MOLECULAR DESIGN OF HETEROCYCLES.

4. USE OF COMPUTERS IN THE CHEMISTRY OF HETEROCYCLES (REVIEW)

D. E. Lushnikov and E. V. Babaev

Fundamental principals of the utilization of computers in the design of heterocyclic structures and reactions are examined and systematized in this review. The authors' original GREN program, which lists the recyclications of heterocycles, and Heterocycland program, which is based on the previously proposed [2] "structure-synthesis" concept and is used for the retrosynthesis of six-membered heterocycles, are examined concisely.

The intent of this review is to achieve several goals: first, to familiarize a wide audience of heterocyclic chemists with the methodology and the most recent advances in the utilization of computers in organic chemistry in general and the chemistry of heterocycles in particular, and, secondly, to place emphasis upon the attention of the specialists themselves in the field of computer chemistry on the specific characteristics of the problems that exist in the chemistry of heterocycles and on the known methods for their solution. Finally, the review can be addressed simply to thinking organic chemists who are not indifferent to research at the boundaries of the sciences. For specialists who have a vague idea regarding the subject of this review we have striven to set forth a maximally approachable exegesis of the idea of computer chemistry.

In examining the utilization of computers in chemistry, one cannot help being amazed at the diversity and wide gamut approaches. Quantum-chemical calculations of polyatomic molecules [3] or optimization of the conformations of polycycles by the methods of molecular mechanics [4] can be realized only by means of computers. Independent reviews [5, 6] have been devoted to the key role of computers in the processing of chemical information, and computer programs for establishing the structures of new compounds on the basis of the spectra and other information available regarding them have been created (and are already being used extensively) [7, 8]. Finally, the entire trend of computer chemistry oriented toward the quest for regularities of the "structure-property" and "structure-activity" type (English abbreviations QSPR and QSAR) has been formulated [9-11].

It goes without saying that in the cited approaches, in which the chief goal is the analysis of the static (physicochemical or structural, for example) properties of substances, heterocyclic substances play a far from minor role (simply by virtue of the significance of heterocycles, let us say, as medicinal preparations, dyes, etc.). Moreover, understanding the chief task and discipline of chemistry as being the study of the utilization of substances in chemical processes, we will strive to concentrate our principal attention on the use of computers in the analysis of precisely chemical transformations, i.e., we will examine the computer aspects of the synthesis and reactions of organic substances, namely heterocyclic compounds. Despite the existence of reviews in this field [12-16], the specific characteristics of the computer chemistry of heterocycles, have evidently not yet been the subject of independent study.

COMPUTER DESIGN OF STRUCTURES AND REACTIONS: CLASSIFICATION OF GOALS AND CONCEPTS

In discussing the use of computers for the solution of synthetic problems one may speak of three major "components" of synthesis — the starting reactants, the reactions, and their products. Usually, one or several of these components are unknown and are the subject of investigation in the use of a computer, while others are known. We will assume that these

*For Communication 3, see [1].
components are independent and, consequently, that they can be combined with one another with a "plus" sign (if starting information is available) or with a "minus" sign (if no information is available or it constitutes the task of the investigation). Simple sorting gives us eight possible combinations, which are presented in Table 1 ("Goals and Concepts of Computer Chemistry"), which is, evidently, presented for the first time. The fundamental difference between the information on the substances and the information on the reactions should be emphasized; the former is subject to quite simple processing, while the latter is not.* We therefore arbitrarily divided the table into four possible types of tasks that reflect the existence or absence of data regarding the reactants and products and the different nuances of knowledge regarding the way in which the reactions occur.

Let us note that the mathematics- and logic-oriented abstract conception of the reactions in the literature is not infrequently typified by the words "formal," "logical," or a combination of these words [12]. The opposite approach is designated as "empirical" or "informational-oriented."

Cases I-III (plus signs in the "Products" column) correspond to retrosynthetic planning of the synthesis, i.e., to finding ways to synthesize the desired structure (subsequently abbreviated DS), regardless of which method and regardless of the substances from which the synthesis is effected. In actual practice, it is necessary to introduce restrictions on the methods (strategy) of synthesis. In case I, moreover, information regarding the available reactants is available.

Case IV (the DS and reactants are available) corresponds to a "lack of freedom" in the selection of the reactants for the DS (suitable reactants should be available in the laboratory or in a catalog). In addition, a certain degree of knowledge regarding the reactions is not explicitly implied; this makes cases IV and I similar.

Cases V and VI (a set of reactants is predesignated). The task of the direct (or synthetic) planning of the synthesis consists in prediction of the products of the reaction between the reactants under predesignated conditions. Elementary knowledge (case VI) regarding the rules under which the reactions proceed is possible.

Case VII (knowledge regarding the reactions is available) corresponds to the classification and quest for new reactions and the study of their mechanisms. The selection of the optimum substances for the realization of abstract ideas is relevant here.

There are idiosyncrasies for extreme cases I and VIII. The first case (the goals are set, the reactions are known, and the reactants are given) correlates with experimental practice. The second case (the lack of reactants, knowledge of the reactions, and the goals of the investigation) also implies the lack of computers.

Below we will discuss all of the cases using heterocyclic structures as the reactants and/or products and placing emphasis upon attention to the specific characteristics of knowledge of heterocyclic reactions (for example, recyclizations or "magic" rules for the cyclizations of heterocycles). The examples presented in the schemes will play the chief role: all of them are predicted using computers (in place of the reactants, above the arrows we will place the name of the program in which the reaction is proposed). The experimental correlation will be specified individually for each case.

RETROSYNTHETIC ANALYSIS (APPROACHES I AND II)

In giving a unique "retrospective of the ideas of retrosynthesis" let us note that the idea to use computers to search for pathways for the synthesis of organic compounds was first proposed back in 1963 by Vleduts, a former coworker in the All-Union Institute of Scientific and Technical Information (VINITI, Moscow). In 1967, Corey, whose name is usually associated with the birth of computer retrosynthesis, proposed the concept of retrosynthetic analysis of a desired molecule and developed the general principles of synthesis planning [18].

In the Corey concept the result of analysis is a set of synthetic precursors, and the application of the retrosynthesis procedure to it leads to the formation of a synthesis tree. Thus two features are fundamental in the development of a plan of synthesis: 1) both to generate synthetic precursors starting from predesignated desired structures (DS) (tactics) and 2) to control the construction of the synthesis tree in order to avoid "combinatorial explosion" and to find the optimum pathways in this tree (strategy).

*Information regarding the substances — the reactants and products — is usually data (either structures or a list of structures such as a catalog of commercially available reagents): one either has them, or one does not. On the other hand, the information entered into the computer regarding reactions is a certain degree of knowledge as to how the transformations of the compounds occur.