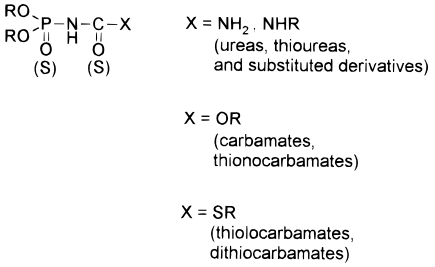

Figure 2. Reaction of phosphorus pseudohalogens with N-, O-, and S-nucleophiles (examples from database).

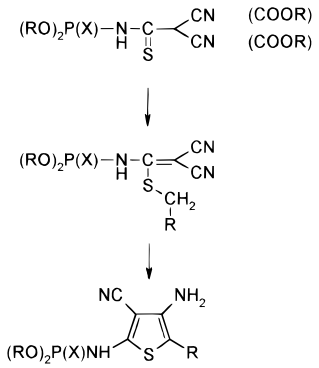

Figure 3. Formation of 2-phosphorylamidothiophenes.

Organic Reactions), whose principles were described in detail in refs 11 and 12. The main idea is the use of a synthon library as knowledge base, which is produced by extracting reactions from the literature. As reaction database we applied the "VINITI/ZIC" database, formerly called SPRESI, which is also the basis for ChemReact and which we modified in some respects for our internal purposes.

At this point we have to make a short excursion into phosphorus chemistry. In order to find new plant protection agents, we were concerned with phosphorus pseudohalogens (cf. Figure 1) as starting materials. The functional groups with their broad and well-known chemistry[13] offer the possibility for synthesizing a large number of new and structural interesting phosphorus compounds. At first we made a search in our reaction database and then a reaction prediction analysis with TRESOR. In extract, some results are given in Figures 2 and 3. The search in the database gave 278 examples, mainly reactions with N-, O-, and S-nucleophiles (Figure 2 shows in a general manner a very small selection). TRESOR generated 57 suggestions; most of them were also found in the reaction database. But there were also some "gaps", e.g., the formation of dithiourethanes was not shown (the appropriate synthons were still lacking in synthon library). New proposals were reactions with C-nucleophiles. As demonstrated in Figure 3, the reaction of phosphorus isothiocyanates with CH-acidic compounds

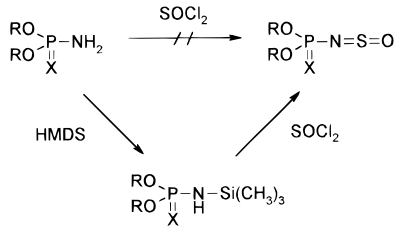

Figure 4. Formation of *N*-sulfinyl phosphorus ester amides.

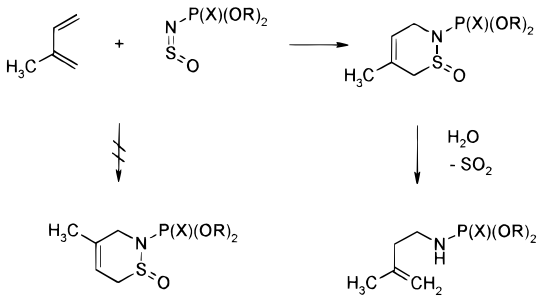

Figure 5. *N*-Phosphoryl-3*H*-thiazin-*S*-oxides.

forms N-phosphorylated thiocarboxamides. After alkylation and under the influence of bases they react further under cyclization to phosphoryl amido thiophenes. Thus, synthesis planning opened the door for the preparation of a new and fungicidally effective class of compounds.[14]

Another type of synthesis building blocks are phosphorus sulfinyl amines. Starting materials for them are phosphorus amides and thionyl chloride. The reaction principle was found in the database, and it was also predicted by TRESOR, but it failed in practice. The reason for it is the low nucleophilicity of nitrogen groups in phosphorus amides compared with amino compounds. Another more general search in the database for thionylation reactions showed that such reactions work with silylated derivatives (silylation was made with hexamethyldisilazane) and now the reaction was successful (Figure 4). But the more interesting part is the following: Sulfinyl amines react with dienes in a hetero-Diels−Alder reaction to 1,2-thiazin-*S*-oxides. With isoprene, two isomers can be formed, but for mechanistic and steric reasons only one resulted (Figure 5). This was specified correctly by the reaction database, while TRESOR showed both isomers. But not in all cases could heterocycles be found, in some others sulfur-free compounds were formed.

A retro-synthetic analysis by TRESOR revealed that thiazin-*S*-oxides can be cleaved hydrolytically under elimination of sulfur dioxide and under formation of acyclic phosphorus amides. In this case synthesis planning gave an example for a not so obvious reaction and showed the way for the correct structure (for experimental details see ref 15).

FINAL REMARKS

With these, admittedly rather simple examples, we wanted to show that reaction databases and synthesis planning are equally valuable tools for the scientific work of a synthesis

chemist and especially that the combined application offers a large number of positive effects. To intensify the interplay between both systems they could be integrated, for instance by linking synthesis planning to large reaction databases in client server systems. Synthesis planning could work not only as a planning server but also as a reaction generator, as a reaction classifier, and as a simulation and modeling system for reaction prediction and synthesis design. The results could be checked for relevance immediately in reaction databases.

ACKNOWLEDGMENT

Thanks are due to G. and H. Fedgenhäuer and S. Doerfelt for preparative work and E. Heinrich and U. Kiedorf for valuable collaboration and discussions.

REFERENCES AND NOTES

(1) Wiggins, G. Chemical Information Sources; McGraw-Hill: New York, 1991.
(2) Ott, M. A.; Noordik, J. H. Computer tools for reaction retrieval and synthesis planning in organic chemistry. A brief review of their history, methods, and programs. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 239−246.
(3) Gasteiger, J.; Ihlenfeldt, W.-D.; Röse, P. A collection of computer methods for synthesis design and reaction prediction. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 270−290.
(4) Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz, A.; Fontain, E.; Gruber, B.; Herges, R.; Knauer, M.; Reitsam, K.; Stein, N. Die computerunterstützte Lösung chemischer Probleme-eine neue Disziplin der Chemie. *Angew. Chem.* **1993**, *105*, 210−239.
(5) Ihlenfeldt, W.-D.; Gasteiger, J. Computer-assisted Planning of Organic Syntheses: The second Program Generation. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2613−2633.
(6) Corey, E. J.; Long, E. K.; Lotto, G. I.; Rubenstein, S. D. Computer-assisted synthesis analysis. Quantitative assessment for transform utilities. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 304−309.
(7) Moll, R. In *Strategies for Computer Chemistry*; Tosi, C., Ed., Kluwer Academic Publ.: Dordrecht, 1989; pp 213−226.
(8) Gasteiger, J.; Marsili, M.; Hutchings, M. G.; Saller, H.; Löw, P.; Rafeiner, K. Models for the Representation of Knowledge about Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 467−476.
(9) 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.
(10) Weise, A. Ableitung organisch-chemischer Reaktionen mit dem Simulationsprogramm AHMOS. *Z. Chem.* **1975**, *15*, 333−340.
(11) Moll, R. Context Description in Synthesis Planning. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 117−119.
(12) Moll, R. CARSA—a computer-assisted system for synthesis research in industry. *Anal. Chim. Acta* **1990**, *235*, 189−193.
(13) Larock, R. C. Comprehensive Organic Transformations. VCH Publishers: New York, 1989.
(14) Fedgenhäuer, H.; Moll, R. Unpublished experimental work.
(15) Doerfelt, S.; Moll, R. Synthesis of N-phosphorylated sulfinyl compounds and reactions with various dienes. *Phosphorus, Sulfur, and Silicon* **1993**, *75*, 119−122.

CI960467M