CHEMISTRY LETTERS, pp. 37-40, 1987.

Computer Assisted Bilateral Generation of Reaction Networks from Educts and Products+

Eric FONTAIN, Johannes BAUER, and Ivar UGI
Institut für Organische Chemie der Technischen Universität München,
D-8046 Garching, West-Germany

A computer program (FORTRAN 77 for IBM-PC) for generating reaction pathways between given educts and products of chemical reactions and their sequences is presented. This program is based on the theory of the BE- and R-matrices. Some applications to known chemical reactions are described as examples.

A novel type of computer programs for the deductive solution of a great variety of chemical problems has been implemented at our laboratory in the recent past.1-3) Unlike the customary expert systems of chemistry,4) e.g. the reaction library based synthesis design programs,4,5) the A.I.-type6) computer programs for the deductive solution of chemical problems are not based on storage, retrieval and manipulation of detailed chemical information.1) The theory of the BE- and R-matrices,7) an algebraic model of the logical structure of constitutional chemistry, is the foundation of the deductive approach to computer assistance in chemistry. Accordingly, our programs are not confined to known chemistry, but they are also capable of generating unprecedented chemistry.

In this paper we present a logic-oriented computer program that generates a network of conceivable reaction pathways, or also mechanistic pathways from the given educts and products of a chemical reaction, or a sequence of chemical reactions. It is of interest for the elucidation of reaction mechanisms, metabolic pathways and the bilateral design of multistep syntheses.8)

There are two distinct ways to use the theory of the BE- and R-matrices as a mathematical foundation for computer programs of the aforementioned type:

On the one hand, it is possible to establish an atom-by-atom mapping of the educts onto the products under the criterion of minimal chemical distance.3,9) The R-matrix of such a conversion of the educts into the products can be decomposed into components that may be used to construct pathways of intermediates between the reactants. A corresponding computer program is being developed.

On the other hand, there is the present approach RAIN (Reaction and Intermediate Network). Here we represent both, the educts and the products, by their BE-matrices without generating an atom-by-atom mapping, and we then use each of these as an independent origin of a tree that is manufactured through transformations of BE-matrices under the mathematical fitting conditions and the chosen valence chemi-

---

+ Dedicated to the 60th birthday of Prof. Teruaki Mukaiyama.

cal boundary conditions. For the construction of the trees no fixed set of transformations (i.e. R-matrices) is used. Instead, any transformations are permissible, as long as their multitude can be restricted by boundary conditions. The valence chemical boundary conditions can refer to stable compounds, or to reactive intermediates (including charged atoms, or radicals.10)) As soon as all branches of the educt tree meet branches of the product tree, the network is complete.

The present approach has more combinatorial degrees of freedom than the rigorously formalized, chemical distance based aforementioned approach. In order to reduce the resulting network to the nodes representing intermediates of interest, the program contains the following options :

1) Statement of an upper bound for the total number of reaction steps.
2) Restriction of the allowable chemical elements by limiting the valence chemical schemes and their interconversions.
3) Restrictions of the R-transformations by stating an upper bound for
    a) the number of reactive centers
    b) the number of participating covalent bonds
    c) the number of atoms whose covalent adjacencies change
    d) the number of changing adjacencies
    e) the absolute values of the off-diagonal entries of the R-matrices.
4) Imposition of structural requirements on the intermediates by limiting
    a) the number of covalently connected heteroatoms
    b) the complexity of ring systems
    c) the covalent bond orders b_{ij} for pairs of atoms (A_i,A_j).
5) Statement on which atoms must, or must not participate as reactive centers.

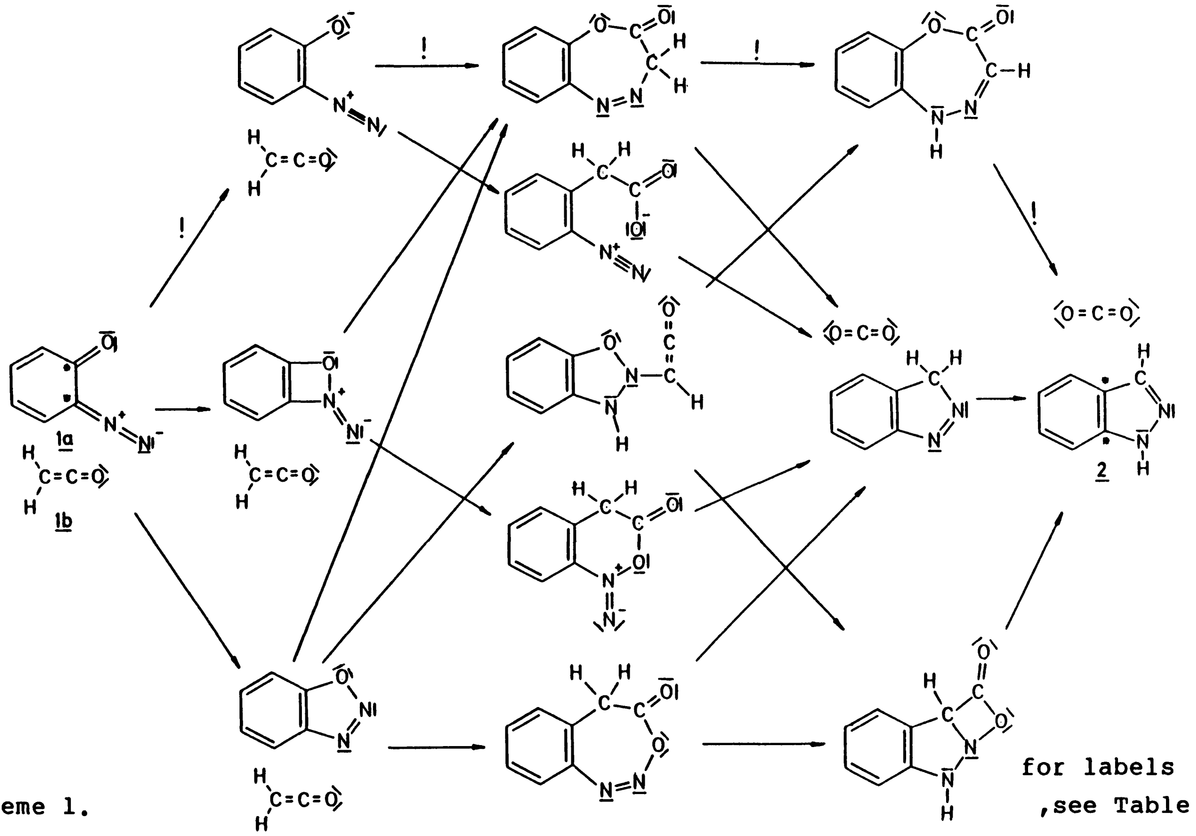
Ried's approach to indazole,11) the thiazole synthesis according to Hantzsch12) and the Favorskii rearrangement13) were used as examples in order to test RAIN. The computing times on an Olivetti PC-M24 (512 KB) were between 30 min. (Hantzsch) and 4-5 hours (indazole). The networks of conceivable pathways were generated with the following restrictions:

Table 1.

Restr.	Indazole	Thiazole	Favorskii
1	5	4	6
2	charged atoms includeda)	no charged atoms	charged atoms includeda
3 a	5	4	4
b	5	4	4
c	5	4	4
d	5	3	2
e	1	1	1
4 a	3	1	1
b	no restriction	no 3,4-membered rings	max. 1 3-membered ring
c	'*' atoms always bonded	no restrictions	no restrictions
5	all atomsb)	all atoms, except 'R'	all atoms, except 'R'

a) max. 2 charged atoms, '*'=sp^2. b) except aromatic C-atoms not labelled with '*'.

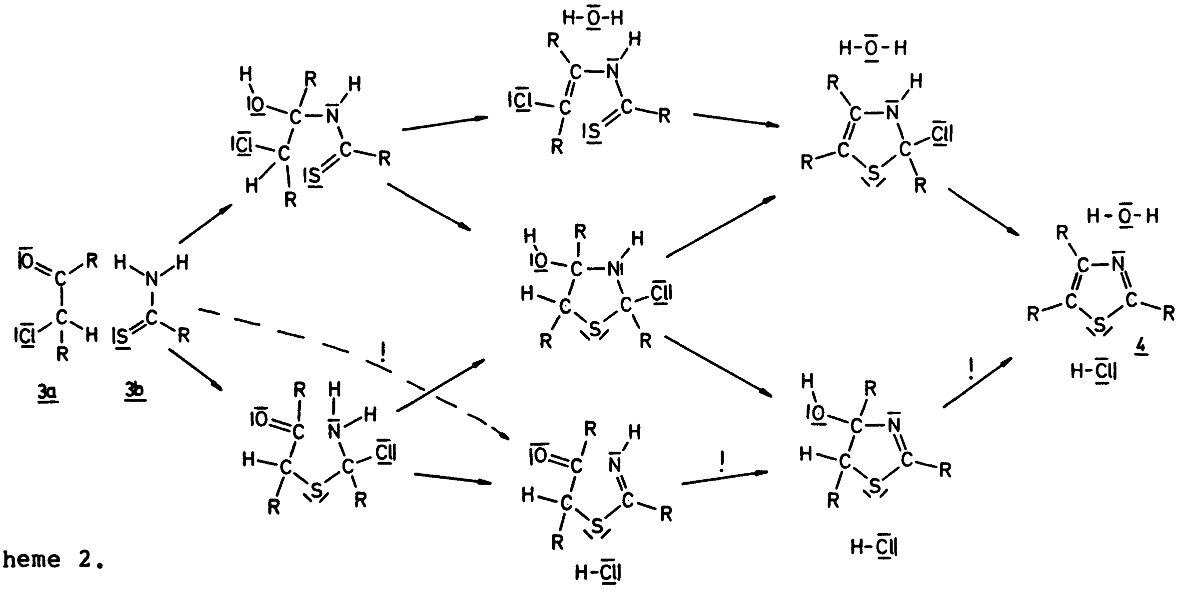
Scheme 1 shows the complete network of the synthesis of indazole 2 from o-quinone diazide 1a and ketene 1b, generated under the above cited restrictions. The generally accepted reaction pathway is labelled with '!'.



Scheme 1.

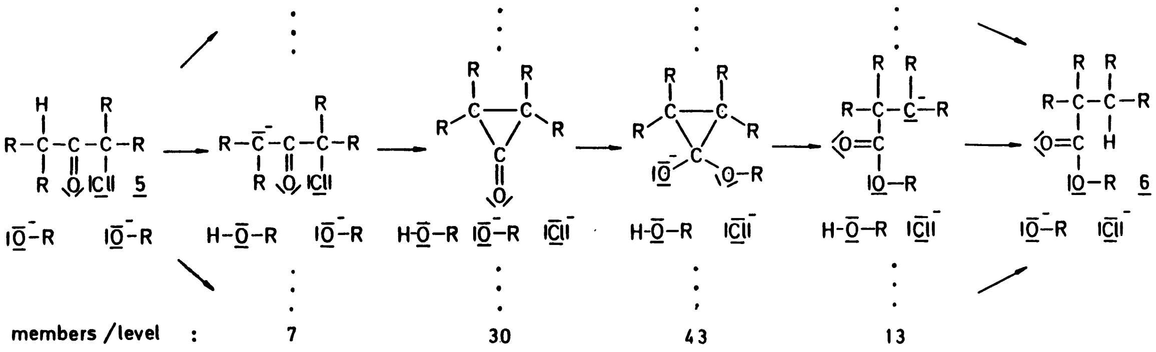
for labels '*', see Table 1.

Scheme 2 shows the RAIN-generated network for the thiazole synthesis according to Hantzsch. The choice of up to 6 reactive centers[14] would integrate the first reaction step that is marked with '!' into the network. However, under this condition, the complete network would be too big to be included here (23 nodes in 4 levels).[15]



Scheme 2.

The Favorskii rearrangement of the haloketone 5 into ester 6, yields a network of 116 nodes and 6 levels; it contains the published reaction mechanism[13] as a subgraph.[15]



Scheme 3. (only a subgraph is shown).

We wish to gratefully acknowledge the financial support of this work by 'Deutsche Forschungsgemeinschaft' and 'Fonds der Chemischen Industrie'.

References

1) I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, and W. Schubert, Angew. Chem., 91, 99 (1979); Angew. Chem., Int. Ed. Engl., 18, 111 (1979).
2) M. Wochner, J. Brandt, A. v. Scholley, and I. Ugi , Acc. Chem. Res, (submitted).
3) J. Bauer, R. Herges, E. Fontain and I. Ugi, Chimia, 39, 43 (1985).
4) M. Wochner and I. Ugi, Chem. Ind., 109, 498 (1986).
5) J. H. Winter, "Chemische Syntheseplanung," Springer, Berlin (1982).
6) "Artificial Intelligence," ed by M. Yazdani, Chapman and Hall, London (1986).
7) J. Dugundji and I. Ugi, Top. Curr. Chem. 39, 19 (1973).
8) I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, and J. Dugundji, "Computational Methods in Chemistry," ed by J. Bargon, Plenum Press, New York (1980), p. 275.
9) C. Jochum, J. Gasteiger, and I. Ugi, Angew. Chem., 92, 503 (1980); Angew. Chem., Int. Ed. Engl., 19, 495 (1980).
10) I. Ugi, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, P. Lemmen, and W. Schubert, Pure Appl. Chem., 50, 1303 (1978).
11) W. Ried and R. Dietrich, Angew. Chem., 75, 476 (1963); Angew. Chem., Int. Ed. Engl., 2, 323 (1963).
12) J. Metzger, "Compr. Heterocycl. Chem.," ed by A.R. Katritzky, C.W. Rees, and K.T. Potts, Pergamon, Oxford (1984), Vol. 6, p. 294.
13) A. Baretta and B. Waegell, "React. Intermed.," ed by R.A. Abramowitch, Plenum Press, New York (1982), Vol. 2, p. 527.
14) RAIN has no upper bound for the number of atoms in the reactive center, but computing time will limit it to 15 to 20 atoms.
15) A listing of the network can be obtained from the authors.

(Received August 25, 1986)