

T04. Chasing molecules that were never there: misassigned molecular structures and the role of computer-aided systems in structure elucidation

M. Elyashberg

Advanced Chemistry Development, Moscow, Russia

The presentation was encouraged by the excellent review [1] of Nicolaou and Snyder entitled "Chasing Molecules That Were Newer There: Misassigned Natural Products and the Role of Chemical Synthesis in Modern Structure Elucidation". The title of our presentation was obviously chosen taking into account to underline connection between the problems discussed in both reports.

According to the review [1], around 1000 articles were published between 1990 and 2004 where originally determined structures were revised. Figuratively speaking, it means that 40-45 issues of the imaginary "Journal of Erroneous Chemistry" were published where all articles contained only mistakenly recognized structures and, consequently, at least the same number of articles was necessary for revision of these structures. The labor (and not only labor) expenses necessary for structural misassignments and subsequent reassignments become at least twice greater than in the case of getting initially correct solution.

It is evident that the stream of publications in which structures of new natural products and new products of organic synthesis determined incorrectly is large enough and there is a problem how to diminish this stream. The authors [1] comment that "there is a long way to go before natural product characterization can be considered a process devoid of adventure, discovery, and, yes, even unavoidable pitfalls".

We suppose that application of modern Computer-Aided Structure Elucidation (CASE) systems (see review [2]) can frequently help the chemist to avoid falling into a pitfall or, if nevertheless the researcher found himself in the pitfall, the expert system can give a signal "caution!" ("it seems you are in pitfall", "be carefully!"). Our hope is based on the fact that the molecular structure elucidation can be formally described as deducing all (without any exclusions) logical corollaries from a system of statements which ultimately form a partial axiomatic theory related to a current spectrum-structural problem. These corollaries are all conceivable structures that meet the initial system of axioms.

The history of CASE systems development convincingly confirmed the point of view suggested 40 years ago [3] that, as a matter of fact, the process of molecular structure elucidation is reduced to logical inferring the most probable structural hypothesis from a set of statements reflecting the interrelation between a spectrum and a structure. This methodology was implicitly used for long time before computer methods appeared. Independently on application or ignoring computer-based methods the way to the target structure is the same. CASE expert systems mimic considerations of human expert. The main advantages of CSASE systems: 1) all statements about interrelation between spectrum and structure ("axioms") are expressed explicitly; 2) all logical consequences (structures) following from the system of "axioms" are deduced completely, without any exclusions; 3) the process of the computer-based structure elucidation is very fast, which gives a tremendous saving of time and labor of the scientist; 4) if the chemist has several alternative sets of

axioms related to a given structural problem, an expert system allowed rapid generating all consequences from each of the sets and identify the most probable structure by comparing the solutions obtained.

In our presentation the main kinds of "axioms" used during the molecular structure elucidation are discussed. The axioms are classified in the following three groups: 1) Axioms and hypotheses reflecting characteristic spectral features; 2) Axioms and hypotheses of 2D NMR spectroscopy; 3) Structural axioms necessary for structure assembling. When the structure elucidation of a new chemical compound is performed with assistance of expert systems all axioms are expressed explicitly. Therefore it becomes possible to investigate dependence of the solution of a structural problem on any change in the initial set of axioms.

We will consider a series of examples in which the original structures suggested by researchers were revised later in successive works. All examples were taken from recent publications in respectable international journals. In each case, we show how the structure could be quickly and correctly identified if reliable MS and 2D NMR data were available and expert system Structure Elucidator [4] was employed. We will also show that if only 1D NMR spectra are available, empirical calculation of ¹³C chemical shifts [5] for suggested structures frequently allow researcher to realize that some structural hypothesis is most probably incorrect. Figuratively speaking, Structure Elucidator can be used as an analytical tool resembling a "polygraph detector".

The considered approach and presented examples allow one to come to conclusion that application of an expert system similar to the Structure Elucidator for the structure elucidation of new complex organic compounds, particularly natural products, can prevent inferring incorrect structures, which is not excluded even for highly qualified and experienced organic chemists. It is possible to expect that worldwide application of CASE systems will reduce the stream of publications containing erroneously elucidated chemical structures.

References:

1. K.C. Nicolaou, S.A. Snyder. *Angew. Chem. Int. Ed.*, 2005, **44**, 1012-1044
2. M. E. Elyashberg, A.J. Williams, G. E. Martin. *Prog. NMR Spectrosc.* 2008, **53**, (1/2) , 1-104
3. M.E. Elyashberg, L.A. Gribov, V.V. Serov. *Molecular Spectral Analysis and Computers* (in Russian). Nauka, Moscow, 1980, 308 P.
4. M. E. Elyashberg, K. A. Blinov, A. J. Williams, S. G. Molodtsov, G. E. Martin. *J. Chem. Inform. Model.* , 2006, **46**, 1643-1656
5. K. A . Blinov, Y. D. Smurnyy, T.S. Churanova, M. E. Elyashberg, A. J. Williams. *Chemom. Intell. Lab. Syst.*, 2009, **97**, 91-97