

Preface

*When studying science,
the examples are more useful than the rules.*

Isaac Newton

The challenge of molecular structure elucidation has been a primary focus in the field of organic chemistry since its origin. A structural formula is the simplest informative model of a molecule. If the structure is known then contemporary semi-empirical and quantum-chemical methods of theoretical chemistry allow for the prediction of many molecular properties (generation of an optimized 3D model, molecular spectra, physicochemical parameters, biological activities, etc.) with an accuracy which usually meets the requirements for the majority of practicing chemists.

The most challenging problem is the structure elucidation of *new* compounds which are obtained by chemists either as products of synthesis or, for example, as compounds isolated from biological species. Especially challenging is the structure elucidation of natural products characterized by unexpected and unprecedented skeletons and the scope of investigations in this field is very broad. For instance, in the past 10–12 years, more than 20,000 and more than 30,000 new marine-derived, and higher plant-derived compounds, respectively, were isolated and structurally characterized with spectroscopic methods (J. Bérdy, J. Antibiot. 2012, 65, 385–395) playing the decisive role in the structure determination of organic molecules. During the mid-1960s researchers realized that the most promising approach to solve this problem would be using a combination of MS, NMR, and IR spectroscopic methods and to perform data analysis using computers. The result was a new form of molecular spectral analysis—Computer-Assisted Structure Elucidation (CASE).

CASE-based computer programs are called *artificial intelligent systems* or *expert systems*. Generally speaking, these systems mimic the expert's way of thinking during the process of structure elucidation using spectroscopic data, even though the computer analysis differs significantly from human reasoning.

As a result of the efforts of many researcher groups worldwide, expert systems have become available which allow chemists to quickly and reliably elucidate the

structures of new complex organic molecules containing a hundred or more skeletal atoms. The state of the art in CASE applications has been described in our previous monograph “Contemporary computer-assisted approaches to molecular structure elucidation,” RSC Publishing, Cambridge, 2012, 482 p.

Among the available expert systems *ACD/Structure Elucidator* is probably the most advanced at this time. It is used in many academic and industrial organizations mainly for the structural characterization of new natural products, drug analysis and the identification of drug impurities and degradants, etc. The system cumulatively employs data sets acquired from MS, NMR (1D and 2D), and IR spectroscopic experiments, however, the 2D NMR spectra play the decisive role as carriers of very rich structural information. Different 2D NMR techniques (HSQC, HMBC, COSY, NOESY/ROESY, etc.) are known to be indispensable for the structure elucidation of complex molecules.

The goal of this book is to help Ph.D. and advanced students and academic and industrial chemists to master the art of structure elucidation with the aid of a contemporary expert system. *ACD/Structure Elucidator* is used as an example of an expert system. We believe that individuals who familiarize themselves with this book will be able to use the program for the purposes of CASE in their everyday work. As far as we know this is the first textbook which explains not only the main ideas associated with CASE but also gives the reader the possibility to understand the different CASE strategies for solving complex real-world problems using a series of examples in the process.

The book is composed of three parts. Part I contains a concise description of the *ACD/Structure Elucidator* flow diagram, its knowledgebase, and the fundamental concepts making up the theoretical basis of CASE (Chap. 1). Different strategies regarding the application of the system depending on the specific features of the problem being solved are discussed in Chap. 2.

Special attention is placed on the explanation of the *axiomatic nature* of the initial information used to logically infer the structure of an unknown. It is shown how *ACD/Structure Elucidator* can deduce correct structures from fuzzy, incomplete, and contradictory statements (set of “axioms”) composing the initial information. It is worth noting that an axiomatic approach is a cognitive basis not only for CASE, but also for organic qualitative analysis. Chapter 2 describes all modes of structure elucidation provided by the expert system and instructs the student when and how each mode can be effectively employed.

Part II can be considered as an introduction to practical approaches used for structure elucidation based on the application of the expert system. For this purpose 22 relatively simple structural tasks adopted from the textbook by M. Reichenbächer and J. Popp, “Challenges in molecular structure determination,” Springer, 2012 serve as examples of the structure elucidation from MS, 1D and 2D NMR, IR and UV spectra. The reader has an opportunity to compare manual solutions to the problems explicitly explained in the cited textbook with those obtained with the aid of *ACD/Structure Elucidator*. A detailed description of the solutions to the problems is available on a Springer server (<http://extras.springer.com/2012/978-3-642-24389-9>). The student is given the unique possibility to repeat all CASE analyses to obtain the

solution to each problem as described in Part I. For this purpose the student can use a limited version of ACD/Structure Elucidator which can be downloaded for free from ACD/Labs (www.acdlabs.com/TeachingSE) server. All spectroscopic data acquired for each task are already presented in electronic formats appropriate for use with the program. We believe that the reader who works through Chap. 3 in combination with solving all challenges will acquire the knowledge and skill necessary to solve complicated real-world problems.

Part III is the most important for those who want to become proficient in routine applications of CASE analysis for solving structural problems which appear in analytical laboratories. Here we fully explain CASE-based solutions to 66 real-world structural problems for which spectroscopic data were adopted mainly from *Organic Letters* and *Journal of Natural Products*, the corresponding articles being published in recent years (2011–2013). For computer-based structure elucidation, we tried to select mainly those problems that were related to molecules possessing unique or unprecedented skeletons. Spectroscopic data for these problems are also available in the form of electronic tables coded in the formats needed by Structure Elucidator. The student therefore has the possibility to repeat the solutions described in Part III and perform additional computational experiments to follow how the results change depending on the composition of the initial axiom set. Moreover, to further test their skills a student can try to solve a problem without the book and then compare the solution obtained with that described.

Part III is divided into two chapters. Chapter 4 describes problems which are solved using *Strict Structure Generation*. This program mode assumes that all HMBC and COSY correlations are of “standard” lengths corresponding to the coupling constants ${}^{2-3}J_{\text{CH}}$ and ${}^{2-3}J_{\text{HH}}$ correspondingly. More challenging problems are collected in Chap. 5. These problems are solved using *Fuzzy Structure Generation*—a very sophisticated approach which allows problems to be solved under the condition that an *unknown number* of correlations of *unknown “nonstandard” lengths* (>4 bonds) are present in the 2D NMR data. It has been shown that this approach allows the researcher to solve problems that would otherwise likely remain unresolved. The Fuzzy Structure Generation approach significantly enhances the ability of a scientist to perform structure elucidation using 2D NMR data.

For those students who want to check their ability to determine molecular structures using ACD/Structure Elucidator without any assistance, a set of prepared tasks are provided (www.acdlabs.com/TeachingSE). Detailed descriptions of their solutions can be found in the textbook by P. Crews, J. Rodriguez and M. Jaspers “Organic Structure Analysis,” Oxford University Press, N.Y., 2010. Taking into account the fact that most readers have no experience in the use of the ACD/Structure Elucidator program we tried to provide detailed explanations for each real-world problem discussed in Part III to allow the reader to solve almost any of the described problems without needing to review previous problems. We hope that this capability will help the student to better understand the reasoning typical for a “CASE equipped” chemist and convince the scientist that an expert system like ACD/Structure Elucidator is already a mature system. The problems considered allow the reader to realize that the ACD/Structure Elucidator is not a robot

intending to exclude a human expert from the process of structure elucidation, but is actually a powerful amplifier of human intelligence, an engine for inferring all logical corollaries (structures) from NMR spectroscopic data, using a formal representation of spectrum-structural knowledge and assumptions introduced by the chemist. The reader will see a large number of examples demonstrating that employing a CASE approach can lead to dramatic acceleration of solving problems in comparison with a traditional approach and makes the solution more reliable. We believe that expert systems similar to ACD/Structure Elucidator are ready to become workhorses in the analytical laboratories.

We hope that this book will attract the attention of university teachers and students as well as those organic chemists whose work is associated with molecular structure determination from NMR spectra.

We are sincerely indebted to our colleagues at Advanced Chemistry Development (ACD/Labs) for their support of this project. The ACD/Structure Elucidator software is more than simply a software “product”. Rather it is the culmination of a vision that was set by management almost two decades ago. The intention was to produce a CASE software package that was more than simply a research project but rather a product that would be used by scientists around the world in different laboratories, would be challenged with a diverse set of structures on an ongoing basis, and would require continued vigilance in terms of supporting the latest experimental techniques. ACD/Structure Elucidator is that product that represents contributions from many ACD/Labs employees over the years. From the simple entry of a chemical structure, through the various modes of NMR prediction (underpinned by expertly curated databases) to the complex data handling required for 1D and 2D NMR, mass spectrometry data, infrared data, and even integration with our many of the other algorithms developed within the organization, ACD/Structure Elucidator represents many hundreds of man years of work at this point as so many people have contributed to the software suite in its entirety.

We also acknowledge our external collaborators for challenging us with their data, for working through some of the technical challenges that have been undertaken and, most of all, for having faith that a solution could be developed.

We especially thank the management at ACD/Labs for allowing us to work on this project as it represents their long-term faith and support in this project. We are especially indebted to one of our colleagues, Rostislav Pol, as he created the custom ACD/Structure Elucidator version supplied to readers of this book. His contribution in enhancing the educational effectiveness of this book is invaluable.

We are eternally thankful to Natasha Elyashberg, wife of Mikhail, for her great assistance in the final preparation of the manuscript for publication.

<http://www.springer.com/978-3-662-46401-4>

Computer-Based Structure Elucidation from Spectral
Data

The Art of Solving Problems

Elyashberg, M.E.; Williams, A.J.

2015, XVI, 447 p. 536 illus., 352 illus. in color.,

Hardcover

ISBN: 978-3-662-46401-4