

Lecture Notes in Chemistry 89

Mikhail E. Elyashberg
Antony J. Williams

Computer- Based Structure Elucidation from Spectral Data

The Art of Solving Problems

 Springer

Lecture Notes in Chemistry

Volume 89

Series editors

Barry Carpenter, Cardiff, UK

Paola Ceroni, Bologna, Italy

Barbara Kirchner, Leipzig, Germany

Katharina Landfester, Mainz, Germany

Jerzy Leszczynski, Jackson, USA

Tien-Yau Luh, Taipei, Taiwan

Nicolas C. Polfer, Gainesville, USA

Reiner Salzer, Dresden, Germany

The Lecture Notes in Chemistry

The series Lecture Notes in Chemistry (LNC) reports new developments in chemistry and molecular science-quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge for teaching and training purposes. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research. They will serve the following purposes:

- provide an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas,
- provide a source of advanced teaching material for specialized seminars, courses and schools, and
- be readily accessible in print and online.

The series covers all established fields of chemistry such as analytical chemistry, organic chemistry, inorganic chemistry, physical chemistry including electrochemistry, theoretical and computational chemistry, industrial chemistry, and catalysis. It is also a particularly suitable forum for volumes addressing the interfaces of chemistry with other disciplines, such as biology, medicine, physics, engineering, materials science including polymer and nanoscience, or earth and environmental science.

Both authored and edited volumes will be considered for publication. Edited volumes should however consist of a very limited number of contributions only. Proceedings will not be considered for LNC.

The year 2010 marks the relaunch of LNC.

More information about this series at <http://www.springer.com/series/632>

Mikhail E. Elyashberg · Antony J. Williams

Computer-Based Structure Elucidation from Spectral Data

The Art of Solving Problems

 Springer

Mikhail E. Elyashberg
Moscow Department
Advanced Chemistry Development
Moscow
Russia

Antony J. Williams
ChemConnector Inc.
Wake Forest, NC
USA

ISSN 0342-4901 ISSN 2192-6603 (electronic)
Lecture Notes in Chemistry
ISBN 978-3-662-46401-4 ISBN 978-3-662-46402-1 (eBook)
DOI 10.1007/978-3-662-46402-1

Library of Congress Control Number: 2015931537

Springer Heidelberg New York Dordrecht London
© Springer-Verlag Berlin Heidelberg 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media
(www.springer.com)

Dedicated to the Scientists whose enthusiasm, clarity of vision, and focused efforts led to the creation and development of a new direction in chemistry—Computer-Assisted Structure Elucidation.

For my wife Natasha, with love and gratitude for patience and assistance.

Mikhail E. Elyashbrg

For my twin boys Taylor and Tyler. Life is complex...but you'll figure it out as you go...

Antony J. Williams

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. Reichenbacher 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.

Contents

Part I The *Structure Elucidator* Expert System

1	Fundamentals of Structure Elucidator System	3
1.1	How Does the Structure Elucidator System Work?	3
1.1.1	Introduction	3
1.1.2	Principle Flow Diagram of Structure Elucidator	11
1.2	Logical Basis of the Spectroscopic Structure Elucidation	13
1.2.1	Axioms and Hypotheses Based on Characteristic Spectral Features	13
1.2.2	Axioms and Hypotheses of 2D NMR Spectroscopy	18
1.2.3	Structural Hypotheses Necessary for the Assembly of Structures	23
1.2.4	Properties of Information Used for the Structure Elucidation	24
1.3	The Knowledge Base of the Structure Elucidator	26
1.3.1	Structure and Content of Factual Knowledge ACD/NMR Database	27
1.3.2	Structure Searching Using a ^{13}C NMR Spectrum	29
1.3.3	Composition of Axiomatic Knowledge	30
1.4	NMR Prediction in the <i>Structure Elucidator</i> System	35
1.4.1	^{13}C NMR Chemical Shift Prediction	37
1.4.2	Prediction of ^1H NMR Spectra	46
1.5	Determination of Relative Stereochemistry	47
	References	49
2	Strategies of Structure Elucidation	53
2.1	Data Input, Processing, and Forming of a Molecular Connectivity Diagram	53
2.1.1	Data Used for Structure Elucidation	54
2.1.2	Molecular Formula	55

2.1.3	Forming the Molecular Connectivity Diagram	58
2.1.4	Checking the MCD for Consistency	61
2.2	Modes of the Structure Generation	65
2.2.1	The “Common” 2D NMR Mode	67
2.2.2	Application of Fragments in Combination with 2D NMR Data	76
2.3	Nonstandard Correlations and Fuzzy Structure Generation	81
2.3.1	Challenge of Nonstandard Spectral Responses	81
2.3.2	Solving Problems Using Fuzzy Structure Generation.	83
2.3.3	Modes of Fuzzy Structure Generation	85
2.3.4	The Strategy of Applying Fuzzy Structure Generation.	90
2.3.5	Is There an Alternative to Fuzzy Structure Generation?	94
	References	94

Part II Getting Started with *Structure Elucidator*

3	Simple Examples of Structure Elucidation	99
3.1	Challenge 1	101
3.2	Challenge 2	109
3.3	Challenge 3	114
3.4	Challenge 4	117
3.5	Challenge 5	121
3.6	Challenge 6	124
3.7	Challenge 7	126
3.8	Challenge 8	128
3.9	Challenge 9	131
3.10	Challenge 10	135
3.11	Challenge 11	138
3.12	Challenge 12	141
3.13	Challenge 13	143
3.14	Challenge 14	145
3.15	Challenge 15	149
3.16	Challenge 16	152
3.17	Challenge 17	155
3.18	Challenge 18	159
3.19	Challenge 19	161
3.20	Challenge 20	167
3.21	Challenge 21	171
3.22	Challenge 22	175
	References	180

Part III Solution of Real World Problems

4	Structure Elucidation Using Strict Structure Generation	183
4.1	Costatol D	183
4.2	Ilesane	187
4.3	Indotertine A	190
4.4	Trefolane A	194
4.5	Spirobacillene A	197
4.6	Spirobacillene B	200
4.7	Lycjaponicum D	202
4.8	Aquatolide	205
4.9	Rubesanolide A	209
4.10	Oxo-agelasine D	213
4.11	Chipericum A	216
4.12	Ascidia SAAF	219
4.13	Hunanamycin A	223
4.14	Periconiasin	226
4.15	Daphmacromine A	229
4.16	Eryngiolide A	233
4.17	Asperjinone	236
4.18	Marinoquinoline F	241
4.19	Schilancitrilactone A	244
4.20	Isocorniculatolide A	248
4.21	Paecilin A	250
4.22	Lycjaponicum A	253
4.23	Virosaine	257
4.24	Geralcin A	261
4.25	Methoxygeissospermidine	266
4.26	Peptidolipin B	270
4.27	Cordyheptapeptide C	274
4.28	Pipestelide C	279
4.29	Tetrabrominated Diphenyl Ether	283
4.30	Indole Alkaloid	288
4.31	Barmumycin	291
4.32	Schizocommunin	295
4.33	Ephelmin A	300
	References	304
5	Problems Solved Using Fuzzy Structure Generation	307
5.1	Gymnopalyne A	307
5.2	Harzianone	311
5.3	Mandelalide A	315
5.4	Puberunine	318
5.5	Ternifolide A	320

5.6	Strophasterol A	322
5.7	Aphanamixoid A	325
5.8	Schiglautone A	329
5.9	Pallambin A	331
5.10	Cristaxenicin A	335
5.11	Juniperolide A	336
5.12	Ligerin	343
5.13	Phosphiodyn A	346
5.14	(2S,3R)-2,3-Epoxy-2,3-Dihydro-8-Hydroxylapachol	352
5.15	Aetheramide	355
5.16	Geranylphenazinediol	358
5.17	Strynuxline A	364
5.18	Hyaladione	369
5.19	Epipancreatistatin	374
5.20	Erythrollic Acid	378
5.21	Farilhydrazone	380
5.22	Physangulidine A	384
5.23	Protuboxepin A	388
5.24	Jatrophalactam	391
5.25	Psammaplysin I	396
5.26	Lasionectrin	402
5.27	Phomentrioloxin	409
5.28	Perenniporide A	413
5.29	Spiroindimicin B	418
5.30	Goniomedine A	423
5.31	Aplidiopsamine A	429
5.32	Polypropionate	432
5.33	Aziridine	438
	References	442
	Glossary	445

Abbreviations and Notations

HSQC	Heteronuclear Single Quantum Coherence
HMQC	Heteronuclear Multiple Quantum Coherence
HMBC	Heteronuclear Multiple Bond Coherence
COSY	COrrrelation Spectroscopy
TOCSY	TOtal Correlation Spectroscopy
ADEQUATE	Adequate sensitivity Double QUAnTum spEctroscopy
INADEQUATE	Incredible Natural Abundance Double QUAnTum Transfer Experiment
NOESY	Nuclear Overhauser Enhancement Spectroscopy
ROESY	Rotating frame Overhauser Enhancement Spectroscopy
MCD	Molecular Connectivity Diagram
FSG	Fuzzy Structure Generation
APCT	Atom Property Correlation Table
PM	Proposed Molecule
<i>k</i>	Number of generated structures
<i>t_g</i>	Processor time for structure generation
<i>r</i>	Position of a structure in the ranked output file
<i>ob</i>	Label which is shown near some atoms in MCD. Denotes necessity of a neighbor heteroatom presence in the first sphere of the given atom environment.
<i>fb</i>	Label shown near atoms in the MCD. Denotes prohibition of a neighboring heteroatom presence in the first sphere of the given atom environment.
<i>n</i> (H)	Total number of hydrogen atoms attached to the carbons which are present in the first sphere of the given carbon atom environment. The value is determined from distinct multiplicities observed in the ¹ H NMR spectrum.

HOSE	Hierarchical Organization of Spherical Environments
d_A	Average chemical shift deviation calculated when a HOSE code based algorithm of the NMR chemical shift prediction was used.
d_N	Average chemical shift deviation calculated when a neural networks based algorithm of the NMR chemical shift prediction was used.
d_I	Average chemical shift deviation calculated when an algorithm of the NMR chemical shift prediction based on incremental approach was used.
d_{complex}	$d_{\text{complex}} = d_N(^{13}\text{C}) + 10d_N(^1\text{H})$
MAE	Mean average error, the same as an average chemical shift deviation
FF	Found Fragments
BadList	List of forbidden fragments
GoodList	List of obligatory fragments
CSB Generator	Correlation Spectroscopy Based Generator
Light blue atom in MCD	A sign of sp^3 or sp^2 hybridization (<i>not</i> sp)
Blue atom in MCD	A sign of sp^3 hybridization
Violet atom in MCD	A sign of sp^2 hybridization
Green atom in MCD	A sign of sp hybridization
Black atom	Atom hybridization is not defined
Green lines in MCD	HMBC connectivities
Blue lines in MCD	COSY connectivities
Violet lines in MCD	Nonstandard connectivities (${}^nJ_{\text{HH}}$ and ${}^nJ_{\text{CH}}$, $n > 3$)
Dotted lines in MCD	Ambiguous connectivities
Red arrows shown on structures	Nonstandard connectivities detected by the Fuzzy Structure Generation
SI	Supporting Information to articles from which spectroscopic data are adopted

Part I
The *Structure Elucidator* Expert System

Chapter 1

Fundamentals of Structure Elucidator System

Abstract This chapter starts with a concise description of the expert system Structure Elucidator flow diagram (selection of possible substructures → structure generation → selection of the most probable structure based on NMR spectrum prediction) and explains how the system works. The fundamental concepts making up the logical basis of spectroscopic structure elucidation are then described. Special attention is placed on an explanation of the *axiomatic nature* of the initial information used to logically infer the structure of an unknown. The following three groups of axioms (statements) are distinguished: (i) axioms of characteristic spectral features in 1D NMR and IR, (ii) axioms of 2D NMR (HSQC, HMBC, COSY, etc.), and (iii) axioms used for the assembly of a structure. The properties of the information used for structure elucidation from 2D NMR data are summarized. It is concluded that to solve real-world problems, Structure Elucidator must be capable of deducing correct structures from a *fuzzy, incomplete*, and frequently *contradictory* set of axioms composing the initial information. It is noted that an axiomatic approach is a cognitive basis not only for CASE, but also for organic qualitative analysis. The knowledge base of Structure Elucidator consisting of *factual* and *axiomatic* knowledge is described and three methods of NMR spectrum prediction included in the system (incremental, neural nets, and HOSE code based) are discussed.

1.1 How Does the Structure Elucidator System Work?

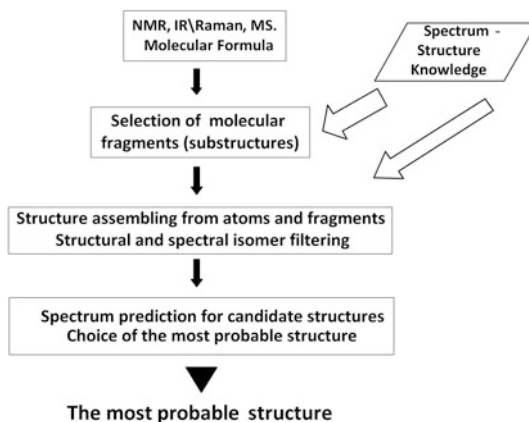
1.1.1 Introduction

When a researcher isolates an unknown chemical compound either as a product of a chemical reaction or as an extract from some material from nature the next step is to determine the structure of the substance. For this goal MS, ^1H NMR, ^{13}C NMR, and IR spectra of the sample may all be registered, and other than a manual structure interpretation approach alternative approaches include using a spectral search

against relevant available databases as well as potentially the data contained within the literature. If, for instance, a spectral search using ^{13}C NMR data leads to the conclusion that the spectrum of the unknown fully coincides with that of a reference spectrum then we can say that the structure of the unknown compound is *identical* to the structure of the reference. This we term as *structure identification*.

If all attempts to identify the structure of an unknown via spectral searching give no positive results, i.e., no reference data can be located or the compound is *new*, then the problem of *structure elucidation* arises. Although spectral searching in databases is by no means a trivial problem, the structure elucidation of a new chemical compound, for instance a complex natural product, is commonly a much more complicated task. Experience shows that to elucidate the structure of a complex organic molecule it is necessary to utilize a wide array of experimental data. The most useful tool in the structure elucidation of complex organic molecules is NMR spectroscopy. The acquisition of ^1H , ^{13}C , ^1H - ^{13}C HSQC, ^1H - ^{13}C HMBC, and ^1H - ^1H COSY spectra is the *general* array of NMR spectra nowadays used for the structure elucidation of complex organic molecules though the problem can frequently be solved also without the application of COSY data. Nowadays it is almost impossible to find a publication devoted to the structure elucidation of a new organic compound without two-dimensional NMR spectroscopy being used. The expert system Structure Elucidator (StrucEluc) [1] was developed to provide the chemist with an intelligent tool capable of assisting in the identification of structures of *new* organic compounds whose properties had not been previously investigated and whose spectra are absent both from the literature and existing databases. To infer a correct structure from the spectroscopic data the expert system has to mimic all stages of the spectroscopist's reasoning process during the structure elucidation procedure. These stages were distinguished as far back as the late 1960s–early 1970s when pioneering works devoted to developing expert systems were initiated [2–6]. A flow diagram illustrating the main steps of the human expert reasoning during the structure elucidation process is shown in Fig. 1.1.

Fig. 1.1 A flow diagram illustrating the main steps of the human expert reasoning process for structure elucidation



During the first stage the molecular formula is determined using mass spectrometry and the available molecular spectra of the unknown which are simultaneously analyzed by the expert using empirical spectrum-structure correlations. The molecular formula is crucial because it sets a limit to the number of all possible isomers whose expected spectra can fit with the experimental data. If the molecular formula is unknown then the number of possible plausible candidate structures is infinite. Nowadays the molecular formula is determined using high resolution mass spectrometry (HRMS).

The spectral analysis based on the spectrum-structure correlations results in the selection of a set of substructures (molecular fragments), whose presence in the molecule under investigation can be suggested on the basis of spectrum-structural correlations and the molecular composition.

The expert then tries to draw structures which can be built from the selected fragments and using “free” atoms not contained within the fragments. This is referred to as “manual structure generation”. In so doing the expert uses their knowledge of characteristic spectral features in the available spectral data to ensure that no structures are produced whose spectra contradict the experimental data. For instance, if the molecular formula contains two oxygens and a singlet is observed in the ^{13}C NMR spectrum at 205 ppm the expert will likely suggest only those structures that contain a ketone group, while all structural hypotheses including an ester group will be rejected. This procedure is equivalent to a manually performed *spectral structure filtering*.

At the same time the expert avoids suggesting structures that contradict chemical rules and chemical common sense. For example, structures that contain 3- and/or 4-membered cycles with a triple bond or structures violating Bredt’s rule will not generally be included in the list of possible candidates. The procedure that provides for the construction of only those structures that do not contradict chemical knowledge is equivalent to *manual structural filtering* of hypothesized molecular structures. It should be obvious that even in those cases when analyzed molecules are of a modest size the number of conceivable structures that can be manually assembled and filtered is much too large to ensure that all structural hypotheses are taken into account.

When an expert selects a set of structures that are equally probable, and none of these structures can be considered as preferred according to their characteristic spectral features, then a more precise structure assessment is necessary. A full spectrum prediction for all hypothetical structures can aid in such an assessment. A calculated spectrum contains not only those features (e.g., NMR chemical shifts, IR frequencies) falling into the spectral intervals, which are characteristic for the structural groups existing in the given molecule (see for example [7]), but it also lists the approximate values of other spectral features (for instance, fingerprint IR frequencies) that are expected in the corresponding spectrum of the evaluated structure. It is worth noting that contemporary quantum-mechanical (QM) methods to perform NMR spectrum calculations are capable of predicting chemical shifts with good enough accuracy even for molecules with unprecedented skeletons.

The structure whose calculated spectrum is closest to the experimental spectrum can be declared as *the most probable structure* of the unknown.

This described methodology was taken as the basis for developing all expert systems intended for molecular structure elucidation from spectroscopic data [6, 8, 9]. At first glance the methodology seems simple and obvious, but realizing the approach using mathematical algorithms and software simulating the reasoning of a human expert is difficult and complex. To formalize all stages of the structure elucidation process algorithms based on mathematical logic, graph theory, and combinatorial analysis were developed. Significant efforts were invested to develop algorithms for structure generation which mimic manual structure assembly from atoms and fragments. The *structure generator* can be considered to be the heart of any expert system and its performance defines, in the final analysis, the power of an expert system.

The structure generator is a program capable of producing all (without any exception) structural formulae of isomers corresponding to a given molecular formula and the imposed structural constraints.

In principle, structure elucidation can be performed using structure generation and structural filtering only, skipping the stage of fragment selection. Figuratively speaking the following analogy can be considered: the expert system acts like a sculptor who removes all superfluous material to produce a beautiful sculpture. The following illustrative example explains this methodology.

Consider that the molecular formula $C_6H_{10}O_2$ for an unknown compound is determined from the HRMS and assume the IR spectrum (Fig. 1.2) is available thereby providing information about typical functional groups in the molecule via characteristic features (frequencies, intensities and half-height bandwidths) of the spectral bands. Careful analysis of the spectrum using reference data from [7, 10] shows that the molecule does not contain alcohol ($-OH$) or vinyl groups

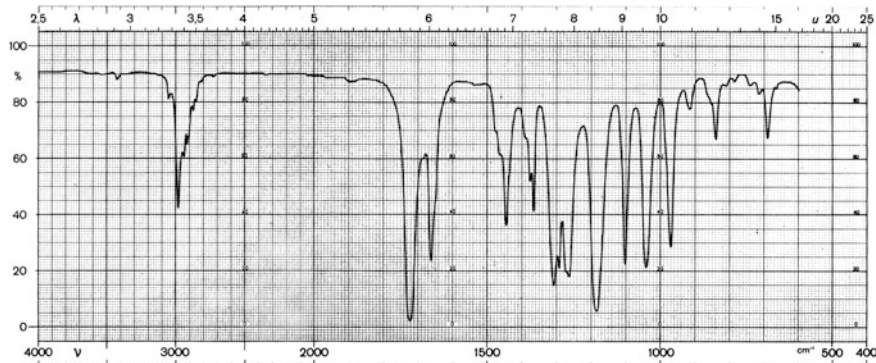


Fig. 1.2 The IR spectrum of the “unknown” $C_6H_{10}O_2$

(no absorptions in the range $3,100\text{--}4,000\text{ cm}^{-1}$ and no doublet between $3,000$ and $3,100\text{ cm}^{-1}$). However, alkenic double bond ($1,660\text{ cm}^{-1}$), carbonyl ($1,730\text{ cm}^{-1}$) and methyl ($1,380\text{ cm}^{-1}$) vibrations are present suggesting the presence of these groups and, in addition, the carbonyl group exists as an ester ($1,730\text{ cm}^{-1}$). The first step is therefore to find all isomers which meet the constraints produced by considering the sets of functional groups that are present and those that are absent. For the molecular formula $\text{C}_6\text{H}_{10}\text{O}_2$ the number of structural isomers is 4,869. This number can be calculated using any of the available structure generators [9].

If we successively impose the constraints obtained from the IR spectrum onto the full set of isomers then the following result is obtained: $4,869 \rightarrow$ (no O–H) $\rightarrow 1,719 \rightarrow$ (C=O present) $\rightarrow 384 \rightarrow$ (C=C present) $\rightarrow 151 \rightarrow$ (CH_3 present) $\rightarrow 147 \rightarrow$ (O–C=O present) $\rightarrow 26 \rightarrow$ (no $\text{C}=\text{CH}_2$) $\rightarrow 10$.

The direct usage of the trivial information regarding the absence of any OH groups reduces the number of candidate structures by a factor of almost three and after the further addition of new constraints only ten candidate structures remain (see Fig. 1.3). Note that as a result of the application of constraints deduced from the IR spectrum, 99.8 % of all conceivable isomers were rejected and the number of potential candidate structures reduces from 4,869 down to 10.

The set of structural isomers derived from the given spectral data is referred to as a *solution to the problem*. The solution is called *valid* if it contains the *genuine structure*, otherwise it is termed *invalid*.

Let N be the theoretically possible number of isomers corresponding to a given molecular formula, k —number of structures which are in agreement with the constraints imposed from the spectroscopic data and I —the amount of *structural information* extracted from the spectrum (or set of spectra). Then, as previously

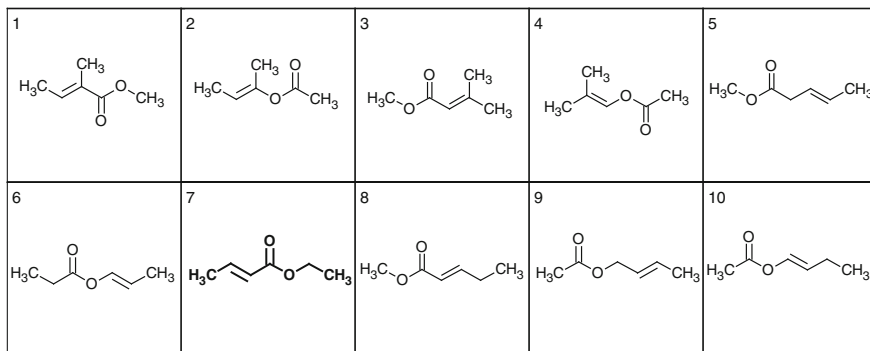
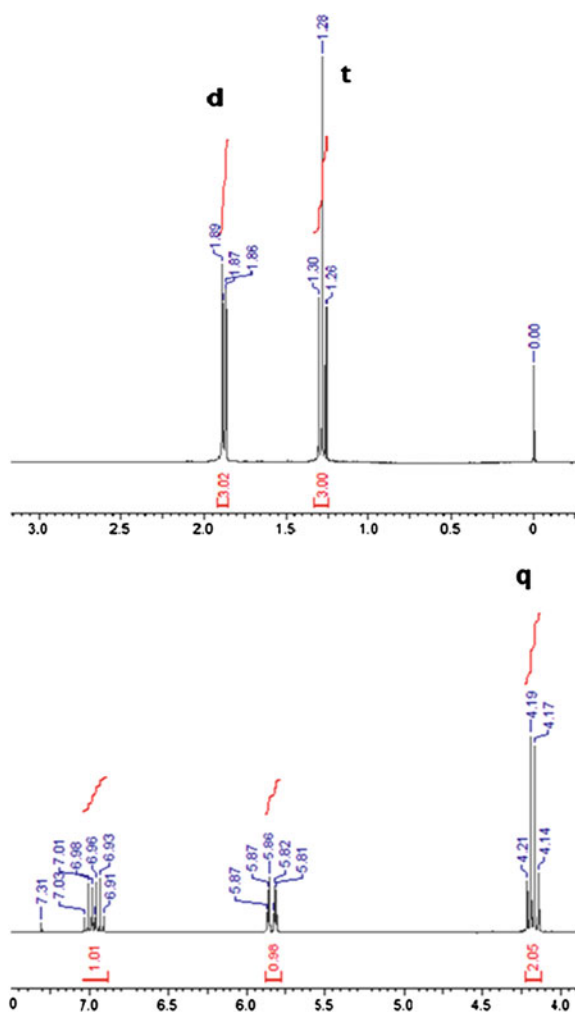


Fig. 1.3 The candidate structures remaining after the imposition of constraints extracted from the IR spectrum of the “unknown” compound $\text{C}_6\text{H}_{10}\text{O}_2$ on the 4,869 conceivable isomers corresponding to this molecular formula

shown [9] $I = \log_2 N - \log_2 k = \log_2(N/k)$ and the moiety μ of the total structural information which was extracted from a given experiment can be assessed as $\mu = (\log_2 N - \log_2 k) / \log_2 N = 1 - \log_2 k / \log_2 N$. For our example $\mu = 0.73$, which means that the IR spectrum carries three-quarters of the total structural information. Therefore, to unambiguously select a single correct structure among the ten (Fig. 1.3), it is necessary to extract almost one-quarter of the total structural information. In some sense, the structural information that has to be additionally extracted for unambiguous structure elucidation looks to be the most valuable.

The remaining ten structures are quite similar and it is not very simple to distinguish the correct structure using the characteristic features of the IR spectrum only. The ^1H NMR spectrum of the compound is shown in Fig. 1.4

Fig. 1.4 The ^1H NMR spectrum of the “unknown” $\text{C}_6\text{H}_{10}\text{O}_2$



The three distinct multiplets observed in the 1–4.2 ppm spectral area of the ^1H NMR spectrum provide unambiguous evidence for the presence of an $\text{O}-\text{CH}_2-\text{CH}_3$ fragment which is absent in all isomers except for structure #7. Hence, with this additional constraint, we can complete the structure elucidation of the unknown using the straightforward application of the general CASE (Computer-Assisted Structure Elucidation) strategy. It turned out that in this case the solution was valid. This strategy works in reality and only 0.4 s execution time for a nominal PC processor running the StrucEluc software were necessary to solve the problem (i.e., perform structure generation under the constraints obtained from the IR spectrum). However, such a direct approach is unfortunately justified only in those cases when the molecular formula of the unknown is rather constrained—when the number of heavy atoms does not exceed 15–20. The number of isomers and, consequently, the generation time of the full set of isomers grows at an almost exponential rate with the increase in the number of heavy atoms in the molecular formula and with the RDBE (Rings and Double Bonds Equivalent) value. This leads to the so-called combinatorial explosion. For instance, the molecular formula $\text{C}_{18}\text{H}_{34}\text{O}$ yields 870,280,931 isomers, most of them are stable molecules. The processor time for structure generation which was run to generate all possible isomers (i.e., without saving the structures to disk) was 1 h 25 min on a standard PC.

Figure 1.5 displays the structures associated with a series of modest sized chemical compounds and the number of conceivable calculated structural isomers [11].

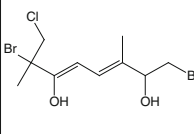
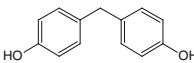
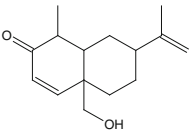
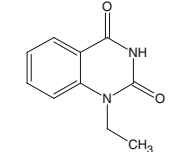
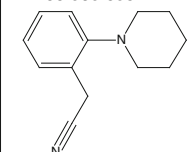
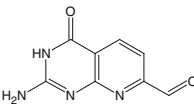
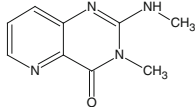
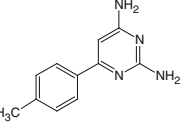
1 50 502 293 	2 104 801 402 125 	3 138 136 211 624 	4 168 958 533 562 
M=359 $\text{C}_{10}\text{H}_{17}\text{Br}_2\text{ClO}_2$	M=200 $\text{C}_{13}\text{H}_{12}\text{O}_2$	M=234 $\text{C}_{15}\text{H}_{22}\text{O}_2$	M=190 $\text{C}_{10}\text{H}_{10}\text{N}_2\text{O}_2$
5 200 939 695 277 	6 326 019 270 108 	7 364 563 117 825 	8 2 258 672 147 012 
M=200 $\text{C}_{13}\text{H}_{16}\text{N}_2$	M=190 $\text{C}_8\text{H}_6\text{N}_4\text{O}_2$	M=190 $\text{C}_9\text{H}_{10}\text{N}_4\text{O}$	M=200 $\text{C}_{11}\text{H}_{12}\text{N}_4$

Fig. 1.5 The structures of a series of small molecules and the theoretical numbers of isomers associated with the related molecular formulae [11]

The figure shows that even the simplest of structures can have hundreds of millions up to trillions of isomers. For the simple structure with the associated molecular formula of $C_{10}H_{17}Br_2ClO_2$, the number of isomers, N , exceeds 50 million and rudimentary inspection suggests that more than 40 million of these could likely exist. It should be noted that the CAS registry contains “only” 90 million known chemical compounds while 67 million are commercially available [12].

The number of heavy atoms in the molecular formula defines the *problem dimension*. There is only one way to overcome the fundamental difficulty associated with the problem dimension and that is to reduce it by providing *molecular fragments* which can encompass (“absorb”) a number of the heavy atoms and form “macro-atoms” as discrete units of the structure.

At first glance it would be attractive to utilize large fragments supplied with their characteristic spectral features. However, in this case another fundamental difficulty appears: it is impossible to create spectrum-structure correlation tables for large fragments due to the combinatorial nature of chemistry—the number of possible fragments is not countable (a “combinatorial explosion” of another origin). Therefore, to provide for the structure generation of real-world molecules some conditions have to be satisfied.

Generally, the chance to successfully perform structure generation depends on the strength and number of constraints imposed on the skeletal atoms.

The constraints can be imposed in two forms. The first form implies introduction of fragments (substructures) where atoms are connected with *chemical bonds* of definite multiplicities. The second form realizes the constraints that follow from structural interpretation of the 2D NMR spectra. *Strict fragments* detected from COSY data and *fuzzy fragments*, which contain skeletal atoms connected with *connectivities* produced from HMBC *correlations*, are used in this case (see Chap. 2). An appropriate combination of strict and fuzzy fragments usually leads to completion of the structure generation in a reasonable time. However, as we will see later (Part III), employing *only* fuzzy fragments suggested by the HMBC data allows one to perform the structure generation and most frequently elucidate structures of large organic molecules successfully. In the history of CASE the development of optimal methods to use fragments for effective structure elucidation took several decades of hard work for a number of research groups [8, 9, 13–15]. This represents how complex the challenge of developing CASE systems actually is.

1.1.2 Principle Flow Diagram of Structure Elucidator

Structure Elucidator was designed to provide the possibility to use all kinds of constraints which can anyhow be imposed on conceivable structures from 1D and 2D NMR, MS, IR/Raman, and UV spectra, as well as from any other sources of structural information (sample origin, chemical knowledge, etc.). A simplified principle flow diagram for the Structure Elucidator expert system is shown in Fig. 1.6. The NMR experimental data that are used for the structure elucidation include ^{13}C and ^1H NMR spectra, ^1H - ^{13}C HMBC (^1H - ^{15}N HMBC if available), ^1H - ^1H COSY, and any other spectra obtained using 2D NMR techniques. A high resolution mass spectrum (HRMS) is added to determine, first of all, an accurate molecular mass and the molecular formula of the unknown. The full MS data can be used during the last stage of the structure elucidation to check the most plausible candidate structures. If information about the presence or absence of some functional groups is available from the IR/Raman and UV spectra it can also be utilized.

The following main stages of the structure elucidation procedure can be distinguished.

- The number of signals, their intensities, and chemical shifts in the ^{13}C NMR spectrum as well as the signal integrals in the ^1H NMR spectrum and IR data are used to determine the minimum numbers of carbons, hydrogens, and heteroatoms in a molecule. Probable heteroatoms are suggested from the analysis of all available spectroscopic information. This information, along with an accurate molecular mass and some characteristic ions observed in the MS spectrum, allows us to suggest a conceivable chemical composition and perform generation of the most probable molecular formula or set of probable molecular formulae.
- The program processes 1D and 2D NMR spectra and performs peak picking. In so doing signal positions, multiplicities, and coupling constants (if possible) are determined. Then the ^1H and ^{13}C chemical shifts and 2D NMR correlations are

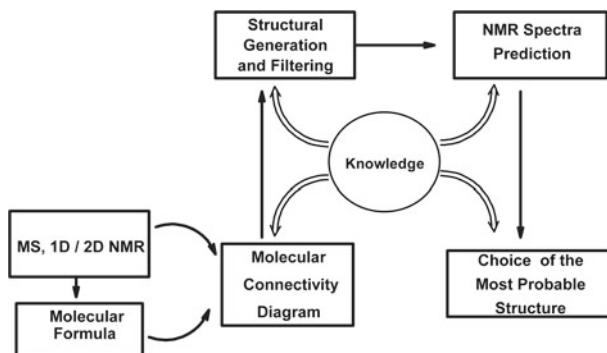


Fig. 1.6 A simplified principle flow diagram of structure elucidator expert system

automatically (or manually if necessary) transferred to data tables which are displayed on the screen.

- ^1H - ^1H COSY and ^1H - ^{13}C HMBC (^1H - ^{15}N HMBC) correlations are transformed into *connectivities* between the corresponding skeletal atoms. The molecular formula and connectivities are used to create a Molecular Connectivity Diagram (MCD) which displays the structural blocks C, CH, CH_2 , CH_3 , NH, NH_2 , OH, and heteroatoms, as well as hydrogens that can be attached to free heteroatoms or carbons whose multiplicities were not determined. Structural blocks are supplied with ^{13}C and ^1H chemical shifts and atom properties (the system assigns the properties automatically), while all connectivities are visualized (as lines or arrows) using different colors specified for the COSY and HMBC correlations as well as for marking the lengths of the connectivities. The information visually displayed in the MCD window can be easily edited by the operator using the program toolbar. This information is read by the program for the structure generation.
- Structure generation is performed automatically in accordance with the options selected by the chemist. A very wide set of options allows the user to initiate the structure generation in several modes and verify the different hypotheses by editing manually the initial information displayed graphically on the MCD. The structure generation is combined with spectral and structural filtering of the generated structures for which the system knowledge is used. Duplicates are removed from the output structural file by a smart algorithm which takes into account ^{13}C chemical shift assignments and deviation values between experimental and predicted spectra.
- To select the most probable structure NMR chemical shift prediction is performed for each generated structure using three different empirical methods of chemical shift calculation (see Sect. 1.4). The calculation uses databases included into the system knowledge. In accordance with the general CASE methodology, the structure which is characterized by the calculated NMR spectrum most similar to the experimental spectrum is declared to be the most probable. The similarity of the NMR spectra is assessed by comparison of the average deviations between the measured and predicted spectra.

As mentioned above, an expert system mimics a spectroscopist's reasoning during the structure elucidation of a new organic compound. Therefore, prior to developing an expert system it is necessary to analyze the methods of logical structure inference that are used by a human expert and try to formalize them since a computer is capable of processing only formalized information. It has been shown [5, 13, 16] that mathematical logic [17] can be used for this purpose. Consequently, the problem is to interpret ("translate") chemical reasoning in terms of symbolic logic. The interpretation of the main approaches to draw conclusions in the process of structure elucidation is important not only for computerizing the process but also to help chemists to avoid annoying logical mistakes (see examples in [18–20]). In the next section we will review the "applied" philosophy of structure elucidation from spectroscopic data.

1.2 Logical Basis of the Spectroscopic Structure Elucidation

The history of the development of CASE systems to date has convincingly demonstrated the point of view [13] that the process of molecular structure elucidation is reduced to the logical inference of 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 a long time before computer methods appeared. Independent of computer-based methods the path to a target structure is the same because CASE expert systems only mimic the approaches of a human expert. The main advantages of the CASE approach are as follows:

- All statements (“axioms”) regarding the interrelation between spectra and a structure are expressed explicitly;
- All logical consequences (structures) following from the system of “axioms” are completely deduced without any exclusions;
- The process of computer-based structure elucidation is fast and provides tremendous saving in both time and labor for the scientist;
- If the chemist has several alternative sets of axioms related to a given structural problem then an expert system allows for the rapid generation of all structures from each of the sets and identification of the most probable structure by comparison of the solutions obtained.

We assume that the reader is familiar with the basic principles of molecular spectroscopy (MS, 1D and 2D NMR, IR/Raman, UV, etc.) and has some experience in structural interpretation of spectroscopic data and the structure elucidation. In this section we will describe the main kinds of statements used during the process of structure elucidation [9, 13, 18]. These can be conventionally divided into several categories.

1.2.1 *Axioms and Hypotheses Based on Characteristic Spectral Features*

In accordance with the definition we refer to “axioms” as those statements that can be considered true based on prior experience. To elucidate the structure of a new unknown compound, the chemist usually uses spectrum-structure correlations established as a result of the efforts of several generations of spectroscopists. In so doing the chemist implies that each correlation is meaningful and *true* though experience suggests the probabilistic origin of these correlations. We will see that statements reflecting the existence of characteristic spectral features can be considered as the basic axioms of structure elucidation theory, the latter possessing all traits of *axiomatic theory* [13, 17, 21]. The general form of typical axioms belonging to this category can be formulated as follows:

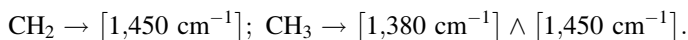
If a molecule contains a fragment A_i , then the characteristic features of fragment A_i are observed in certain spectrum ranges $[X_1], [X_2], \dots [X_m]$ which are specific for this fragment.

This axiom can be presented symbolically as a logical function

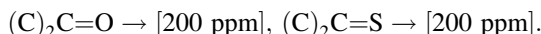
$$A_i \rightarrow [X_1] \wedge [X_2] \wedge \dots \wedge [X_m] \quad (1.1)$$

Here an arrow denotes a logical operation referred to as an *implication* which corresponds to an expression “if..., then...”. The symbol \wedge denotes the logical operation referred to as the *conjunction* which corresponds to the junction “and”.

For example, if a molecule contains a CH_2 group then a vibrational band around $1,450 \text{ cm}^{-1}$ is observed in the IR spectrum (a scissoring vibration of CH_2 group). If a molecule contains a CH_3 group then two bands around $1,450$ and $1,380 \text{ cm}^{-1}$ appear (antisymmetric and symmetric bending vibrations of a CH_3 group). According to (1.1) these axioms can be presented formally in the following way:



Analogously, for characteristic ^{13}C NMR chemical shifts the following implications are also exemplars of axioms:



When characteristic spectral features are used for the detection of fragments that can be present in a molecule under investigation, then the chemist usually forms statements for which a typical “template” is as follows:

If a spectral feature X_j is observed in a spectrum range $[X_j]$ then the molecule contains at least one fragment of the set $A_i(X_j), A_k(X_j), \dots A_l(X_j)$, where $A_i, A_k, \dots A_l$ are fragments for which the spectral feature X_j observed in the range $[X_j]$ is characteristic. The fragments $A_i, A_k, \dots A_l$ form a finite set A .

This assertion is symbolized as

$$[X_j] \rightarrow A_i(X_j) \vee A_k(X_j) \vee \dots \vee A_l(X_j) \quad (1.2)$$

The symbol \vee denotes the logical operation referred to as *disjunction* which corresponds to the junction “or”. The expression (1.2) means that the presence of the feature X_j can be accounted for by any combination of fragments belonging to the set A .

It should be pointed out that this statement is a *hypothesis*, not an axiom, because: (i) the feature X_j can be produced by some fragment which is not known as yet, (ii) the feature X_j can appear due to some intramolecular interaction of known fragments. Therefore, if an absorption band is observed at $1,450\text{ cm}^{-1}$ in an IR spectrum then the molecule can contain either CH_2 or CH_3 groups, or both of them (band overlap at $1,450\text{ cm}^{-1}$ is allowed), or the $1,450\text{ cm}^{-1}$ band can appear as a result of the presence of another unrelated functional group or substructure. This statement can be expressed formally using the symbol for logical *disjunction*: $1,450\text{ cm}^{-1} \rightarrow \text{CH}_2 \vee \text{CH}_3 \vee \alpha$, where α is a “sham fragment” denoting an unknown cause of the feature origin. For our ^{13}C NMR examples we may obviously formulate the following hypothesis: $200\text{ ppm} \rightarrow (\text{C})_2\text{C}=\text{OV}(\text{C})_2\text{C}=\text{S}$. It is important to have in mind that if $A_i \rightarrow X_j$ is *true*, then in accordance with formal logic the inverse implication $X_j \rightarrow A_i$ can be true or not true. In other words:

The presence of a signal $X_j \subset [X_j]$ belonging to the range of a fragment characteristic feature in a spectrum does not still imply the presence of a corresponding fragment $A_i (X_j)$.

According to symbolic logic rules [17] a *true* implication equivalent to expression $A_i \rightarrow X_j$ is

$$\overline{X_j} \rightarrow \overline{A_i}$$

This implication means that *if the characteristic spectral feature $X_j \subset [X_j]$ does not occur in the $[X_j]$ range of a spectrum, then the corresponding fragment A_i is absent from the molecule under investigation.*

The latter statement can be considered as another equivalent formulation of the basic axiom $A_i \rightarrow [X_j]$ and it is in good agreement with chemical common sense. For instance, if no IR absorption band occurs in the range of $3,700\text{--}3,000\text{ cm}^{-1}$ then an alcohol group is absent from the molecule.

During the process of structure elucidation, to make a conclusion about the presence or absence of some fragment its characteristic features in all available spectra are taken into account.

In the general case, it can be written as

$$A_i \rightarrow w_1(i) \wedge w_2(i) \wedge \cdots \wedge w_n(i) \quad (1.3a)$$

$$\{\bar{w}_1(i) \vee \bar{w}_2(i) \vee \cdots \vee \bar{w}_n(i)\} \rightarrow \bar{A}_i \quad (1.3b)$$

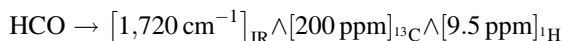
Here the variables w_j in the relations 1.3a and 1.3b denote the characteristic features of the fragment A_i assigned in all available spectra of different kinds (for instance, IR, ^1H NMR and ^{13}C NMR).

Equation (1.3b) is equivalent to Eq. (1.3a) and is obtained from (1.3a) by application of the de Morgan laws [9, 17] to conjunction $w_1(i) \wedge w_2(i) \wedge \cdots \wedge w_n(i)$. An interpretation of Eq. (1.3b) can be described as follows:

If at least one characteristic feature w_1, w_2, \dots, w_n of the fragment A_i is not observed in at least one of the available experimental spectra, then the fragment A_i is absent from the structure of unknown.

It is necessary to underline that the “template” presented by Eq. (1.3b) allows chemists to obtain the most reliable conclusions during the structural analysis of an unknown compound. The reason is that we use an *axiom* (1.3a) in the equivalent form (1.3b) for rejecting the presence of a fragment, while, as mentioned, the inverse implication $X_j \rightarrow A_i$ is only a hypothesis and it can be true or not true. In agreement with the philosophy of science, the truthfulness of a hypothesis is verified in practice. In our case the truthfulness of a suggested hypothesis is determined when the genuine structure is established.

For instance, the following characteristic spectral ranges are specific for the aldehyde group: IR [1,720 cm^{-1}], ^{13}C NMR [200 ppm] and ^1H NMR [9.5 ppm]. Then the corresponding logical relation is



and the equivalent representation of the axiom can be written as

$$[\overline{1,720 \text{ cm}^{-1}}]_{\text{IR}} \vee [\overline{200 \text{ ppm}}]_{^{13}\text{C}} \vee [\overline{9.5 \text{ ppm}}]_{^1\text{H}} \rightarrow \overline{\text{HCO}}$$

It is evident that the absence of any IR band in the range [1,720 cm^{-1}] or the absence of any resonances in the corresponding intervals of the ^{13}C or ^1H NMR spectra leads to the conclusion that the molecule under investigation cannot contain the aldehyde group. It is worth noting that in some specific cases Eq. (1.3b) may be violated. For instance, the ^{13}C NMR signals of quaternary carbon atoms are sometimes of such low intensity that they cannot be distinguished against the level of noise. The human expert analysis of 1D and 2D NMR spectra can help to resolve the contradiction between the molecular formula and the number of signals in a ^{13}C NMR spectrum.

When a newly synthesized or isolated compound has to be identified the first stage is commonly to characterize an unknown by the structural groups that may exist in the molecule. The investigation of an expert’s approach to reasoning leads

to the conclusion that the structural-group spectral analysis, SGA, (both manual and computerized), is based on the following assumptions in addition to the ones mentioned above:

- Each fragment $A_i \in \mathbf{A}$ has definite ranges for its characteristic spectral features if the fragment's *permissible environment* in a molecule is strictly defined.
- The analyzed molecule must be *additive* or at least contain some closed self-contained groups. *Additive molecules produce additive spectra.*
- Only $A_i \in \mathbf{A}$ fragments can be involved in the solution of the problem. All other conceivable fragments, $A_x \notin \mathbf{A}$, will not be considered, but they may be involved in the solution during structure generation.
- While solving the problem *all spectral features* (chemical shifts, frequencies) of the experimental spectra are considered as potential particular attributes of the structural elements, $A_i \in \mathbf{A}$, and may be interpreted only in terms of the A_i fragments.
- The chemical composition and degree of unsaturation, U , for the fragments selected during the SGA must be consistent with the molecular empirical formula.
- While solving the problem every possible fragment set resulting from the structural group analysis must provide an interpretation of all spectral features perceived as characteristic in the experimental spectra.

All these assumptions are used (in the explicit or implicit form) by experts in the process of SGA. The expert system, based on the mathematical model of structural group analysis, uses them only in the explicit form because these “axioms” are put in the knowledge and algorithms of the system.

Consequently, all those “real world” contradictions which the chemist may encounter in the process of molecular structure elucidation are typical of the expert system as well.

In this section we showed that structural group analysis of organic molecules using characteristic spectral features (chemical shifts in NMR, frequencies in IR) could be formalized on the basis of symbolic logic. It turned out that the interrelations between definite substructures and their characteristic features were described by logical functions. It has been shown [9] that as a result of joint logical analysis of these interrelations with an experimental spectrum the algorithm produces all combinations of substructures that can be present in the structure of an unknown. We would like to note that application of logical calculus for solving a spectrum-structure problem is not a mathematical technique only, but an aid for more in-depth understanding the way of expert reasoning during the structure elucidation of an unknown compound. Being implemented in an expert system, the algorithms of structural group analysis may reveal contradictions in the set of

“axioms”. To overcome the contradictions the system needs the assistance of a chemist who has diverse aids to do it, working in the interactive mode.

1.2.2 Axioms and Hypotheses of 2D NMR Spectroscopy

2D NMR spectroscopy is a method which, in principle, allows a molecular structure to be inferred from the available spectral data ab initio without using any spectrum-structure correlations and additional suppositions. In some cases the 2D NMR data provide sufficient structural information to suggest a manageable set of plausible structures. This is a fairly common situation for small molecules with a lot of protons contained within the molecule. In practice, the structure elucidation of large molecules by the ab initio application of 2D NMR data only (without usage of any axioms based on spectrum-structure correlations) is generally impossible. The 1D and 2D NMR data are usually combined synergistically to obtain solutions to real analytical problems in the study of organic compounds.

Experience has shown that the size of a molecule is not a crucial obstacle for a CASE system based on 2D NMR data. The number of hydrogen atoms responsible for the propagation of structural information across the molecular skeleton and the number of skeletal heteroatoms are the most influential factors. An abundance of hydrogen atoms and a small number of heteroatoms generally eases the structure elucidation process rather markedly.

There exists the so-called Screws' rule which says: a structural problem becomes a challenge if the ratio $n(X)/n(H) > 2$, where $n(X)$ and $n(H)$ are numbers of skeletal and hydrogen atoms in the molecule correspondingly.

Any specific dependence between molecular composition and the number of plausible structures deduced by an expert system was not established so far, because the different modes for solving a problem are chosen according to the nature of the specific problem (see Sects. 1.2.2 and 1.2.3). Moreover, the complexity of the problem is associated with many factors that cannot be identified before attempts are made to solve the problem. For instance, the complexity of the problem depends on whether the heavy atoms and their attached hydrogen atoms are distributed “evenly” around the molecular skeleton. If at least one “silent” fragment (i.e., having no attached hydrogens) is present in a molecule then it can interrupt a chain of HMBC and COSY correlations. As a result the number of structural hypotheses will increase dramatically as reported, for example, in the cryptolepine family [22, 23].

When 2D NMR data are used to elucidate a molecular structure then the chemist (or an expert system which mimics the manner of a chemist's reasoning) deduces conceivable structures from the molecular formula and a set of hypotheses matching

the data from two-dimensional NMR spectroscopy. When we deal with a *new* organic compound we must interpret a *new* 2D NMR spectrum or spectra. In this case we have no possibility to rely on “axioms” valid for the given spectrum-structure matrix, so hypotheses that are considered as the most plausible are formed. These hypotheses are based on the general regularities, which are the significant *axioms of 2D NMR spectroscopy*.

Prior to considering these axioms it would be appropriate to briefly discuss the nature of CASE as a so-called inverse problem. When we elucidate the structure of a new compound using 1D and 2D NMR spectra, we eventually establish all chemical bonds (and their multiplicities) which connect the atoms in the molecule analyzed. A problem of this kind is a typical *inverse problem*. Inverse problems are met rather frequently in different sciences; their main properties are discussed in series of monographs, for instance, [24].

The most important peculiarity of inverse problems is that they usually have ambiguous solutions. A single solution is selected by utilizing additional information (imposing constraints).

Analysis of a general CASE strategy based on 2D NMR techniques (HSQC, HMBC, COSY, etc.) leads to the conclusion that the entire problem can be presented as a *combination of two inverse problems* which should be solved in consecutive order.

The *first problem* can be formulated in the following manner: determine *all* (if possible) *pairs of atoms (nuclei)* in the molecule for which there exist spin couplings observed in the NMR spectra.

This goal is achieved by peak picking in the 1D and 2D NMR spectra and enumeration of all pairs of interacting nuclei using rules for the spectrum-structure interpretation of 2D NMR spectra.

The *second problem* is to generate *all structures* that meet the results of solving the *first problem* (structural constraints) and in addition satisfy known interrelations between a chemical structure and its characteristic spectral features (correlation tables), as well as in agreement with chemical knowledge.

A single structure (a single solution to the problem) is selected, as discussed above, by successively imposing different new constraints on the generated structures.

We see that solution of the second (main) inverse problem strongly depends on the solution of the first problem. If erroneous spin couplings (correlations) were

included into the solution of the first problem the possibility to obtain a correct solution to the second problem becomes problematic. To deeply understand the nature of CASE problems it is important to master the art of structure elucidation.

Now we will attempt to express the axioms of 2D NMR spectroscopy in an explicit form and classify them. There are of course various forms of 2D NMR spectroscopy, the most important and common of these being homonuclear ^1H - ^1H and heteronuclear ^1H - ^{13}C spectroscopy. Although heteronuclear interactions of the nature X_1 - X_2 (X_1 and X_2 are magnetically active nuclei but not ^1H or ^{13}C) are possible such spectra are rare and, except for labeled materials, very difficult to acquire in general.

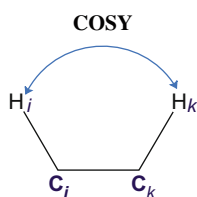
A *necessary* condition for the application of 2D NMR data to computer-assisted structure elucidation is the chemical shift assignment of all proton-bearing carbon nuclei, (i.e., all CH_n groups where $n = 1-3$). This information is extracted from the HSQC (alternatively HMQC) data using the following axiom:

If a peak (δC_i , δH_i) is observed in the HSQC (HMQC) spectrum then the hydrogen atom H_i with chemical shift δH_i is attached to the carbon atom C_i having chemical shift δC_i .

Using notations of symbolic logic we can present this axiom as follows:

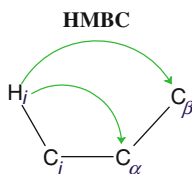
$$\text{peak}(\text{C}_i, \text{H}_i) \rightarrow \text{bond}\{(\text{C}_i)-(\text{H}_i)\}$$

The main sources of structural information are COSY and HMBC correlations which allow for the elucidation of the backbone of a molecule. We refer to *standard* correlations as those that satisfy the following axioms reflecting the experience of NMR spectroscopists:



If a peak (δH_i , δH_k) is observed in a COSY spectrum, then a molecule contains the chemical bond $(\text{C}_i)-(\text{C}_k)$, which is symbolized by the implication

$$\text{peak}(\delta\text{H}_i, \delta\text{H}_k) \rightarrow \text{bond}\{(\text{C}_i)-(\text{C}_k)\}$$



If a peak $(\delta H_i, \delta C_k)$ is observed in an HMBC spectrum, then the atoms C_i and C_k ($k = \alpha, \beta$) are separated in the structure by one or two chemical bonds:

$(C_i)-(C_\alpha)$ or $(C_i)-(C_\alpha)-(C_\beta)$, which is symbolized by the implication

$$\text{peak}(\delta H_i, \delta C_k) \rightarrow \text{bond}\{(C_i)-(C_\alpha)\} \vee \text{bonds}\{(C_i)-(C_\alpha)-(C_\beta)\}$$

In a general case, the carbon C_α can be replaced by any heteroatom X .

The standard COSY and HMBC correlations are characterized by coupling constants ${}^3J_{\text{HH}}$ (COSY) and ${}^{2-3}J_{\text{CH}}$ (HMBC) so the maximum topological distance between interacting nuclei is three bonds in standard correlations.

By analogy, the main axiom associated with employing the NOE effect (NOESY/ROESY spectra) for the purpose of structure elucidation can be formulated in the following manner:

If a peak $(\delta H_i, \delta H_k)$ is observed in a NOESY (ROESY) spectrum, then the distance between the atoms H_i and H_k through space is less than 5 \AA , which is symbolized by the implication $\text{peak}(\delta H_i, \delta H_k) \rightarrow \{\text{dist} [(H_i, H_k)] < 5 \text{ \AA}\}$

It is important to note that there is a principal difference between the logical interpretations of 1D and 2D NMR axioms. For instance, for COSY there exists the second equivalent form of the main axiom:

If a molecule does not contain the chemical bond $(C_i)-(C_k)$, then no peak $(\delta H_i, \delta H_k)$ is observed in a COSY spectrum:

$$\overline{\text{bond}\{(C_i)-(C_k)\}} \rightarrow \overline{\text{peak}(\delta H_i, \delta H_k)}$$

Note that according to the property of logical implication, the logical expression

$$\overline{\text{peak}}(\delta H_i, \delta H_k) \rightarrow \overline{\text{bond}}\{(C_i)-(C_k)\}$$

produced from the true COSY axiom can have any value of truthfulness. In this case, the interpretation of logical implication allows us to conclude that the absence of a peak $\delta(H_i, H_k)$ in the COSY spectrum says *nothing* about the existence of the chemical bond $(C_i)-(C_k)$ in the molecule: the bond may exist or may not exist. There are many examples where an expected COSY peak is not observed in the spectrum due to some physical reasons (see Figs. 1.7 and 1.8 in Sect. 1.2.4). Consequently, the expert system does not use the absence of some 2D NMR peak $\delta(H_i, H_k)$ in order to reject structures containing the bond $(C_i)-(C_k)$. Analogous conclusions are also applicable to HMBC and NOESY spectra. Note that formal conclusions following from the symbolic presentation of the main 2D NMR axioms fully correspond to the conclusions, which are usually made by experts during the structural interpretation of 2D NMR data.

While it is known that the listed axioms hold in the overwhelming majority of cases, there are many exceptions and these correlations are referred to as *non-standard correlations* (NSCs) [25, 26].

The NSCs are those for which the topological distance between interacting nuclei exceeds three bonds, i.e., $n > 3$ for coupling constants ${}^nJ_{\text{HH}}$ and ${}^nJ_{\text{CH}}$.

Since standard and NSCs are not easily distinguished the existence of NSCs is the main hurdle to logically inferring the molecular structure from the 2D NMR data. If the 2D NMR data contains both undistinguishable *standard* and *nonstandard* correlations then the total set of “axioms” derived from the 2D NMR data will contain logical *contradictions*.

This means that the correct structure cannot be inferred from these axioms and in this case the structural problem either has no solution or the solution will be incorrect: the set of suggested structures will not contain the genuine structure (in the best case a correct structure with erroneous chemical shift assignment will be generated). Numerous examples of such situations will be considered in Chap. 5 where we will show how the problem of the presence of contradictions in the initial data is resolved within Structure Elucidator.

Unfortunately, as yet there are no routine and reliable NMR techniques which distinguish between 2D NMR signals belonging to standard and nonstandard correlations. Nevertheless, the application of time-consuming INADEQUATE and different versions of ADEQUATE experiments helps to overcome the presence of contradictions but these techniques are also based on their own axioms which can be violated.

1.2.3 *Structural Hypotheses Necessary for the Assembly of Structures*

When spectral peaks in 1D and 2D NMR spectra are assigned and all 2D correlations are transformed into *connectivities* between heavy atoms in the skeletal framework, then feasible molecular structures should be assembled from *strict fragments* (suggested on the basis of the 1D NMR, 2D COSY, and IR spectra, as well as those postulated by the researcher) and *fuzzy fragments* determined from the 2D HMBC data. To assemble the structures it is necessary to make a series of responsible decisions, equivalent to constructing a set of axiomatic hypotheses. At least the following choices should be made:

- Allowable *chemical composition(s)*: CH, CHO, CHNO, CHNOS, CHNOCl, etc. The choice is made on the basis of NMR and optical spectra, the pattern of the molecular ion cluster, characteristic m/z values in MS, chemical considerations and rules (for instance, the so-called nitrogen rule [10]), and other information that may be available (e.g., sample origin, etc.). It is evident that if the potential possibility of some chemical element that really exists in an unknown will be declined by the researcher the genuine molecule will never be identified (see examples in [18] and Sect. 5.13)
- The *possible molecular formula* (formulae) is selected from a set of possible accurate molecular masses. The suggestion of a correct molecular formula is crucial for CASE systems and is highly desirable in order to perform dereplication.
- The *possible valences* of each atom having variable valence: N(3 or 5), S(2 or 4 or 6), P(3 or 5). If ^{15}N and ^{31}P NMR spectra are not available then, in principle, all admissible valences of these atoms should be tried. Obviously it is practically impossible to perform such a complete search manually. The application of a CASE system allows, in principle, the verification of all conceivable valence combinations.
- The *possible hybridization* of each carbon atom: sp ; sp^2 ; sp^3 ; *not defined*. The choice is made on the basis of characteristic chemical shifts in ^{13}C and ^1H NMR spectra, the chemical composition and degree of unsaturation of the molecule.
- The *possible neighborhoods* for each carbon atom with heteroatoms: *forbidden*, *obligatory*, *not defined*. An example of a typical challenge: does C($\delta = 103$ ppm) indicate a carbon in the sp^2 hybridization state or in the sp^3 hybridization state but connected with two oxygens by ordinary bonds?
- The *possible number of hydrogen* atoms attached to carbons that are the nearest neighbors to a given carbon (determined, if possible, from the signal multiplicity in the first-order ^1H NMR spectrum). This decision may be rather risky and therefore such constraints should be used only with great caution and in those cases where no signal overlap occurs and signal multiplicity can be reliably determined as, for instance, in the case of methyl group resonances showing distinctive singlets, doublets, or triplets.

- The *maximum allowed bond multiplicity*: 1 or 2 or 3. The main challenges are related to a carbon neighboring with heteroatoms (for instance, C(sp^2) or O–C(sp^3)–O), as well as to the presence or absence of a triple bond. Strictly speaking, the latter can be solved reliably only based on combined utilizing IR and Raman spectra: an IR spectrum enables a decision regarding the presence or absence of terminal $\text{–C}\equiv\text{CH}$ and $\text{–C}\equiv\text{N}$ groups, while a central $\text{–C}\equiv\text{C–}$ group is usually revealed as a strong band in a Raman spectrum and this band may be absent from the IR spectrum.
- The list of *fragments that can be assumed to be present* in a molecule according to chemical considerations or based on a fragment search using the ^{13}C NMR spectrum, the fragment database being used for this goal. The chemical considerations usually arise from careful analysis of the NMR spectra related to known compounds, particularly to natural products, that have the same origin and similar spectra. The presence of the most significant functional groups (C=O, OH, NH, NH_2 , C \equiv N, C \equiv C, C \equiv CH, etc.) can be suggested from both IR and Raman spectra taking into account the NMR data and molecular formula of the unknown. If the unknown is a product of a chemical reaction then the presence of some fragments can be postulated from the chemical knowledge.
- The list of *fragments which are forbidden* within the given structural problem. These include fragments unlikely in organic chemistry: for example, a triple bond in small cycles or an O–O–O connectivity, etc. Additionally, highly strained substructures which are rarely met in chemistry. NMR, IR, and Raman spectra can also hint at the specification of forbidden fragments, and the axiom $\bar{X}_j \rightarrow \bar{A}_i$ is usually a rather reliable basis for making a particular decision. For example, if no characteristic absorption bands of benzene rings are observed in the IR, ^{13}C , and ^1H spectrum, then this group may be included into the list of forbidden fragments. Structural constraints which can be obtained very simply frequently lead to the rejection of a huge number of conceivable structures. It should be evident that at least one poor decision based on the points listed above would likely lead to a failure to elucidate the correct structure.

1.2.4 Properties of Information Used for the Structure Elucidation

If we generalize all axioms and hypotheses forming the partial axiomatic theory of a given molecule structure elucidation then we will arrive at the following properties of initial information which should be logically analyzed:

- Information is *fuzzy by nature*, i.e., there are either 2 or 3 bonds between pairs of H_i and C_k atoms associated with a two-dimensional peak (δH_i , δC_k) in the HMBC spectrum.

- Not all theoretically possible correlations are observed in the 2D NMR spectra, i.e., information is *incomplete*.
- The presence of NSCs frequently results in *contradictory* information.
- The number of NSCs and the true lengths of each of them (four, five or six bonds?) are *unknown*. Signal overlap in NMR spectra leads to the appearance of *ambiguous* correlations. Therefore the information is otherwise *uncertain*.
- Information can be *false* if a mistaken hypothesis is suggested by the researcher.
- Information contained within the “structural axioms” reflects the opinion of the researcher and the information is, therefore, *subjective*, and typically based on biosynthetic arguments.

The incompleteness of the 2D NMR data is illustrated by Figs. 1.7 and 1.8 [27]. The figures show the relationship between the number of observed and theoretically possible correlations in both the COSY and HMBC spectra of natural products which were used as problems for challenging the StrucEluc system. The red line corresponds to a ratio of one.

Taking into consideration the information properties mentioned above we can understand why the human expert is frequently unable to search *all plausible structural hypotheses*. Therefore, it is not surprising that different researchers arrive at different structures from the same experimental data and as a result, articles revising previously reported chemical structures are quite common [18–20]. Considering the potential errors that can combine in the decision-making process associated with structure elucidation it is actually quite surprising that chemists are so capable of processing such intricate levels of spectrum-structure information and successfully deducing very complex structures at all. To assist the chemist to logically process the initial information a computer program that would be capable of systematically generating and verifying all possible structural hypotheses from

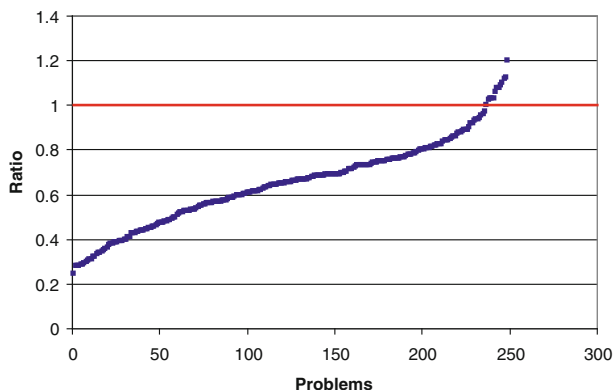


Fig. 1.7 The ratio of the number of COSY correlations versus the number of theoretically possible correlations across the 120 natural products used as problems to challenge the StrucEluc system. The problems are ordered in ascending order of the ratio. The values exceeding 1 are accounted for the presence of NSCs in COSY data [26]

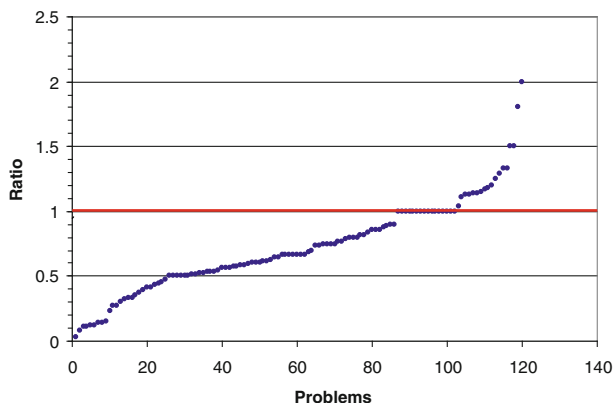


Fig. 1.8 The ratio of the number of experimental HMBC correlations versus the number of theoretically possible correlations across the 250 natural products used as problems to challenge the StrucEluc system. The problems are ordered in ascending order of the ratio. The values exceeding 1 are accounted for the presence of NSCs in HMBC data [26]

ambiguous information would be of value. The expert system Structure Elucidator (StrucEluc) comprises a software program and series of algorithms which were specifically developed to process *fuzzy, contradictory, incomplete, uncertain, subjective*, and even *false* spectrum-structure information. We will see in later chapters that the Structure Elucidator is not a robot, but a powerful amplifier of human intelligence, which works like an engine to infer all logical corollaries (structures) from the NMR spectroscopic data and from a set of axioms specific for the given problem.

In the next sections we will describe the knowledgebase of this system and its spectrum prediction aids and explain the methodology of its application. We believe that the awareness of readers with these materials will allow them not only to understand the manifold examples of the program application to the structure elucidation of complex natural products but will enable them to solve these problems on their own, using the program version adjusted for this goal. We will start with a description of the system knowledgebase.

1.3 The Knowledge Base of the Structure Elucidator

If the unknown under study is not new and its structure and spectra are already present in existing databases, then the identification of the chemical compound can be successfully performed by searching the spectra against the database. Even in those cases when the unknown is absent from the library of reference compounds a spectral search can yield a set of structures similar to that under analysis according

to the common spectroscopic principle “similar structures produce similar spectra.” Note that the inverse statement can either be true or not true depending on the case considered. The spectral search is also helpful for dereplication which answers the first question that arises when a sample is isolated: Is the compound new? Databases described in the literature are usually utilized for the goals outlined and are used as standalone resources. We will describe the ACD/NMR database in detail because it is an integral part of the CASE system Structure Elucidator.

As mentioned above (Sect. 1.2.2) structure elucidation using 2D NMR data is, in principle, is possible with an *ab initio* approach. In practice, this method fails without the application of axiomatic knowledge including different spectrum-structure correlations and chemical rules. Molecular structure elucidation is a complex logical-combinatorial process including the treatment of experimental data combined with both *factual* and *axiomatic* knowledge. Therefore, a high-performance CASE expert system must rely on a knowledge base (KB) that contains both *factual* and *axiomatic* knowledge.

In this chapter we will consider the main features of the knowledge base implemented into ACD/Structure Elucidator and explain how it is applied.

1.3.1 Structure and Content of Factual Knowledge ACD/NMR Database

The Factual Knowledge of the StrucEluc system contains the following set of databases:

- Database I contains more than 290,000 chemical structures with their ^{13}C and ^1H NMR chemical shifts assigned to the corresponding carbon and hydrogen atoms.
- Database II contains 355,000 structures supplied with ^{13}C and ^1H NMR chemical shifts assigned to their corresponding carbon and hydrogen atoms. The database is adjusted to support ^{13}C and ^1H chemical shift prediction using a HOSE code-based algorithm. The chemical shift prediction from Database II also allows for the calculation of coupling constants J_{HH} and graphical simulation of an NMR spectrum.
- Databases containing structures supplied with ^{15}N (23,000), ^{19}F (42,000), and ^{31}P NMR (>36,000) chemical shifts assigned to nitrogen, fluorine, and phosphorus atoms correspondingly. The databases include data for ^{15}N , ^{19}F , and ^{31}P NMR chemical shift predictions calculated using HOSE code-based, incremental (PLS), and neural networks algorithms.
- A Fragment Library containing 2,375,000 fragments supplied with their ^{13}C and ^1H NMR subspectra with chemical shifts assigned to the corresponding carbon and hydrogen atoms.

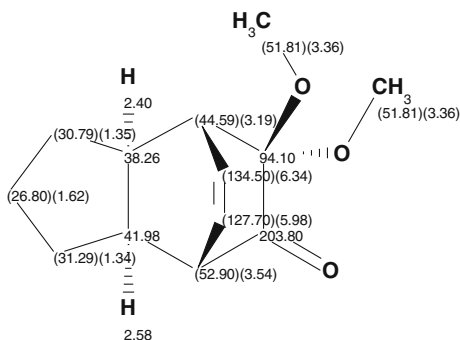
Databases I and II are repositories of factual information and can be used independently for either structural or spectral (^{13}C and ^1H NMR spectra) searching. The database contains data from various sources including natural products and compounds produced by organic synthesis. All data were taken from scientific articles and they encompass practically all compounds whose NMR spectra were published in the leading chemical journals since 1990. This work continues iteratively and the number of reference structures constantly increases. To provide high reliability of the data all information related to a structure-spectrum pair is carefully checked using both NMR spectrum prediction and human expert inspection. If any contradictions between a structure and its spectrum are revealed the structure is declared as questionable and is not included into the DB. The number of structures stored in the DB is small in comparison with just the total number of registered chemical substances, about 90 million, based on the recent announcement by the Chemical Abstracts Service (i.e., less than 1 % of this total number and this is only *registered* substances). However, practically all basic classes of natural products as well as many artificially synthesized organic molecule classes are presented in the database.

The following information is given for each compound:

- Structural formula, molecular mass, solvent, and references to the articles from which the data were obtained.
- ^{13}C chemical shift ranges specified for each group of equivalent carbon atoms existing in a molecule. For a given compound, a definite chemical shift associated with the group of equivalent carbon atoms is shown. Spectra of the same compound registered under different experimental conditions (kind of spectrometer, solvent used, etc.) are usually not fully coincident, and this is the reason why spectral intervals are used. The chemical shift is also characterized by the intensity of the corresponding signal which is quantified by the number of equivalent carbon atoms.
- ^1H chemical shift ranges are specified for each group of equivalent hydrogen atoms as well as the number of hydrogen atoms assigned to each chemical shift. Database II also contains coupling constant values associated with the ^1H chemical shifts.
- Indexes are generated regarding the chemical composition and chemical shift values in the ^{13}C and ^1H spectra of a compound. These are used to accelerate searching of structures using spectral data.

An example of a Database I structure with the assigned ^{13}C chemical shifts and accompanying information is shown in Fig. 1.9. Note that the stereochemistry of the molecule is represented by the accepted up/down convention of wedge and hashed-wedge bonds.

Fig. 1.9 An example of a structure presented in the ACD/NMR DB with accompanying information. ^{13}C and ^1H chemical shifts, as well as stereobonds are displayed (if a molecule contains stereocenters)



Ref 01 Org. Biomol. Chem./2006/4/-/2267
 Solvent 01 Chloroform-D
 In Internal DB 1
 Formula C13H18O3
 FW 222.2802
 Nominal Mass 222
 Monoisotopic Mass 222.125594
 M+ 222.125046
 M- 222.126143
 [M+H]⁺ 223.132871
 [M+H]⁻ 221.118318
 [M-H]⁺ 221.117221
 [M-H]⁻ 221.118318

1.3.2 Structure Searching Using a ^{13}C NMR Spectrum

The structure search algorithm to allow searching of the database was developed to take into account the properties of an experimental ^{13}C NMR spectrum as much as possible. The signals in a ^{13}C NMR spectrum are generally narrow and well-resolved peaks. An assumption can be made that each signal in the spectrum corresponds to one group of equivalent carbon atoms. On forming a search request proper weighting is given to the following peculiarities of carbon NMR spectra:

- The values of chemical shifts can change based on the conditions under which the spectra were acquired (temperature, solvent, etc.). A chemical shift tolerance is therefore postulated.
- Relative integral peak intensities and peak heights do not always correspond to the number of related carbon atoms. This causes difficulties in determining the number of carbon atoms assigned to a given signal. To circumvent this problem the possibility to indicate the number of potential nuclei was provided.
- Spectra can contain signals produced by impurities and different artifacts. At the same time signals from some carbon atoms may not appear in the spectrum, a condition that is rather common for quaternary carbons in low sensitivity spectra. To account for these issues, a lack of or excess number of signals compared with the number of carbon atoms in the molecular formula can be defined in the search results.

- If the information about the number of hydrogen atoms attached to a given carbon atom is available from the experimental data then it is also defined in the search results.

A structure search can be carried out under the following additional constraints:

- (a) Possible elemental compositions of the unknown compound, for example, $C_{20-22} H_{42-46} O_{1-2}$. (b) A tolerance in the molecular mass value.

1.3.3 Composition of Axiomatic Knowledge

1.3.3.1 Fragment Libraries

The axiomatic knowledge of the system is formed from a series of fragment libraries, correlation tables, algorithms for the generation of spectral constraints based on different levels of spectrum prediction, as well as definite chemical rules. This knowledge can be systematized in the following way:

- A fragment library containing substructures (mainly functional groups) accompanied by the appropriate characteristic spectral feature ranges in ^{13}C , 1H , and IR (optionally) spectra. The library is used mainly for filtering the output structural files of StrucEluc (we will call it a **Filter Library**).
- The Atom Property Correlation Table (**APCT**) containing substructures and their characteristic spectral ranges in ^{13}C and 1H NMR spectra. **APCT** is used for the assignment of carbon atom properties—atom hybridization state and the possibility of neighboring with a heteroatom. This table is also used during structure generation.
- A library containing the most typical functional groups in organic chemistry. The library is used for preliminary structural-group analysis and for building a “generalized portrait” of the unknown compound.
- A library of fragments that are considered as unlikely in organic chemistry under most conditions.
- Common sense chemical rules used for imposing constraints on the structures generated by the expert system (Bredt’s rule and some geometrical constraints).

Spectral constraints imposed by 1H , ^{13}C , ^{15}N , ^{19}F , and ^{31}P NMR chemical shift prediction using HOSE code-, neural nets-, and incremental-based approaches have an axiomatic nature and can be included into the axiomatic knowledge of the system.

The Filter Library comprises a set of molecular fragment libraries ordered in a hierarchical manner. Every fragment is accompanied by intervals of characteristic feature variations in ^{13}C NMR, 1H NMR, and IR spectra (optionally). In the NMR spectrum the chemical shift intervals, multiplicities, coupling constants, and integrals are used as characteristic attributes of a fragment. It should be noted that IR data have limited application as in complex organic molecules such as natural

products characteristic vibration frequencies often go outside of the common ranges. The characteristic spectral features of the fragments were taken from various spectroscopic sources and then checked carefully using Database I of StrucEluc.

The libraries (there is no limitation on their number) can be classified into the two following categories: libraries containing well-known chemical functionalities and those which consist of substructures chosen depending on common approaches to the structural interpretation of spectra of different types.

1.3.3.2 Universal Libraries

1. *Library of principal functionalities*

The library of principal functionalities contains the most important functional groups such as $>C=C<$, $>C=O$, $N-H$, aromatic rings, etc., with the intervals of the characteristic features in all three different kinds of spectra (if applicable). For example, the carbonyl group, contained in any environment, can have a ^{13}C NMR signal in the range 150–220 ppm and a strong IR absorption band within the interval of 1,630–1,880 cm^{-1} . This library is used at the first stage of filtering to reject structures containing the main functional groups in any environments if the structure is not confirmed by spectra.

2. *Library of common functional groups with defined environment.*

This library also contains common functional groups, but in this case a more precise description of the group and its surrounding environment is given (for instance, $C-CO-C$, $C-CO-O-C$, $CO-OH$, $C-CO-C=C$, $C-CO-Ar$, etc.).

1.3.3.3 Specialized Libraries

1. *Library specialized for ^{13}C -NMR.* This library includes fragments consisting of, as a rule, a central carbon atom in different environments. The library has proven to be very effective in the filtering of generated structures.
2. *Three libraries specialized for 1H and ^{13}C -NMR*

These libraries (CH_3 -Lib, CH_2 -Lib, and CH -Lib) are adjusted for the NMR detection of CH_3 , CH_2 , and CH groups with regard to different environments and the states of carbon atom hybridization. If the fragment environment includes some functional groups (for instance, $C=C-CH_2-C=O$, $CH_3-C=O$, $Ar-CH-C$), then the NMR features of these groups are considered as essential spectral properties of the corresponding fragments.

1.3.3.4 Structural Filter

The structural filter enables the chemist to select the structures which are both in accordance with constraints imposed by the user and some restrictions following

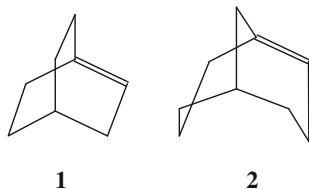
from the main regulations of structural chemistry and stereochemistry. Three groups of constraints are provided for structural verification:

Substructural constraints. This group of constraints consists of three fragment libraries, two of them (GOODLIST and BADLIST) being empty before the problem starts running, and the third library is a List of Unlikely Fragments. The GOODLIST and BADLIST libraries may contain obligatory (GOODLIST) and forbidden (BADLIST) fragments introduced by the chemist according to a priori information, experimental data, or theoretical considerations during problem solving. The input of fragment structural formulae is performed either using an integrated structure editor, ACD/ChemSketch, or by drawing fragments directly in a corresponding window—GOODLIST/BADLIST.

The constraints imposed on the molecular skeleton can be introduced in the generalized form of “abstract fragments,” for example $A=A$, $A=A=A$, $A=A-A=A$, where “A” denotes a skeletal atom. One should bear in mind that the GOODLIST/BADLIST fragments are available to be overlapped.

The *Unlikely Fragment Library* is a list of a limited number of fragments which are unlikely to exist in terms of general organic chemistry and stereochemistry. For example, triple bonds and the allene substructure in small cycles, different highly strained unsaturated polycyclic substructures and some fragments containing heteroatoms (O–O–O, OH–C–OH, etc.). If some unlikely fragments seem to be probable in the context of a given problem, the chemist may delete them from the library before the check is started or disable the library during filtering.

Control using Bredt’s rule. If the molecular formula, degree of unsaturation, and the generator fragment set allow for the generation of bridged structures with double bonds, then the appearance of structures defying Bredt’s rule is quite possible. Bredt’s rule states that bi-cycles containing a double bond at the nodal atom of the bridge can exist if the larger of the two rings containing the double bond includes not less than eight atoms in the cycle. According to this rule structure **1** is unrealistic, while structure **2** in principle may exist.



Structure filtering by Bredt’s rule eliminates all contradictory structures.

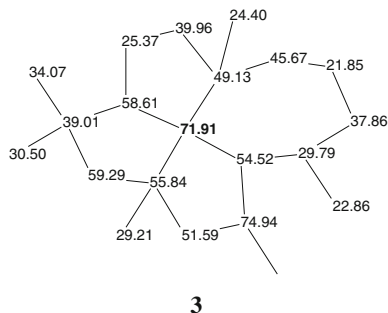
The structure filter is useful not only as a verification aid, but also as a tool for “step by step” selection of the most probable structures when the answer file is large.

1.3.3.5 Atom Property Correlation Table (APCT)

An Atom Property Correlation Table (APCT) was generated from the system knowledge base. The table contains carbon atom-centered fragments with the corresponding intervals of the ^{13}C NMR chemical shift variation for the central carbon atom and the ranges of ^1H chemical shifts corresponding to hydrogens attached to the central atom. The program uses this table for the automatic assignment of the hybridization (sp^3 , sp^2 , sp , sp^2 or sp^3 , *not defined*) to all carbons and for assessing the possibility of their neighboring heteroatoms (*forbidden*, *obligatory*, *not defined*). The mark “*not defined*” is assigned to a parameter if several conceivable possibilities are equally probable.

1.3.3.6 Optimization of Library Spectral Ranges

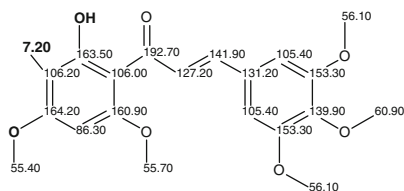
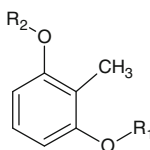
To optimize the chemical shift intervals assigned to the fragments present in the axiomatic knowledge (Filter Library and APCT), structures contained within the database of assigned ^1H and ^{13}C NMR spectra were used. The Database I structures were filtered with the help of the fragment libraries. If any contradictions were detected between a structure and the characteristic spectral intervals associated with a fragment then the program provided a corresponding message. Based on these messages the intervals were either modified to resolve the contradiction or were left unchanged. Changes were not made in those cases where the contradiction was caused by the presence of “anomalous” chemical shifts in a structure, a so-called exotic structure. An example of such a structure [28] is illustrated below:



In structure **3** the chemical shift of the quaternary carbon atom that is common to all four cycles is 71.91 ppm, a value which is typical for a carbon neighboring an oxygen atom.

The assumption was made that such anomalous shifts were quite rare. The risk of overlooking the correct structure in rare cases was justified by the possibility of solving a great number of problems using chemical shifts corresponding to the

common values similar to those known from the literature. Some fragments characterized by specific chemical shift values were placed into a *Library of Exceptions* that was applied as part of the structure filtration process using a specially derived algorithm. For example, in structure **4** the anomalously low carbon chemical shift value of the CH₃ group (7.20 ppm) is typical of a methyl group attached to a benzene ring with two neighboring substituents of the type O–R. The reason for this is that oxygen atoms influence carbon shifts at the β position moving the resonances upfield. These effects are cumulative in a manner similar to β branching. The fragment **5** shown below was therefore introduced into the library of exceptions:

**4****5**

Analysis of the fragment tables revealed that spectral filtering libraries and correlation tables (APCT) allowed for the solution of tasks with a negligible risk of overlooking the correct structure. 98 % of the structures present in the system database withstood a verification challenge by both ¹H and ¹³C NMR spectra using spectral filters and ACPT. Taking into account the high degree of diversity of the structure library it can be expected that spectral filtering of the output file is a procedure that offers only a small risk of losing the actual structure. Nevertheless, if a researcher does not find expected structures in a filtered structural file then filtering can be switched off during the structure generation. As we will see (Sect. 1.4) the algorithms for NMR chemical shift prediction implemented into Structure Elucidator are so fast that spectral and structural filtering may be skipped if a modern high speed PC is used.

1.3.3.7 Library of Typical Functional Groups

To help chemists to initially relate an unknown to some chemical classes the user can use a function of the program to display a “generalized portrait” of the analyzed molecule, i.e., the distribution of typical functional groups with the frequency of their occurrence in the fragments found as a result of a Fragment Library search by ^{13}C NMR spectrum. A special Library of Typical Functional Groups is used to form this generalized portrait (see examples in Fig. 1.10). The high frequency of occurrence of a group indicates that it is present in a molecule with higher probability. The groups that are completely absent in the found fragments can be placed in the User Defined Badlist if desirable.

Filtering the fragments found in a database allowed us to use this procedure for getting hints to the presence of functional groups in a molecule that cannot be detected directly from the ^{13}C and ^1H NMR spectra. For instance, if it becomes evident that a significant number of fragments contain an $-\text{NO}_2$ group, then the possibility of the presence of this group in the molecule under analysis should be taken into account.

1.4 NMR Prediction in the *Structure Elucidator* System

Depending on the rigorous nature of the structural constraints imposed by the experimental data, the output file of StrucEluc may contain tens to tens of thousands of structural formulae. A correct structure cannot easily be distinguished by taking

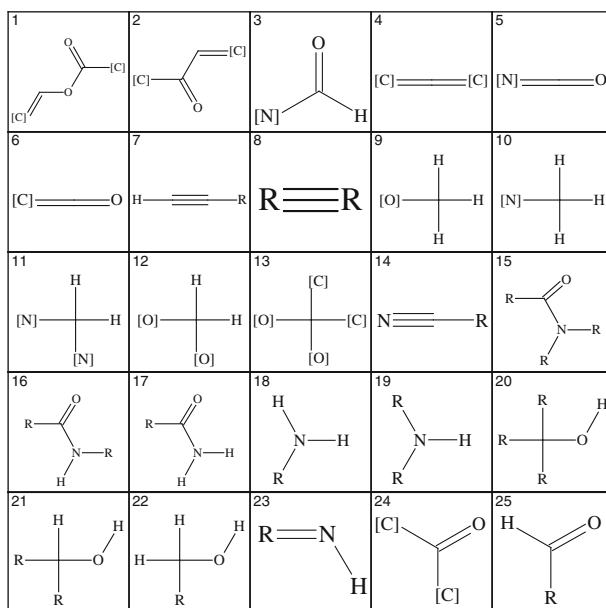


Fig. 1.10 Examples of substructures included in the library of typical functional groups

into account changes in the characteristic spectral features of the functional groups and fragments existing in the probable structures. Therefore, the selection of the most probable structure is carried out by comparing experimental to predicted spectra and this step is generally at the conclusion of the Structure Elucidator workflow (see Fig. 1.6).

In the early days of the development of CASE systems a question was posed: What kinds of molecular spectra would be most amenable to prediction for distinguishing the “best” structural hypothesis among other competing structures? The appropriate methods of spectrum prediction would need to meet, at least, the following requirements:

- (a) The speed of spectrum calculation must be fast enough to be applied to large structural files.
- (b) The calculations should be as automated as possible and, preferably, fully automated.
- (c) The application of the methods must not depend on molecular size.
- (d) The methods must be of high enough accuracy to discriminate between similar candidate structures.

It quickly emerged that only *empirical* methods of NMR prediction had an opportunity to address the goals. MS spectral prediction cannot provide theoretical spectra of arbitrary chemical structures, though there are many variables in the mass spectral conditions that can lead to various fragmentation pathways, and fragmentation prediction is used more as a support filter for structure identification. Examination of the potential of IR spectrum prediction shows [29] that semi-empirical methods based on the valence-optical scheme [30] as well as quantum-chemical methods do not satisfy the requirements (a)–(c) and, consequently, they should be declined. In addition, it is challenging to compare experimental IR spectra with calculated IR spectra when band intensities are taken into account: IR spectra of similar isomers frequently differ only in the shape and intensity of some of the absorption bands.

In the 1970s the first programs for empirical ^{13}C NMR chemical shift prediction [31] based on additive rules appeared. These methods of spectral prediction simulation, in contrast to MS and IR spectra, supported the requirements outlined in points (a)–(d) and are amenable to further improvements in the accuracy and speed of chemical shift calculation. During the last two decades quantum-mechanical (QM) methods of NMR chemical shift calculation were shown to be capable of predicting the NMR spectra of mid-sized molecules with accuracy, which is sufficient to select the preferable structures among those suggested by the researcher. This improvement was possible due to the development of DFT (Density Functional Theory) methods for nuclear magnetic shielding calculations, for which GIAO (Gauge Independent Atomic Orbital) approximation is the most frequently used [32]. These methods are, however, rather time-consuming so far and satisfy only requirement *d*.

As a result of the efforts of many research groups a series of programs for NMR prediction are now available to the chemical community. ^1H and ^{13}C spectra are

clearly the primary nuclei of interest and are most utilized by chemists for structure verification. The development of NMR prediction tools have thus focused on these nuclei. NMR prediction software has been incrementally improved over the past two decades and these tools can now provide reasonable to excellent accuracy in the quality of NMR prediction (*vide infra*). Spectroscopists also regularly make use of other nuclei during their structure elucidation efforts and the predictions of chemical shifts and, in some case, coupling constants for other nuclei have also been pursued. Along with chemical shift prediction for ^1H and ^{13}C nuclei, *Structure Elucidator* is also capable of predicting chemical shifts for ^{15}N , ^{19}F , and ^{31}P nuclei.

Within the ACD/NMR Predictor software three main algorithms are used for NMR spectrum prediction *independent of the type of nucleus*.

Method based on additivity rules. An additive linear model for calculating chemical shifts was suggested [33, 34]. This approach is the simplest and it has been used since the 1960s when calculations were carried out manually. Now the model uses PLS (Partial Least Squares) and increments accounting for the effects of atoms existing in the environment of the prediction center [35, 36].

Method based on HOSE code (fragmental method). To predict chemical shifts, extensive databases containing the structural formulae of organic compounds along with experimental chemical shifts assigned to the corresponding atoms (^{13}C , ^1H , etc.) are used. The environment of each atom is described by a so-called HOSE (Hierarchical Ordering of Spherical Environments) code [37], which allows for the detection of such atoms in the database that have environments similar to those in the investigated molecule.

Methods based on algorithms of artificial neural networks [38]. These methods appeared during the last two decades and they are based on program learning to predict chemical shifts. During the learning process training sets are used. As a result, such rules are produced which allow the program to predict chemical shifts automatically with high speed and reasonable accuracy.

ACD/NMR Predictor is an integral part of the *Structure Elucidator* software. We believe that the incremental method of chemical shift prediction is well known to chemists so here we will focus only on the HOSE code and neural nets-based approaches.

1.4.1 ^{13}C NMR Chemical Shift Prediction

1.4.1.1 Method Based on HOSE Code (Fragment-Based Method)

Databases containing chemical structures and the assigned carbon chemical shifts form the foundation data set to allow for the derivation of prediction algorithms. For every carbon atom associated with each chemical structure contained in the database, atom-centered fragments (ACF) with a prescribed number of concentric layers are generated according to the HOSE code. These fragments, and their corresponding chemical shifts, are stored as an ordered list for use in the prediction

algorithms. To predict the spectrum for a candidate structure, the program selects all possible ACFs present in a structure, performs a search for their analogs in the database, and after statistical processing ascribes the chemical shifts taken from the reference fragments to the carbon atoms being predicted. If an ACF is not found in the database then the program interpolates using the most similar structural environments available. The results obtained using this approach are generally in good agreement with the experimental data when using a large database containing a diverse set of structures.

The ^{13}C NMR shift prediction program, ACD/CNMR [39], has an internal data file containing over 3,018,000 experimental ^{13}C chemical shifts and over 123,000 coupling constants characterizing interaction with different nuclei. Each δC value is preassigned to a specific carbon nucleus in its unique environment. The data file has been generated from the assigned spectra of over 355,000 chemical structures. The program uses many different algorithms for estimating δC for those fragments not represented in the internal database of experimental values. Both the algorithms and the details of their derivation constitute proprietary knowledge and have not been described in the literature although the general concepts are described below.

When a new chemical structure is input into ACD/CNMR the program automatically splits the structure into a set of unique fragments that are then compared to the structural fragments from the internal database.

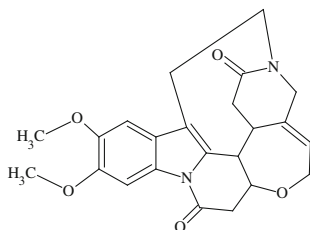
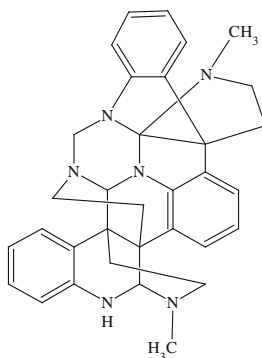
- If a fragment from the drawn structure *coincides* with a fragment contained within the database, the program will use this experimental δC value as part of the final set of chemical shifts for the structure. For such δC values the program will not show confidence intervals in the table of chemical shifts. The program utilizes a reference structure of up to 16 spheres in radius for a particular carbon atom. As a result the size of the fragment is defined by the size of the largest fragment common to both the predicted and the reference structure, the fragments being centered on the given carbon atom.
- If some fragments from the structure cannot be found in the internal database, then the program will search for the most similar fragments in the database. *First*, the program composes sets of fragments from the database that are structurally similar to each of the fragments generated from the analyzed structure. *Second*, the program estimates the δC values for the fragments from the structure using secondary algorithms and compares them to the estimated δC values of fragments selected from the database. This allows the program to narrow down to a set of similar fragments from the database. *Third*, the program calculates both the average values (δ_{Av}) of the experimental data and produces estimated δC values after application of the second criterion described above. The resulting δC value is calculated using both the estimated δC value of the given fragment and the average δ_{Av} values. The obtained δC values are used to compose the final set of chemical shifts for the structure. After composing the final lists of chemical shifts, ACD/CNMR generates the exact number, location, intensities, and assignment of the spectral lines associated with the structure.

The array of chemical shift, coupling constants, and line width parameters describing an NMR spectrum are influenced by many external factors including solvent, concentration, temperature, relaxation times, concentration of paramagnetic impurities, shimming, and observation frequency to cite just a few. Many of these parameters are simply too complex to take account of during a prediction but certainly solvent dependence can be accounted for to a certain extent. The ACD/CNMR Predictor provides the ability to perform solvent-specific predictions. The user can select from a list of common NMR solvents and predict a solvent-specific NMR spectrum. The stereochemistry of a particular structure is crucial in determining the molecular properties and when the stereochemistry of an atom is included in the submitted chemical structure the information is utilized during the prediction process.

Since many basic chemistry research programs are focused primarily on the development of new and novel chemical structures it is possible that specific fragments are not yet described in the literature and consequently are not contained within the databases used as the foundation of the algorithms. The ideal situation would be to allow a scientist to not only capture and catalog their own structures and assignments but to use this information directly in the prediction algorithms.

To this end ACD/CNMR allows the chemist to create a *user database* of structures and assignments. In this case, the program again splits the structures into unique fragments. As new chemical shifts are assigned to atoms the program treats these data as an update to its internal database. If a spectrum is predicted for a new chemical structure while the user database is open, the program performs all of the same actions described above, but pays primary attention to the data that have been entered in the **User Database**. This form of user database training can have a dramatic impact on the ability of an organization to predict NMR spectra for a diverse array of compounds containing structural moieties that have been represented at a fairly minimal level of 1–2 structures in the training database. With this capability not only are the legacy data generated as a result of the structure elucidation process available for reference through searching but also it forms the foundation for a large number of organizational scientists to improve or benefit from their own predictions.

^{13}C NMR spectral prediction relying on the fragmental approach produces predicted spectra for diverse structures with a precision that is most frequently sufficient for distinguishing the correct structure in the set of candidates included into the valid solution (i.e., a solution containing the genuine structure). The main shortcoming of this approach, however, is the speed of prediction. For instance, ACD/CNMR Predictor takes about 3 s to calculate the ^{13}C NMR spectra of the typical natural products shown below when both structures are absent from the database (PC 3.4 GHz):

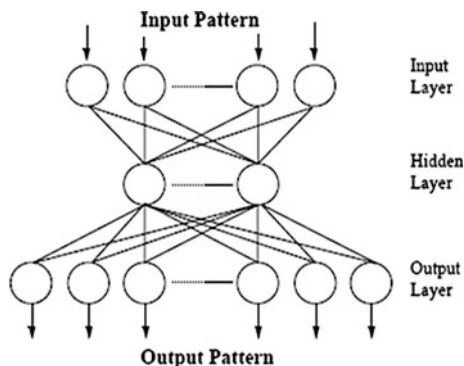
**6****7**

This speed of calculation is quite acceptable for single calculations or even for a batch predictions when the number of structures is not large, for example 50–100 structures. However, if the intent is to apply a spectral-prediction program to thousands of hypothetical structures, then achieving the same high calculation precision becomes a time-consuming issue. As we will see later (Sect. 2.2.2), this difficulty is readily circumvented in Structure Elucidator by application of a multistep strategy of chemical shift prediction when two very fast methods are used in the first steps. These fast methods utilize incremental and neural net-based approaches.

1.4.1.2 Artificial Neural Networks (ANN)

Beginning in the early 1990s, the attention of chemists was drawn to the possibilities of promising new mathematical tools developing in computer-based chemistry, viz., artificial neural networks (ANN). There was a rapid increase in the number of studies on the application of ANN to the interpretation, classification, and prediction of spectral data, including NMR chemical shift prediction.

Fig. 1.11 Schematic diagram of a neural network



A neural network can be considered as a simplified computer model of the human brain consisting of several layers of neurons that send signals to other neurons as a function of the input signals received. A flowchart representing a neural net is presented in Fig. 1.11. Typically, a neural network consists of an Input Layer containing a certain number of *input* neurons, Hidden Layer(s) containing hidden neurons, and an Output Layer. Such networks have a “black box” nature and possess the common ability to construct empirical models of the systems for which theoretical dependencies between the input and output are extremely complicated or even unknown. Models are obtained as a result of training the network. In the course of training, the network is represented in the form of input–output pairs related by a simulated transformation. A network trained using these examples is able to predict the output signals from input examples not presented originally in the training set. The training procedure may be time-consuming (tens of hours) but a network, once trained, generates the result almost instantaneously. For instance, a network can be trained to generate structural information (output) retrieved from a spectrum (input) or to predict a spectrum (output) from structural information (input). The theory of ANN and examples of their application in chemistry and spectroscopy can be found in a monograph by Zupan and Gasteiger [38].

Artificial neural networks are trained to predict the spectra of compounds via a training set including encoded structures and their associated spectra. In the course of training, the network uses reference structures as input information and the output signals are compared with the ^{13}C NMR spectra of these structures. The training process is complete if the deviations of the predicted spectra from the reference set are less than a chosen threshold. In general, artificial neural networks have somewhat better performance than linear models in terms of spectral prediction.

The results of the computational experiments performed with ACD/CNMR Predictor employing HOSE code based, NN, and PLS algorithms produced the statistical data necessary for comparison of performance of all three methods [35, 36, 40]. The mean error of all three methods varies between 1.6 and 1.8 ppm,

while NN and PLS can provide results of similar quality after being properly optimized.

The fragmental approach fails with structures that are underrepresented in the database and this raises the average deviation and the maximum error. In such cases, PLS and NN methods perform better. If structures necessary for chemical shift calculation of all carbon atoms are present in the database, the HOSE code-based spectrum prediction is of good accuracy.

1.4.1.3 Speed of ^{13}C Chemical Shift Prediction

As mentioned above, Structure Elucidator is based on the utilization of 1D and 2D NMR data and this allows for the identification of newly isolated materials such as natural products or synthesized organic molecules. The system is capable of elucidating large molecules containing 100 or more skeletal atoms. Since the initial structural information extracted from 2D NMR spectra is fuzzy by nature (see Sect. 1.2) the number of structures that are consistent with the 2D NMR data can be rather large (up to tens and even hundreds of thousands). As a result the selection of the most probable structure from a large output file requires an approach whereby the expert systems can utilize both *accurate* and *fast* approaches for NMR chemical shift prediction.

The prediction speed was estimated by performing the spectral prediction of thousands of candidate structures generated by the program. It was found that the average speed of ^{13}C chemical shift prediction by the incremental method was about 30,000 shifts per second on a 3.4 GHz PC computer while the neural network-based algorithm was also fast, but approximately 2.5–3 times slower. The combination of this high speed of prediction with an appropriate accuracy (an average deviation of 1.60–1.80 ppm) makes both of these approaches powerful tools for computer-aided structure elucidation. The strategy of combined application of all three algorithms of NMR chemical shift prediction for selection of the most probable structure will be explained in Sect. 2.2.2.

1.4.1.4 Comparison of Chemical Shift Prediction Methods

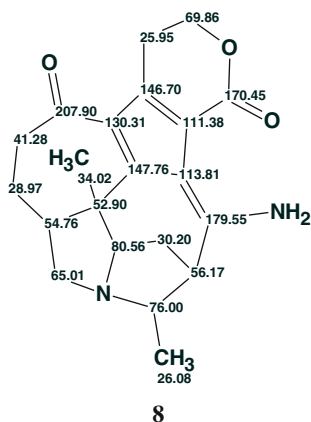
It should be noted that the HOSE code methods relying on fragment databases are readily capable of incorporating new structures, while neural networks would need to be retrained dynamically to adjust the prediction algorithms to novel chemical classes. In contrast, quantum-chemical methods do not need any adjustment for the prediction of new classes of substances.

The important advantage of the HOSE code-based approach is the large diversity of structures used for the spectrum prediction. The approach also takes into account the stereochemical configuration of the molecule being analyzed. This can provide a higher probability of obtaining the smallest deviation for the ^{13}C NMR

spectrum calculated for the right structure if the Structure Elucidator output file represents a valid problem solution.

A particularly useful feature of HOSE code-based software programs includes the ability to review the details of how a predicted spectrum is generated. In the Structure Elucidator program a **Chemical Shift Calculation Protocol** can be displayed. When a prediction is performed on a chemical structure that is absent from the database then the different structures used to produce the predicted spectrum are indicated. This protocol allows for comparison of the environments of the carbon atoms for which the chemical shifts are calculated with the environments and chemical shifts of the related structures are used for prediction. The analysis of the spectrum-structure information presented in the protocol dialog box allows the chemist to validate the accuracy and reliability of the chemical shift prediction for a given atom. To see the protocol, a right mouse click on the corresponding carbon atom in the structure is enough.

For example, for the carbon nuclei at 65.01, 76.00, and 179.55 ppm in the complex structure of the natural product daphnipaxinin **8**, HOSE code-based chemical shift prediction showed significant discrepancies between the experimental and calculated values (10–15 ppm).



The ACD/CNMR program predicted values of 55.26, 61.71, and 166.21 ppm respectively. The cause for the large discrepancies between the experimental and predicted shifts was clarified by the Calculation Protocol. The protocol produced by the program for the carbon atom at 65.01 ppm is presented in Fig. 1.12.

The protocol indicates that 28 hits were selected from the database for prediction of the chemical shift of the carbon nucleus under consideration. The histogram indicates the distribution of hits with chemical shifts assigned to representative carbon nuclei used for prediction. In the left part of the figure an example structure from the set of hit structures is shown. The fragment centered on the carbon atom of interest (in square) is colored red. The histogram shows that chemical shifts

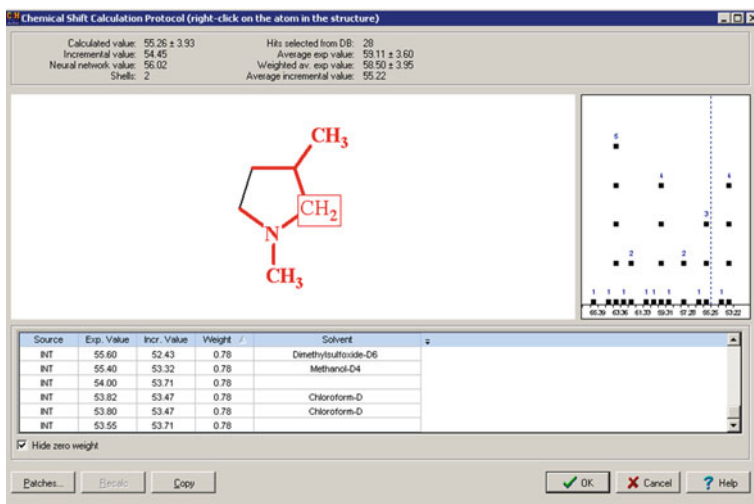


Fig. 1.12 The chemical shift calculation protocol corresponding to C 65.01

assigned to the atom under consideration are scattered between 53 and 66 ppm and the calculated value is 55.26 ± 3.93 ppm. Therefore, the great spread of data is the reason for the poor chemical shift prediction.

When the protocol related to the carbon atom with an experimental shift of $\delta C = 76.00$ ppm was requested, the program displayed the following protocol (Fig. 1.13):

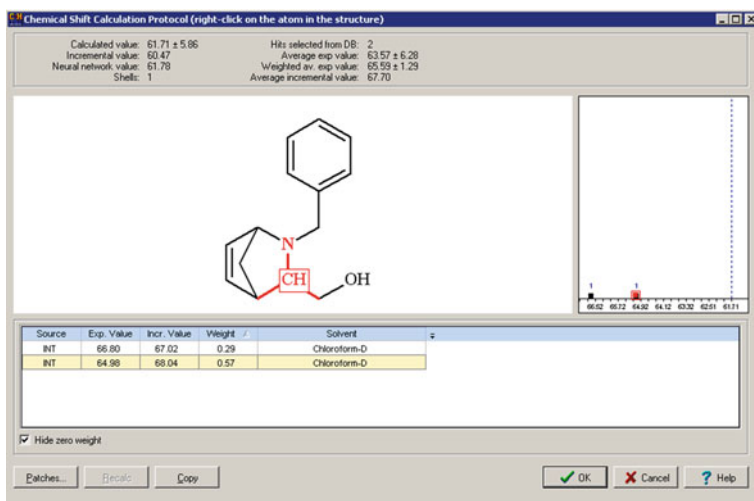


Fig. 1.13 The chemical shift calculation protocol corresponding to C 76.00

The protocol shows that only two hits were found in the database, while the “working” fragment manifests a similarity with the environment of the investigated atom. The following protocol message was delivered for C 179.55:

only within one chemical bond around the center and the reference chemical shifts are 65 and 66.8. The calculated chemical shift value is 61.71 ± 3.93 ppm.

Calculated value: 166.21 ± 10.81	Hits selected from DB: 0
Incremental value: 166.21	Average exp value: n/a
Neural network value: 162.32	Weighted av. exp value: n/a
Shells: 0	Average incremental value: n/a

It follows from this message that no structure was found that contains a carbon centered fragment suitable for chemical shift calculation for the given atom. Therefore, the chemical shift calculation was performed using an incremental approach that yielded $\delta C = 166.21$ ppm and the consequent discrepancy (~ 15 ppm) relative to the assigned shift. It should be kept in mind that the accuracy of prediction significantly depends on the degree of environment similarities intrinsic for an atom center under analysis and the associated reference structures. For instance, for C 41.28 the following protocol was produced (Fig. 1.14 shows a part of it):

We can see that the environment of the center carbon atom in structure **8** is similar to that in the single reference structure, as a result of which the calculated chemical shift (41.8 ppm) is very close to the experimental one (41.28).

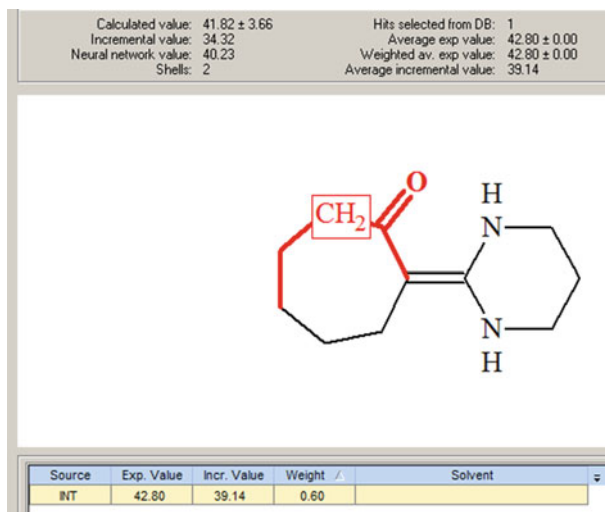


Fig. 1.14 The chemical shift calculation protocol corresponding to C 41.28

As mentioned, the neural networks function as a “black box” and do not provide any access to the details associated with how the chemical shift for a given carbon atom is calculated. However, as stated previously neural networks are significantly faster for ^{13}C NMR chemical shift calculation in comparison with the fragmental approach, and this makes them more effective in the prediction of assignments for large files of chemical structures. Nevertheless, as we will see later (Part III), a rational combination of the different approaches—an incremental approach, an approach based on HOSE code, and ANN methods—provides an optimal strategy for identification of the correct structure even in very large output files.

1.4.2 Prediction of ^1H NMR Spectra

^{13}C NMR spectra are both relatively simple to predict as well as to compare with experimental data since carbon NMR spectra consist almost exclusively of singlets or clearly defined multiplets due to heteronuclear coupling with magnetically active nuclides. The prediction of ^1H NMR spectra, in contrast, is much more complex because they contain both first-order and higher order multiplets. The relative complexity of proton spectra accounts for the greater difficulty in comparing the experimental and predicted spectra in an automated fashion. The latter observation likely explains why there are a limited number of software programs for proton NMR relative to carbon chemical shift NMR prediction.

The algorithms that were developed for ^{13}C chemical shift prediction using HOSE code, neural nets, and additive rules (PLS)-based models were also implemented into the ACD/HNMR predictor software for calculation of ^1H NMR chemical shifts. The ACD/HNMR Predictor (like ACD/CNMR predictor) is embedded into Structure Elucidator as an integral part.

Here we will dwell on the main traits of the HOSE code-based approach of ^1H chemical shift prediction. As was mentioned previously chemical shift prediction is carried out using large databases.

The internal database contains almost 2,176,000 experimental ^1H chemical shifts and over 824,000 coupling constants. The algorithm used by the program to quantify spin–spin interactions is based on parameters determined for more than 3,000 structural fragments. Each δH value in the database is assigned to a particular proton nucleus and each coupling constant is assigned to a pair of interacting nuclei. The internal database of assigned δH and J_{HH} values has been extracted from the analysis of 261,000 experimental ^1H NMR spectra. The δH and coupling constant values for proton nuclei in fragments not contained in the database are calculated using proprietary algorithms.

^1H chemical shift calculations can be performed as a function of the spectrometer frequency using an optional parameter setting. Calculated chemical shifts and coupling constants are provided with 95 % confidence intervals to provide a measure of the reliability of the calculated values. The calculated spectrum accounts for second-order interactions and long-range coupling constants thereby allowing

for the simulation of spectra with strongly coupled spin systems containing up to eight magnetically inequivalent nuclei. For magnetically equivalent nuclei a larger number of interacting spins can be taken into account. The algorithm recognizes *cis/trans* isomers for alkenes, *syn-anti* isomers of amides, oximes, hydrazones, and nitrosamines, and axial-equatorial isomers of ring systems. The chemical shifts of protons belonging to OH, NH, and SH groups can also be predicted. The program takes into account solvent effects in the following way: if a solvent is specified the program selects for the predictions only those reference structures whose ^1H NMR spectra were recorded in the specified solvent in the database. The algorithm for ^1H chemical shift prediction is similar to that used for the prediction of ^{13}C NMR chemical shifts. After input of a structure for prediction, the program generates a set of unique fragments that are compared with the fragments contained in the internal database. If any generated structural fragment coincides with a fragment from the database then its experimental δH values are included into the result set of chemical shifts generated for the structure. If some fragments are not identified in the database then the chemical shifts are extrapolated using parameters associated with similar fragments. Once again a **Chemical Shift Protocol** dialog window, similar to those shown in Figs. 1.12, 1.13 and 1.14, allows inspection of the calculation associated with each chemical shift and can be valuable in probing any discrepancies observed.

After composing the final lists of chemical shifts and coupling constants the program composes and diagonalizes the spin Hamiltonian matrix to calculate the number of lines, their positions, and intensities in order to assign the δH values to the hydrogen atoms of the structure under consideration. The program generates a spectral plot including the positions of the lines and the integral. In the process, the program takes into account the spectrometer frequency, which provides a visual display to allow comparison of the calculated spectral pattern with experimental spectrum. The calculations of the chemical shifts, spectral contours, and integral curve can take from several seconds to tens of seconds on a standard PC. ^1H chemical shift prediction is applied to structures placed at the beginning of a rank-ordered final list to help select the “best” structure. The mean error of the ^1H chemical shift prediction is 0.18 ppm. In order to reduce the error to that experienced in experimental determinations further optimization such as the detailed description of the 3D geometry might be necessary. The speed of prediction is comparable with that achieved for ^{13}C chemical shift calculation (30,000 shifts/s for incremental approach, 10,000—for neural nets and 15–20—for the HOSE code).

1.5 Determination of Relative Stereochemistry

The biological activity of natural products and drug molecules is generally dependent on the stereochemistry of a molecule. Indeed, there are many examples known where one stereoisomer can exhibit vastly different pharmacological activity from the other stereoisomer. Stereochemical considerations can also influence

reaction pathways and certainly reaction kinetics, with one form being favored over another. Generally, the final step in contemporary structure characterization efforts is to define the relative and, if possible, absolute stereochemistry. NMR methods are generally well suited to the former while the latter is generally obtained using chemical structure modification combined with NMR studies [41] or by X-ray crystallographic methods. Determination of relative stereochemistry from NMR data is based on the nuclear Overhauser effect (NOE), which is dependent on the distance separating the cross-relaxing nuclides [42]. Typically, NOESY or ROESY two-dimensional NMR experiments or their selective 1D analogs [43] are used to provide the data for this analysis in rigid molecules. In the case of flexible molecules, considerable effort has been devoted to the development of J-based NMR methods that are used to measure long-range heteronuclear coupling constants that can then be used to assign the relative stereochemistry.

With this in mind, the capabilities of Structure Elucidator were enhanced by a software program that allows for the determination of the relative stereochemistry of a molecular structure based on the nuclear Overhauser effect constraints [44]. The program extracts NOE information from either NOESY and/or ROESY spectra and determines the molecular stereochemistry accordingly. The results of selective NOE or ROE experiments can also be used for input to the program. This process can be carried out for several of the most likely structures produced during a structure determination using Structure Elucidator or performed on a chemical structure proposed by the chemist.

The utility of NOESY/ROESY spectra for relative stereochemistry determination is based on a direct correlation between both the cross peak volume integration and the internuclear distance. Peak intensity in NOE/ROE measurements has an inverse sixth power relationship. Consequently, what can be referred to as a “strong” NOE is generally observed between pairs of hydrogens which are 1.8–2.5 Å apart. Responses of “medium” intensity usually correspond to an internuclear distance of 2.5–4.0 Å while “weak” NOEs will generally be observed for larger distances if they are observed at all. NOE responses are not commonly observed for nuclei farther than 6.0 Å apart.

Minimization algorithms deal with *numerical values* and, in this case, these numerical values are extracted from a set of NOEs overlaid on a 3D structure and examined for goodness of fit. The function describing this goodness of fit is called a *penalty function*. The better the solution, the lower the value of the function. The function must exhibit the lowest value for the best-matching stereoisomer.

An appropriate function [44] was suggested that can be minimized by calculation for *all* stereoisomeric structures or by using a genetic algorithm [45] to limit the number of stereoisomers that need to be investigated. Suggested methods for determining the relative stereochemistry of structures as well as the calculation of their 3D configurations were examined using complex structure examples.

Along with the described method based on NOE and molecular mechanics, we considered another, simpler approach, which can be helpful for determining the relative stereochemistry of the molecule. Taking into account the advantages of a fragmental approach for ^{13}C chemical shift prediction we hypothesized that it could

help in preliminary selection of a set of the most probable stereoisomers for their subsequent verification by additional experimental techniques, QM chemical shift prediction and 3D optimization using NOE as described above. This may be possible since the stereocenters associated with structures are included into the ACD/CNMR database and stereochemistry is taken into account by the NMR chemical shift prediction algorithms. The incremental and neural nets-based algorithms of chemical shift prediction also use the stereochemistry information related to the atoms included in 3–6-membered cycles [35].

This hypothesis was verified [46] using a series of examples taken from the literature for novel structures for which the relative stereochemistry was reported. Hence these structures were deliberately absent from the ACD/CNMR database. As a result, the application of empirical methods of ^{13}C NMR chemical shift prediction is shown to indeed allow for the selection of a limited set of the most probable stereoisomers which always includes the genuine stereoconfiguration.

References

1. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ, Martin GE (2004) Structure elucidator: A versatile expert system for molecular structure elucidation from 1D and 2D NMR data and molecular fragments. *J Chem Inf Comput Sci* 44:771–792
2. Lederberg J, Sutherland GL, Buchanan BG, Feigenbaum EA, Robertson AV, Duffield AM, Djerassi C (1968) Application of artificial intelligence to chemical inference. I. The number of possible organic compounds. Acyclic structures containing C, H, O and N. *J Am Chem Soc* 91:2973–2976
3. Nelson DB, Munk ME, Gasli KB, Horald DL (1969) Alanylactinobicyclon. An application of computer techniques to structure elucidation. *J Org Chem* 34:3800–3805
4. Sasaki SI, Abe H, Ouki T, Sakamoto M, Ochia SI (1968) Automated structure elucidation of several kinds of aliphatic and alicyclic compounds. *Anal Chem* 40:2220–2223
5. Elyashberg ME, Gribov LA (1968) Formal logic method of infrared spectrum interpretation. *Zh Prikl Spectrosk* 8:296–300
6. Gribov LA, Elyashberg ME (1979) Computer-aided identification of organic molecules by their molecular spectra. *Crit Rev Anal Chem* 8:111–220
7. Pretsch E, Bühlmann P, Affolter C (2000) Structure determination of organic compounds—tables of spectral data. Springer, Berlin
8. Elyashberg ME, Williams AJ, Martin GE (2008) Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation. *Prog NMR Spectrosc* 53(1, 2):1–104
9. Elyashberg ME, Williams AJ, Blinov KA (2012) Contemporary computer-assisted approaches to molecular structure elucidation, vol 1. New developments in NMR. RSC Publishing, Cambridge
10. Reichenbacher M, Popp J (2012) Challenges in molecular structure determination. Springer, Heidelberg
11. Elyashberg M, Blinov K, Molodtsov S, Smurnyy Y, Williams AJ, Churanova T (2009) Computer-assisted methods for molecular structure elucidation: realizing a spectroscopist's dream. *J Cheminform* 1:3. doi:10.1186/1758-2946-1-3
12. <http://www.cas.org/content/chemical-suppliers>
13. Elyashberg ME, Gribov LA, Serov VV (1980) Molecular spectral analysis and computer (in Russian). Nauka, Moscow
14. Gray NAB (1986) Computer-assisted structure elucidation. Wiley, New York

15. Munk ME (1998) Computer-based structure determination: then and now. *J Chem Inf Comput Sci* 38:997–1009
16. Gribov LA, Elyashberg ME (1970) Symbolic logic methods for spectrochemical investigations. *J Mol Struct* 5:179–198
17. Stoll RR (1961) Sets, logic and axiomatic theories. W.H. Freeman and Company, San Francisco
18. Elyashberg ME, Williams AJ, Blinov KA (2010) Structural revisions of natural products by computer assisted structure elucidation (CASE) systems. *Nat Prod Rep* 48:571–574
19. Nicolaou KC, Snyder SA (2005) Chasing molecules that were never there: misassigned natural products and the role of chemical synthesis in modern structure elucidation. *Angew Chem Int Ed* 44:1012–1044
20. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ (2013) Structure revision of asperjinone using computer-assisted structure elucidation methods. *J Nat Prod* 76:113–116
21. Elyashberg ME (1998) Infrared spectra Interpretation by the characteristic frequency approach. In: Schleyer PvR, Allinger NL, Clark T et al (eds) *The encyclopedia of computational chemistry*. Wiley, Chichester, pp 1307–1312
22. Blinov KA, Carlson DV, Elyashberg ME, Martin GE, Martirosian ER, Molodtsov SG, Williams AJ (2003) Computer-assisted structure elucidation of natural products with limited 2D NMR data: application of the StrucEluc system. *Magn Reson Chem* 41(5):359–372
23. Blinov KA, Elyashberg ME, Martirosian ER, Molodtsov SG, Williams AJ, Sharaf MMH, Schiff PLJ, Crouch RC, Martin GE, Hadden CE, Guido JE, Mills KA (2003) Quindolinocryptotackeine: the elucidation of a novel indoloquinoline alkaloid structure through the use of computer-assisted structure elucidation and 2D NMR. *Magn Reson Chem* 41:577–584
24. Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. SIAM, Philadelphia
25. Molodtsov SG, Elyashberg ME, Blinov KA, Williams AJ, Martin GM, Lefebvre B (2004) Structure elucidation from 2D NMR spectra using the StrucEluc expert system: detection and removal of contradictions in the data. *J Chem Inf Comput Sci* 44:1737–1751
26. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ, Martin GE (2007) Fuzzy structure generation: a new efficient tool for computer-aided structure elucidation (CASE). *J Chem Inf Model* 47(3):1053–1066
27. Elyashberg ME, Blinov KA, Williams AJ, Molodtsov SG, Martin GE (2006) Are deterministic expert systems for computer-assisted structure elucidation obsolete? *J Chem Inf Model* 46(4):1643–1656
28. Weavers RT (2001) Laurenanes: fenestranes with a twist. *J Org Chem* 66:6453–6461
29. Elyashberg ME (1999) Expert systems for the determination of structures of organic molecules by spectral methods. *Russ Chem Rev* 68(7):525–547
30. Gribov LA, Orville-Thomas WR (1988) *Theory and methods of calculation of molecular spectra*. Wiley, Chichester
31. Clerc J-T, Sommerauer HA (1977) A minicomputer program based on additivity rules for the estimation of ^{13}C NMR chemical shifts. *Anal Chim Acta* 95:33–40
32. Lodewyk MW, Siebert MR, Tantillo DJ (2012) Computational prediction of ^1H and ^{13}C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem Rev* 112(3):1839–1862. doi:10.1021/cr200106v
33. Paul EG, Grant DM (1963) Additivity relationships in carbon-13 chemical shift data for the linear alkanes. *J Am Chem Soc* 85:1701–1702
34. Grant DM, Paul EG (1964) Carbon-13 magnetic resonance. II. Chemical shift data for the alkanes. *J Am Chem Soc* 86:2984–2990
35. Smurnyy YD, Blinov KA, Churanova TS, Elyashberg ME, Williams AJ (2008) Toward more reliable ^{13}C and ^1H chemical shift prediction: a systematic comparison of neural-network and least-squares regression based approaches. *J Chem Inf Model* 48(1):128–134

36. Blinov KA, Smurnyy ED, Curanova TS, Elyashberg ME, Williams AJ (2009) Development of a fast and accurate method of ^{13}C NMR chemical shift prediction. *Chemometr Intell Lab Syst* 97:91–97. doi:10.1016/j.chemolab.2009.01.010
37. Bremser W (1978) HOSE—a novel substructure code. *Anal Chim Acta* 103:355–365
38. Zupan J, Gasteiger J (1993) *Neural networks for chemists*. VCH, Weinheim
39. Advanced Chemistry Development. ACD/NMR predictors. Prediction suite includes ^1H , ^{13}C , ^{15}N , ^{19}F , ^{31}P NMR prediction. <http://www.acdlabs.com>
40. Blinov KA, Smurnyy YD, Elyashberg ME, Churanova TS, Kvasha MP, Steinbeck C, Lefebvre BA, Williams AJ (2008) Performance validation of neural network based ^{13}C NMR prediction using a publicly available data source. *J Chem Inf Model* 48(3):550–555
41. Seco JM, Quinoa E, Riguera R (2004) The assignment of absolute configuration by NMR. *Chem Rev* 104:17–117
42. Neuhaus D, Williamson M (2000) *The nuclear overhauser effect in structural and conformational analysis*, 2nd edn. Wiley, New York
43. Berger S, Braun S (2004) *200 and more NMR experiments: a practical course*. Wiley, New York
44. Smurnyy YD, Elyashberg ME, Blinov KA, Lefebvre B, Martin GE, Williams AJ (2005) Computer-aided determination of relative stereochemistry and 3D models of complex organic molecules from 2D NMR spectra. *Tetrahedron* 61:9980–9989
45. Mitchell M (1999) *An introduction to genetic algorithms*. MIT Press, Cambridge
46. Elyashberg ME, Blinov KA, Williams AJ (2009) The application of empirical methods of ^{13}C NMR chemical shift prediction as a filter for determining possible relative stereochemistry. *Magn Reson Chem* 47:333–341

Chapter 2

Strategies of Structure Elucidation

Abstract Different strategies regarding the application of the system depending on the specific features of the problem under analysis are discussed in this chapter. The input of 1D and 2D NMR spectroscopy data (automated and manual) to the program and the creation of electronic data tables are outlined. Special attention is focused on developing a Molecular Connectivity Diagram (MCD) as well as its editing and checking for consistency. An MCD represents visually *all initial information* employed by the system for structure generation. The application of the *Common* and *Fragment* modes of *Strict Structure Generation* depends on the particular peculiarities associated with a problem (molecule size, deficit of hydrogen atoms, etc.) and the methodology of the most probable structure selection from the output file is described. Special consideration is given to the problem of resolving logical contradictions in 2D NMR data arising from the presence of “nonstandard” *correlations* (those for which ${}^nJ_{\text{HH, CH}, n} > 3$). To this aim, a *Fuzzy Structure Generation* (FSG) algorithm is implemented into Structure Elucidator which allows for identification of the structure of the compound under analysis in the presence of an *unknown number* of nonstandard correlations of *unknown lengths*. Different modes of FSG are expounded and strategies for its application are discussed, instructing the student when and how each mode can be effectively employed.

2.1 Data Input, Processing, and Forming of a Molecular Connectivity Diagram

We assume that the reader is skilled and experienced enough to manually process raw 1D and 2D NMR data using traditional approaches (peak picking, structural interpretation of spectral features, etc.). These abilities are also necessary in all stages of CASE problem solving, especially during the preparation of spectral data for input into a computer. This stage is very important and it can be considered as a first step in forming the primary “axioms” and hypotheses (Sect. 1.2). It should be strongly emphasized that the application of the CASE system does not release the

chemist from having the necessary knowledge and experience in NMR spectroscopy. At the same time the program takes responsibility for the automatic creation of the majority of axioms and hypotheses necessary for structure elucidation, leaving their approval to the chemist. Once the initial set of “axioms” and hypotheses is adopted by the chemist further structure inference is performed by the system automatically. We also assume that the chemist has knowledge and experience in mass spectrum interpretation. Skills in the analysis of a peak cluster observed around the molecular ion and in deducing the molecular formula from an accurate molecular mass are crucially important for utilization of a CASE system. Therefore, in this chapter we will briefly describe the main steps of initial NMR spectral data preprocessing and data input into the program using the facilities and interface developed for Structure Elucidator. For this goal, a typical example will be used where the 1D and 2D spectra are of good quality and the molecule under investigation is a natural product of common complexity.

In the examples presented in Part III of this textbook, we will use 1D and 2D NMR data that were already processed and saved as electronic tables in the format common for Structure Elucidator. Primary attention will be placed on the methods of overcoming uncertainty, incompleteness, and contradictions in the initial data and the different operating modes provided in Structure Elucidator for these goals will be explained.

2.1.1 Data Used for Structure Elucidation

There are various types of data that can be used to perform computer-assisted structure elucidation. In particular, as described in the previous sections, 2D NMR is the essential technique for the elucidation of complex chemical structures and, due to its inherent complexity in terms of processing and manipulation, it is this form of NMR that puts significant demands on the software. The 2D NMR structure generator requires as input a set of atoms and the connectivities between them. The generator also takes into account the associated chemical shifts of the atoms as well as a series of different structural constraints. The base set of atoms is usually obtained from the molecular formula, while the connectivities between the atoms are revealed by combining the data encoded into the 1D and 2D NMR spectra.

The following spectra are commonly used in structure elucidation:

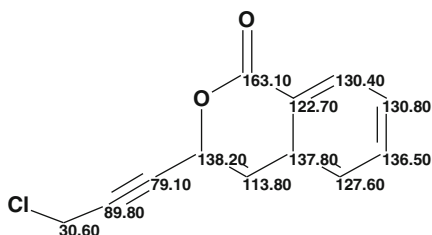
- **Mass spectra** mainly provide the molecular mass and elemental composition (in combination with NMR and IR spectra) and, as a result, access to the molecular formula for the compound under examination. Note that the determination of a molecular formula is crucial for the structure elucidation both manually or using an expert system. The fragmentation of the compound represented by a mass spectrum provides access to further detail regarding the molecular composition in terms of key molecular fragments present in the compound under study.

- **1D NMR spectra** contain information about atoms included in the structure in terms of their electronic environments, their proximity to each other both in terms of skeletal connections and through-space interactions, as well as details regarding internal barriers of rotation.
- **2D NMR spectra** can provide information about both through-bond and through-space interactions between atoms and are the most informative, especially when multiple types of 2D NMR spectra are acquired and analyzed in parallel.

Data preparation is a key part of both manual and automated structure elucidation. Almost any error made during this procedure can lead to erroneous structures being derived as a result of the elucidation process. For input into Structure Elucidator, data preparation should therefore be done as carefully as possible. Data preparation consists of two main steps—the determination of an atom list and the determination of the connectivities between atoms. Algorithms for structure generation can automatically correct inconsistency and some errors in connectivities between atoms but it is almost impossible to rectify mistakes in the list of atoms. A dialog window **Spectrum Parameters** is used to specify the main spectrometer parameters which are set for corresponding spectrum registration.

The chemist has the possibility to postulate values of chemical shift **User Defined Tolerance** (ppm) for the F1 and F2 axes. These parameters significantly influence the problem complexity (size of the output file, time of structure generation, etc.).

Most procedures described in this chapter will be illustrated using the example of a small molecule, gymnopalynes [1], which contains carbon atoms of diverse properties and several heteroatoms. Gymnopalynes has the molecular formula $C_{12}H_7O_2Cl$. Its structure **2.1** with the assigned ^{13}C chemical shifts is shown below.

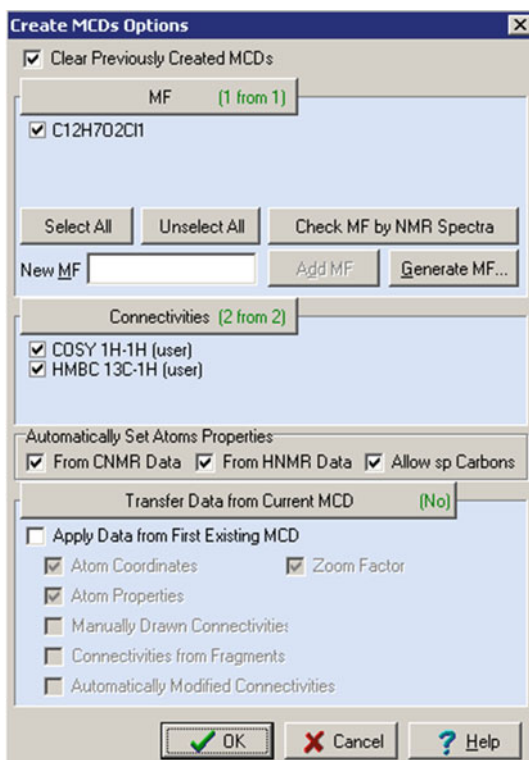


2.1.2 Molecular Formula

The structure generation algorithm requires a list of atoms as initial data input. The molecular formula (MF) is a compact representation of this list and is an absolute requirement in order to perform structure elucidation. Usually a monoisotopic

(accurate) mass is used to determine the molecular formula and is determined by analyzing the mass spectrum for the mass (m/z) of the molecular ion. It should be noted that a molecular ion peak is not always present in the mass spectrum. This is more common in electron impact ionization spectra but spectra obtained using other (more “mild”) ionization methods, for example, electrospray, usually do contain the molecular ion peak. Mass spectra obtained using the positive ion electrospray ionization (ESI) method contain the adduct peak of the protonated molecular ion. In addition to the protonated ion, usually called $[M+H]^+$, other adducts include ions such as Na^+ , K^+ or NH_4^+ , denoted as $[M+Na]^+$, $[M+K]^+$, and $[M+NH_4]^+$. This information should be taken into account when generating molecular formulae, i.e., the monoisotopic mass should be corrected as appropriate. Structure Elucidator is supplied with the Molecular Formula Generator which allows for the generation of all molecular formulae corresponding to a given molecular mass and postulated mass tolerance tol_m . To get to the **Molecular Formula Generator** it is necessary first to activate the command **Structure Elucidation/Create Molecular Connectivity Diagram**, as a result of which the dialog window **Create MCDs Options** will appear (Fig. 2.1). To open the dialog window **Molecular Formula Generator** (Fig. 2.2) it is necessary to click on the key **Generate MF**.

Fig. 2.1 The dialog window **Create MCDs Options**



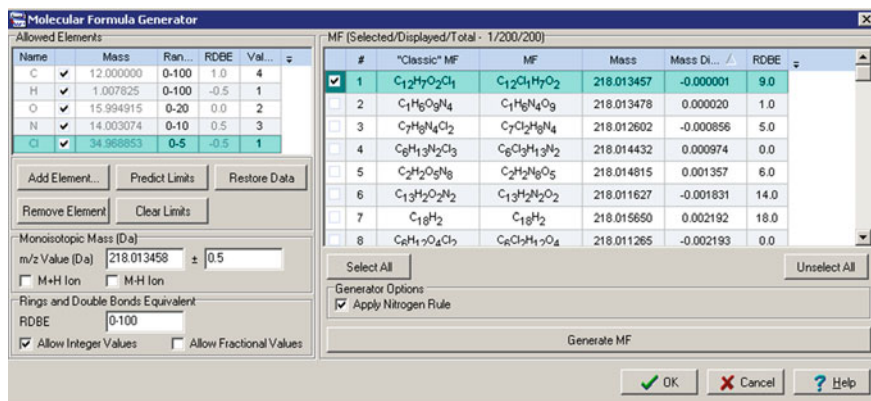


Fig. 2.2 The dialog window of the **Molecular Formula Generator**

The top of the molecular formulae list corresponding to the molecular ion of gymnopylyne $m/z = 218.013458$ at $tol_m = 0.5$ Da are shown in the right part of the window. In the left upper corner, the chemical elements assumed to be present in a molecule are shown along with the limits of the numbers of corresponding atoms allowed by user. The limits are postulated by the user on the basis of data extracted from 1D NMR and IR spectra, as well as by taking into account the pattern of the molecular ion cluster. In particular the number of signals in the ^{13}C NMR spectrum, the values of the integrals in the ^1H NMR spectrum and the characteristic IR absorption bands observed in the region of $3,700\text{--}1,300\text{ cm}^{-1}$ serve for this goal. If no constraints regarding the number of chemical elements can be imposed, the default settings are as follows: C (0–100), H (0–100), O (0–20), N (0–10). In the left lower part of the window the monoisotopic mass and a tolerance tol_m are set. The possible limits for the RDBE value (Rings and Double Bonds Equivalent) can be input in the relevant field (0–100 as default). If chemical ionization is used to acquire the mass spectrum then the check boxes **M+H Ion** or **M–H Ion** are selected. To use the “nitrogen rule” during molecular formulae generation, the check box **Apply Nitrogen Rule** should be selected. The generation of the molecular formulae is started by clicking on the **Generate MF** key. Generated molecular formulae are displayed in the right part of the window.

The number of molecular formulae corresponding to a given monoisotopic mass can vary depending on the accuracy of the mass determination and the possible elemental composition. For example, 200 molecular formulae correspond to the monoisotopic mass of gymnopylyne (218.013458 Da) if the accuracy of the mass determination is within 0.5 Da and only the elements C, H, N, O, and Cl are allowed. As shown in Fig. 1.17 the molecular formula of gymnopylyne is on the top of the list (#1) because generated molecular formulae are automatically ranked in increasing order of differences. If the number of carbon atoms is restricted to 12, the corresponding number of peaks in the 1D NMR carbon spectrum, then the number of molecular formulae is reduced to 9. A similar decrease in the number of potential

formulae occurs when the accuracy of mass determination increases. When the limits of carbon atom numbers are set as the default and the accuracy is 0.05 Da then the number of molecular formulae is 107. However, when measured to an accuracy of 0.005 Da the number of formulae decreases to 8 and two molecular formulae (a true one and unrealistic formula $C_1H_6N_9O_4$) can be found when the accuracy is 0.0005 Da. Only one and correct molecular formula was generated when the tolerance was set to 0.000005. In practice both methods, the restriction of elemental composition using other data and increasing mass accuracy, are used simultaneously to identify a single molecular formula in most cases. In some relatively rare cases, when unambiguous determination of a molecular formula is impossible, the structure elucidation process can be run several times using different molecular formulae.

2.1.3 Forming the Molecular Connectivity Diagram

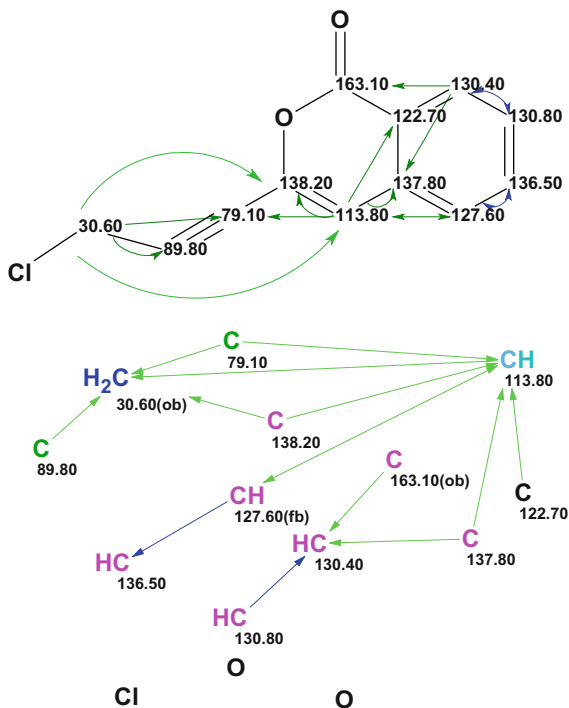
To provide a complete and clear pattern of the properties of the skeletal atoms and the connectivities between them the program places skeletal atoms together with hydrogen atoms attached to skeletal atoms (CH_3 , CH_2 , CH , and C groups, as well as OH and NH if identified by the user from 1H NMR and 2D NMR spectra) in a display. This pattern is referred to as the Molecular Connectivity Diagram (MCD). As mentioned above, to create the MCD it is necessary to activate the command **Structure Elucidation/Create Molecular Connectivity Diagram**. The dialog window **Create MCDs Options** (Fig. 2.1) provides for execution of the following functions:

- Use one MF or a set of possible MFs for creating the MCDs. If n formulae are selected, then n MCDs will be created.
- Input a new molecular formula directly in the field **New MF**.
- Select the types of 2D NMR data input into the program which will be used for MCD creation.
- Use the Atom Property Correlation Table (APCT) for atom property setting from the ^{13}C and 1H NMR data.
- Transfer data from the first existing MCD to a new MCD which is created by the user. Transferable properties are set using a selection of corresponding check boxes.

The MCD created from the 1H , ^{13}C , HMBC, and COSY spectra of gymnopalynes is shown in Fig. 2.3. A copy of the MCD is created simultaneously in the dialog window **Auto MCD**. All changes made in the Tables of Data are automatically transferred into the **Auto MCD** window. This window allows viewing of the initial MCD as the User MCD is edited by the user.

HMBC and COSY connectivities are shown in the MCD as “fuzzy” subgraphs (fragments) connecting carbon atoms and/or carbon and nitrogen atoms by arrows or lines when the corresponding 1H - 1H COSY and ^{15}N HMBC data are available.

Fig. 2.3 The MCD created from ^1H , ^{13}C , HSQC, HMBC, and COSY spectra of gymnopalynes whose structure along with 2D NMR connectivities is shown in the upper segment of the picture. Atom properties are artificially adjusted to the structure of gymnopalynes for illustrating different conventional signs



If known then the different possible distances between the connected atoms are marked on the MCDs with specific colors. In accordance with the main axioms (see Sect. 1.2.2) the lengths of the COSY and HMBC connectivities are taken by *default* (standard connectivities) to be 1 and 1–2 bonds correspondingly, this indicating the number of bonds between the *skeletal* atoms. In StrucEluc the standard length HMBC connectivities are colored *green* and the COSY connectivities are colored *blue* (see Fig. 2.3). If nonstandard connectivities are detected in the 2D NMR data by the chemist or by the program then they are marked in *violet* color. Ambiguous correlations (those which have ambiguous lengths) are distinguished by *dotted lines*. This scheme permits “at-a-glance” knowledge of whether two atoms are linked by exactly one bond, exactly two bonds, or a specific range of bonds. Information from different 2D NMR experiments can be viewed or suppressed by clicking the appropriate buttons on the toolbar.

The program analyzes the ^{13}C and ^1H chemical shifts of the CH_n groups ($n = 0\text{--}3$) and automatically sets, if possible, the parameters (atom properties) to show the most probable hybridization and the possible heteroatom neighborhood for each carbon atom. To carry out this procedure special Atom Property Correlation Tables (see Sect. 1.3.3.4) are used. The relationships to neighboring heteroatoms are marked as “forbidden” (*fb*), “at least one”, “at least two”, “at least three”, “four”, and “not defined” (*nd*). The possible states of atom hybridization are designated as sp^3 ,

sp^2 , sp , *not sp*, and “not defined”. To ease the visual recognition of the type of hybridization of a given atom each type is marked in its own specific color: sp^3 —blue, sp^2 —violet, undefined (sp^3 or sp^2)—light blue, sp —green (see Fig. 2.3). It is essential to note that both ^{13}C and ^1H NMR chemical shifts are taken into account by the program when setting the atom parameters. These descriptors for the carbon atoms allow the system to analyze 2D NMR data and to efficiently apply constraints during the process of structure generation.

If a distinct multiplet is observed in the ^1H NMR spectrum from a structural block $(\text{C}_i)\text{H}_n$ then the total number of hydrogen atoms attached to carbons adjacent to the (C_i) carbon is set. This property is determined by the chemist after visual analysis of the ^1H spectrum pattern and after taking into account the coupling constants (if measured). The atom properties should be set and edited with great caution because an erroneous assumption (a wrong “axiom”) leads to the exclusion of the correct structure from the output file. All structural constraints presented in the MCD are used during structure generation. Figure 2.4 shows a window where all properties of a particular CH_2 group are presented as an example, while Fig. 2.5 displays the pull-down menus for setting the possibility of neighboring with a heteroatom and atom hybridization.

Fig. 2.4 A window showing an example of setting the properties for a CH_2 group

Edit Properties of Atom # 1

Number of Current Atom = 1

<< Previous (15) Next (2) >>

Assigned Shift(s)

Atom's NMR Shift (nucleus 13C, shifts from -10 to 280 ppm)

Experimental 30.6 ± 0 Clear

Calculated 49.64 ± 0.000

Atom's QM calculated NMR Shift (ppm)

QM (GIAO/DFT) ± Clear

Attached Hydrogen's NMR Shift(s) (-2 - 20 ppm)

1st Experimental 4.55 ± 0 Clear

Calculated 3.02 ± 0.000

2nd Experimental ± Clear

Calculated n/a ± n/a

Atom Properties

Connection with Heteroatoms at least one

Number of Hydrogens on Neighbor Atoms 0

Hybridization State sp3

Charge 0

Valency not defined

Allow Non-default Valences

OK Cancel Help

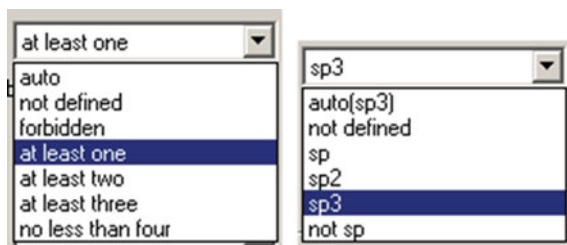


Fig. 2.5 The pull-down menus for setting the possibility of neighboring with a heteroatom (*left*) and atom hybridization (*right*)

Here a carbon atom with a chemical shift of 30.6 ppm is in a sp^3 hybridization state, and its connection with a heteroatom is obligatory. Since the signal at 4.55 ppm is a distinct singlet in the ^1H NMR spectrum the number of hydrogens attached to the carbon atoms closest to the C (30.6) carbon is set by the user to be equal to zero. The latter constraints speed up the structure generation process significantly because the generation of structures where this constraint is violated will be suppressed. The chemist can then edit these parameters using other available information. For this purpose, a set of buttons is shown on the MCD toolbar. The functions of these buttons are explained by screen tips and are intuitively clear. For example, if the molecule belongs to the CHNO class and the sp^2 hybridized carbon atom C(163.1) is marked by the user as sp^2 at least two then the system will only generate (if possible) O–C=O, N–C=O, O–C=N, and O–C \equiv N fragments on the basis of that atom. The chemist is also offered the opportunity to draw bonds of any multiplicity between the atoms to introduce suggested fragments into the process and to set some proposed functional groups (for instance C=O, O–C=O, C \equiv C, etc.). If a molecule contains heteroatoms and there are free H atoms displayed on the MCD, then O–H, N–H, NH₂, etc., groups may also be drawn in. This provides a quick and intuitive mechanism for entering structural information evident from the ^1H NMR and/or IR/Raman spectra. Lengths of connectivities can also be edited by drawing connectivities of definite lengths between selected atoms. In addition, a forbidden connectivity between a pair of atoms may be drawn on the MCD. All structural constraints presented in the MCD are used during the structure generation process. Edits of the MCD are carried out easily using the toolbar where all commands are intuitively clear and supplied with screen tips.

2.1.4 Checking the MCD for Consistency

As we will see later Structure Elucidator is capable of solving complicated problems if the spectral data are free of contradictions. The system is adjusted by default to account for the coupling constants $^{2-3}J_{\text{HH}}$ and $^{2-3}J_{\text{CH}}$ which are common for the corresponding COSY and HMBC correlations (referred to as “standard” correlations

in Sect. 1.2.2) and contradictions will appear when at least one correlation of >3 bonds results in a response in the 2D NMR data. Although the intensities of the 2D NMR peaks corresponding to “nonstandard” correlations are, in general, somewhat weaker than those corresponding to the standard correlations, the origin of both kinds of peaks is difficult to distinguish. Despite recent developments to aid in the identification of the correlation lengths [2–5] there is presently no routine NMR technique that is capable of distinguishing couplings of different lengths in a reliable fashion. Therefore, the development of theoretical methods for 2D NMR data analysis that identifies the presence of “nonstandard” correlations is of considerable importance.

The StrucEluc system is supplied with algorithms and programs [6] that are able to detect the presence of nonstandard correlations in 2D NMR data in the majority of cases. Algorithms that help to remove contradictions by lengthening certain connectivities have been delivered. A more general method to overcome the presence of contradictions in 2D NMR data, referred to as Fuzzy Structure Generation (FSG) (see Sect. 2.3), was also developed [7] and implemented into StrucEluc. In any case, the first step of structure elucidation using StrucEluc is to check the MCD for the presence or absence of contradictions in the 2D NMR data, i.e., to check data for consistency. The data checking algorithm is based on logical analysis of the full set of connectivities derived from the available 2D NMR spectra. The algorithm is sophisticated and is based on utilizing a method of logical proof by reduction *ad absurdum*. For instance, an indication of the presence of contradictions in two-dimensional data can serve a conclusion: the data could be considered consistent only in those cases where the valence of at least one carbon atom was assumed to be five or six, which is impossible. Because the algorithm is based on some heuristic statements it gives no guarantee that contradictions will be detected in any case. Experience has shown that analysis is successful in approximately 90 % of those problems where nonstandard connectivities existed in the data.

For MCD checking the command **Structure Elucidation/Check Current MCD...** is activated in the menu **Structure Elucidation** (Fig. 2.6).

As a result the dialog window **Check MCDs Options** (Fig. 2.7) is opened. Typical options which are used for the first program run are displayed in Fig. 2.7.

The options should correspond to the conditions and assumptions postulated by the user as being true during structure generation. If the check box **Automatically Resolve Contradictions** is selected the program will try to elongate all connectivities emanating from “suspicious” atoms by one bond. The relevance of the other check boxes is easily interpreted, but some explanations are necessary. The Atom Property Correlation Table (APCT) is commonly used for automatic atom property setting with “standard intervals” as shown in Fig. 2.7. The “wide intervals” can also be selected by the user or, if necessary, the APCT may be switched off by selecting the option “not used” (see Fig. 2.8, left part). For 2D NMR spectrum processing the option **Real Spectrum** is selected when real problems are solved (Fig. 2.8, right part).

Fig. 2.6 The **Structure Elucidation** menu



Fig. 2.7 The dialog window for **Check MCDs Options**

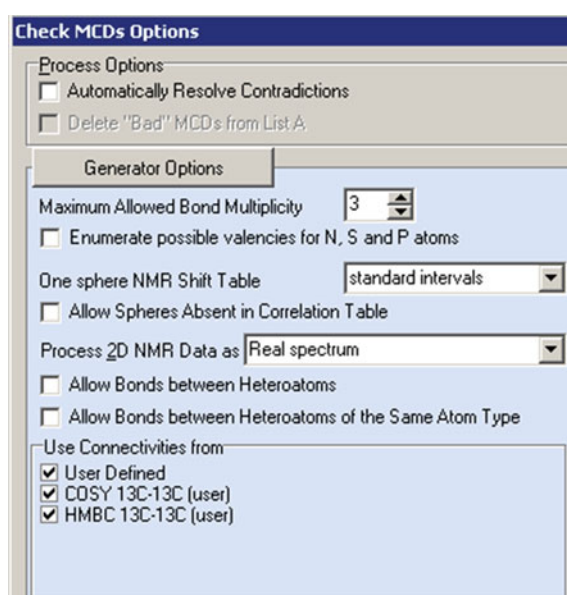




Fig. 2.8 The options of APCT (*left*) and possible selections for the field **Process 2D NMR** (*right*)

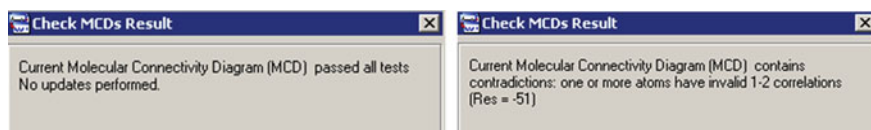


Fig. 2.9 Check MCD Results messages. The *left* message is delivered by the program if no contradictions were detected in 2D NMR data. If contradictions were detected the *right* message appears

When MCD checking is completed (it usually takes 1–3 s) the program displays a message containing information about the presence or absence of contradictions in the 2D NMR data. Examples of **Check MCD Results** messages are shown in Fig. 2.9.

When pressing on the key **More...** the user can see information about “suspicious” atoms and program suggestions about the minimum number of nonstandard connectivities. If no contradictions were detected then Strict Structure Generation can be performed, otherwise some different mode of the structure elucidation software is utilized (see Sect. 2.3).

A situation can be realized, however (see examples in Chap. 5), when generation of structures corresponding to all of the defined connectivities, including the non-standard ones, is *possible*. Generation is possible because the program fails to identify the nonstandard connectivities. This leads to either an *invalid* solution or to a valid solution in which chemical shift assignment for the right structure is incorrect. Therefore it can happen, even after seemingly successful checking of the data for contradictions, that nonstandard connectivities will nevertheless remain unnoticed by the algorithm. Their presence can only be identified a posteriori in indirect ways, for which the following conditions exist:

- large value(s) for the ^{13}C NMR spectral deviations calculated for the most probable structure(s);
- inconsistencies between the most probable structure and additional experimental data (for instance, NOESY, ROESY, etc.);
- disagreement between the chemical shifts and multiplicities of the experimental and calculated data of ^1H NMR spectra;
- the structures contradict IR/Raman correlations and/or interpretation of a mass spectrum.

The nonstandard connectivities that do not prevent the structure from being built are termed *implicit* nonstandard connectivities.

As the identification of implicit nonstandard connectivities cannot be guaranteed, since any of the given connectivities may be nonstandard, a method for removing nonstandard connectivities was developed which in the majority of cases allows for the identification of a valid solution even in this situation. As the method will be described in detail in Sect. 2.3, here we will explain only the main idea of the approach allowing the problem of implicit nonstandard connectivities to be circumvented.

Some connectivities are declared, as a series, as *suspicious* and are lengthened or eliminated. Each time structure generation is initiated with a renewed connectivity set. This process is termed as *Fuzzy Structure Generation* (FSG).

Let n be the total number of connectivities in the 2D NMR data and m be the number of connectivities that are suggested to be nonstandard. In this case it is necessary to consider $N = \binom{n}{m} = \frac{n!}{m!(n-m)!}$ different sets of m connectivities that will be declared as suspicious. If all N combinations of connectivities were used for structure generation, then the calculation time would increase dramatically as the number of tasks resulting in structure generation sharply increases with the rise in the m value.

Declaring members of the connectivity sets to be suspicious is used to search for atoms and pairs of atoms with nonstandard connectivities as well as for the direct determination of the presence of nonstandard connectivities. In these cases the program usually lengthens or deletes all connectivities belonging to the atoms selected during data analysis. In so doing, both nonstandard and standard connectivities are deliberately lengthened or deleted, which correspondingly leads to an increase in the number of structures generated. If only the definite connectivities (related to atoms for which the presence of nonstandard connectivities are revealed), are lengthened or deleted, then the number of generated structures will be considerably lower. The methodology and strategy of FSG will be discussed in Sect. 2.3.

2.2 Modes of the Structure Generation

In this section we will consider the main modes of the StrucEluc system. The program is capable of elucidating the chemical structure of much larger molecules, up to a mass of 1,500 amu to date, and containing more than 100 skeletal atoms. Typically, this task is accomplished from the analysis of 2D NMR spectral data. In general the system has been designed to elucidate structures containing up to 250 skeletal atoms. The capabilities of the *StrucEluc* system in terms of general utility as a tool for the structure elucidation of complex molecules, especially natural

products, has been demonstrated in many publications [8–12] (see reviews [11, 13] and the monograph [14]).

It should be emphasized that a large number of problems can be solved using only a molecular formula, heteronuclear (HSQC/HMQC, HMBC, or COLOC) and homonuclear (H–H COSY) 2D NMR correlations and without using any additional structural information. In this mode of operation, referred to as the *Common* mode, the system creates connectivities from the spectral data and generates all possible structures in accordance with the default settings for the number of intervening bonds between corresponding skeletal atoms and with atom properties including the state of hybridization and the possibility of taking neighboring heteroatoms into account.

However, it turned out that there were problems that could not be solved or proved to be very time-consuming due to a lack of information in the 2D NMR data (see, for example, Sect. 4.34). In these cases it proved necessary to introduce additional structural information, if available, to facilitate the elucidation process. In the real world, it is common for a chemist or spectroscopist faced with elucidating a structure to have prior knowledge of reaction components in a synthesis, knowledge of the class of compounds that may have been isolated, or even hypothetical structures for validation rather than full elucidation from no information.

It has been shown [14] that the utilization of molecular fragments found from the system knowledge base, or potential substructures proposed by the chemist, can be helpful to circumvent the difficulties. Such a *fragment approach* has been used in a number of first-generation expert systems based on correlation tables containing substructures and their associated characteristic intervals for specific spectral features. In contrast, StrucEluc employs a database containing substructures and their associated ^{13}C NMR subspectra. At present the StrucEluc database contains more than 290,000 chemical structures and more than two million substructures. The database continues to grow as further literature data is added. The value of including substructures directly into the elucidation process is that a fragment, considered as a macro atom, can absorb a significant number of the skeletal atoms and leads to a reduction in the complexity of the problem. This results in acceleration of the structure generation procedure, which is typically the most time-consuming stage of the structure elucidation process.

Nevertheless, in those cases when 2D NMR data is employed, the usage of molecular fragments is hampered by the fact that *all* carbon atoms existing in a fragment utilized in solving the problem *must* be supplied with chemical shifts. Moreover, the values of these chemical shifts must be as close as possible to the observed values for the atoms of the corresponding fragments in the experimental ^{13}C NMR spectrum of the unknown under study. Particularly, the approximate chemical shift values of carbon atoms can be found using ACD/C NMR Predictor. Before structure generation all approximate chemical shift values set for fragment carbon atoms should be replaced by experimental chemical shifts closest to those ascribed to fragment atoms. The reason for this requirement is obvious: the utilization of 2D NMR correlations implies the possibility to use only observed experimental chemical shifts. The accommodation of one or more fragments within

a set of connectivities derived from the 2D NMR data is a problem that requires the development of new algorithms. In this chapter we will discuss different strategies for applying the StrucEluc system. Depending on the initial data available, and the complexity of the molecule being analyzed, the system offers a wide range of methods for solving a problem.

2.2.1 The “Common” 2D NMR Mode

The *StrucEluc* system is based on a number of programs developed for elucidating a molecular structure from a combination of 2D NMR spectra. The most typical combination providing the basis for structure determination includes H–H COSY, HSQC/HMQC, and HMBC. The StrucEluc system also operates with additional 2D NMR methods: ROESY, NOESY, TOCSY, ADEQUATE, and INADEQUATE [15]. Other methods can also be used by the system through a flexible procedure that allows input and processing of experimental 2D NMR data.

Prior to the structure generation the MCD is checked for the presence of contradictions (see Sect. 2.1.4).

The data collected in connectivity tables and graphically presented as an MCD are used as the input information for the 2D structure generator. If MCD checking shows that the 2D NMR data are consistent then Strict Structure Generation is initiated. Structures are generated under constraints determined from the molecular formula, the MCD, and any additional constraints which may be introduced by the chemist. The structure generator is based on mathematical algorithms developed by Molodtsov, who enhanced them during the Structure Elucidator program development [6, 14].

In general, structure generation is initiated by the command **Structure Elucidation/Run CSB Generator** (CSB, Correlation Spectroscopy Based). Structure generation will be performed from *all* existing MCDs in series upon executing this command. The command **Structure Elucidation/Run SCB Generator from Current MCD** allows structure generation to be performed from one MCD selected by the chemist. Figure 2.10 shows the dialog window associated with the CSB Generator Options.

The meanings of the majority of options are intuitively clear, but some of them require explanation. **Estimate Generation Time Only** is selected if the initial data is fairly uncertain and it is desirable to provide an estimate of the generation time and the number of generated structures is expected to be manageable.

The options **Add Generation Results to User Notes** and **Save Project After Generation is Completed** are recommended to be selected if it becomes clear that structure generation will be time-consuming (for instance, if the program is left to work overnight).

A group of options **Use Connectivities from** allows the chemist utilizing different combinations of 2D NMR spectra during structure generation. For instance, the presence of 3–5 nonstandard connectivities of 4J length in the COSY data may

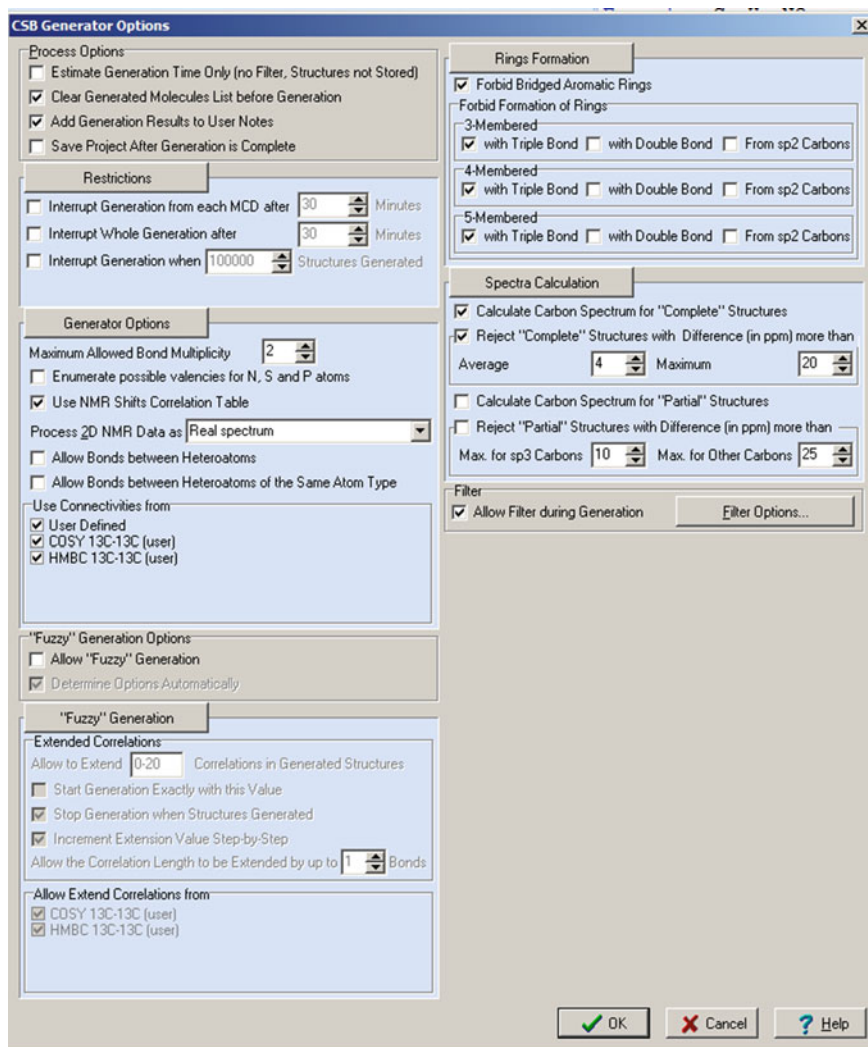


Fig. 2.10 CSB Generator Options

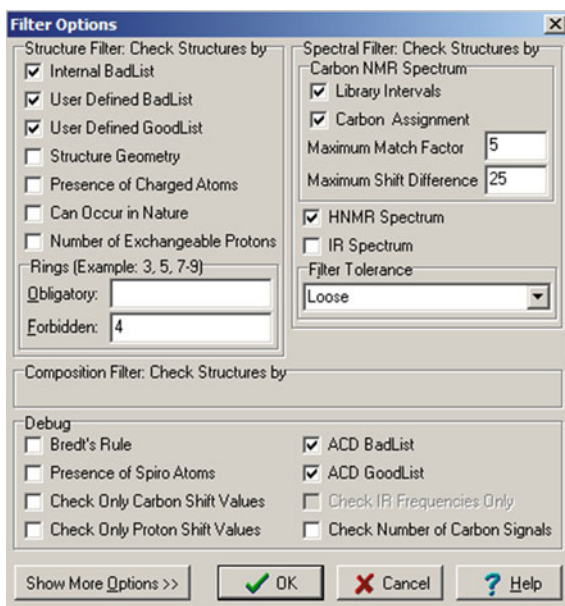
markedly elongate the time of FSG which makes it more advantageous to deselect the COSY check box and proceed with structure generation from the HMBC data only.

The options **Spectra Calculation** and **Allow Filtering during Generation** are used in combination and play an important role in the structure elucidation process. The purpose of these options is to reject generated structures for which predicted chemical shifts differ dramatically from the experimental shifts. On the level of the spectral filter, a *rough* spectrum prediction is realized by intervals of characteristic spectral features. If at least one ^{13}C or ^1H NMR chemical shift assigned to an atom

of a generated structure does not meet the corresponding spectral interval, the structure is rejected by the filter. As it is impossible to take into account the influence of all conceivable environments of a given atom in a molecule on an atom's chemical shift sometimes the filter can reject a correct structure. Therefore, to minimize risk of the correct structure loss the **Spectrum Calculation** option can be used. For this aim it is necessary to select the commands **Calculate Carbon Spectrum for Complete Structures** and **Reject Complete Structures with Difference (in ppm) more than** and indicate the average and maximum deviations which will be used as thresholds for the current structure rejection. ^{13}C chemical shift prediction is performed by the *Incremental* approach which shows a high speed of calculation ($\sim 30,000$ chemical shifts/s). In reality, it is practically impossible to notice if the ^{13}C chemical shift calculation was used or not used during the structure generation. Experience has shown that the thresholds $d = 4\text{--}5$ and $d_{\text{max}} = 20\text{--}25$ ppm are optimal values providing an output structural file of a manageable size. When the option **Spectrum Calculation** is used, in the dialog window **Filter Options** a check box **Carbon Assignment** is automatically selected and the fields **Maximum Match Factor** and **Maximum Shift Difference** are automatically filled in as shown in Fig. 2.11.

Thus, when ^{13}C spectrum calculation is activated the filter can be used both as a mechanism intended only for structure rejection in correspondence with the calculated ^{13}C NMR chemical shifts and as a “standalone” spectral and structural filter. If the filter is used *only* as a facility of the ^{13}C chemical shift calculation procedure then the check boxes **Library Intervals** and **HNMR Spectrum** must be deselected, while all structural constraints shown in the left part of the dialog window can be

Fig. 2.11 Dialog box **Filter Options**



used at the same time. If the filter is used in a mode where **Spectrum Calculation** is disabled it is necessary to check if the option **Carbon Assignment** is also deselected, otherwise the filter will reject all structures because they have no deviations for comparison. It is expected that if **Spectrum Calculation** and spectral filtering are used simultaneously (check boxes **Library Intervals** and **HNMR Spectrum** are selected) the output structure file will be of minimum size.

A question arises: Which method of reducing the output structural file and rejecting deliberately invalid structures is optimal? We suppose that the sequence of operation which was traditional for all expert systems “Fragment Selection → Structure Generation → Structure Spectral Filtering” can be modified now. The Structure Spectral Filtering can be replaced by Spectrum Calculation and utilization of the *spectral filtering* only for rejection structures which do not satisfy the threshold criteria $d \leq 4-5$ and $d_{\max} \leq 20-25$ ppm. As a result only structures that went through this stage will be saved. In this case, checking structures by ^{13}C and ^1H NMR characteristic spectral intervals can be used as an additional aid for verification of the saved structures.

Generated structures can be inspected by clicking on the key **MOL** in the toolbar of the main window of the program. Then ^{13}C NMR spectrum prediction is performed for all structures included in the output file and the structures are ranked in order of increasing d value using the method described in detail in the following section. To perform the NMR spectrum calculation it is necessary to press on the key **Tools** on the Toolbar and select a spectrum that should be predicted in the drop-down menu (Fig. 2.12).

If the user wants to predict chemical shifts for several kinds of NMR spectra by different methods of spectrum calculation, the command **All Spectra...** should be activated. In the dialog window **Select Spectra to Calculate** (Fig. 2.13) the corresponding spectra are chosen by check box selections. As a default chemical shifts will be calculated for all structures in a given structure file, but a **Maximum Number of Processed Structures** can also be specified.

If the user has his own structural hypothesis the proposed structure can be drawn in the **PM** (Proposed Molecule) window and ^{13}C and ^1H chemical shift assignment can be performed (use the tool **Edit Atom Properties**). Then the system can be used for verification of the proposed structure and the associated ^{13}C and ^1H NMR signal assignments by all available two-dimensional NMR spectra. For this purpose, a command **Structure/Check by 2D NMR Data (1)** is activated. If any nonstandard connectivities are found, the program displays a textual message detailing the cause(s) of the conflict(s). At the same time all connectivities are shown on the structure in graphical form. To ease visual analysis, nonstandard connectivities are marked in red.

2.2.1.1 Selection of Preferable Structure

As previously mentioned, in general, selection of the preferable structure is reduced to NMR chemical shift prediction for structures included into the output file and

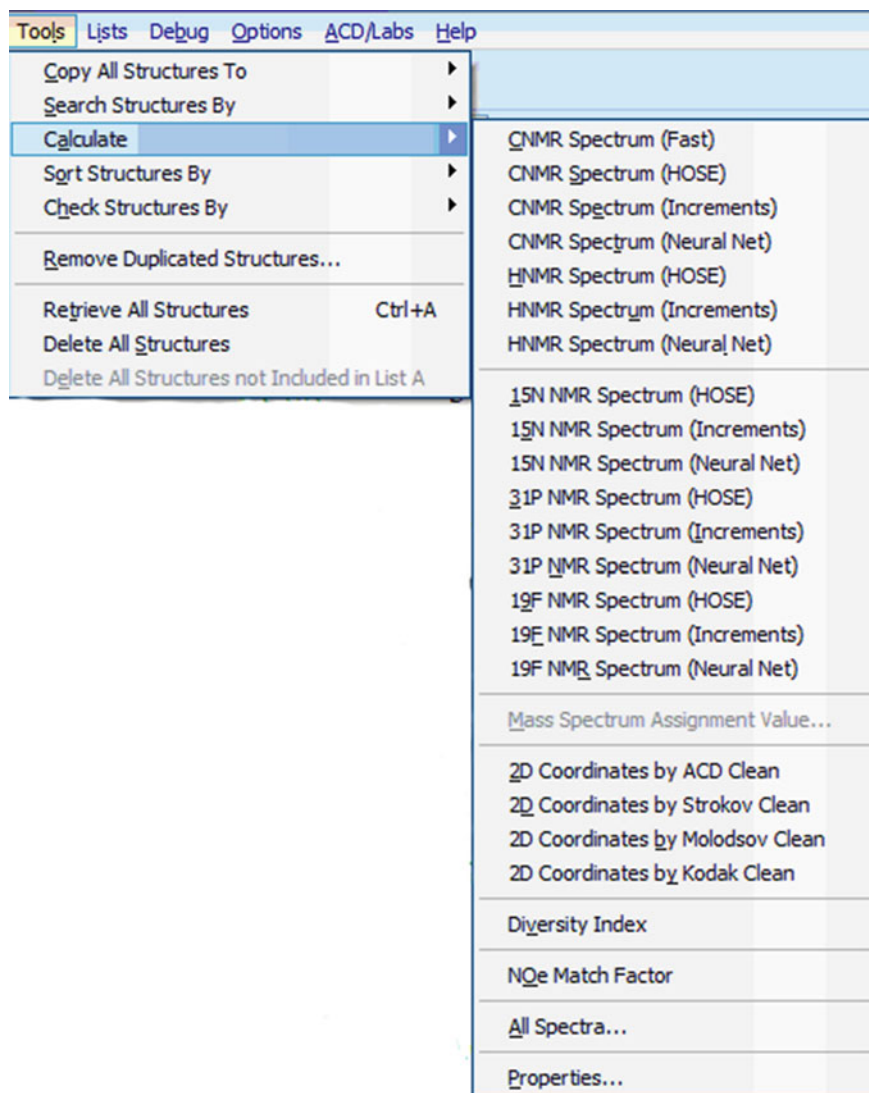
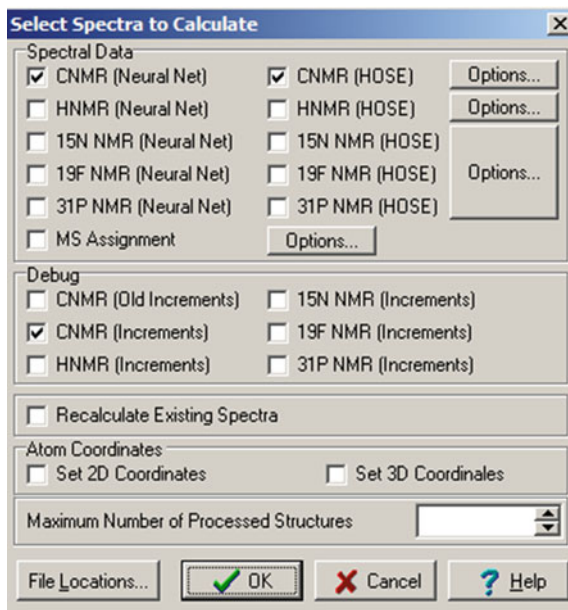


Fig. 2.12 Drop-down menu **Tools**

structure ranking in increasing order of deviation values. The StrucEluc system provides the following procedure for identifying the most probable structure in the output file.

First step ¹³C NMR spectra are predicted for all generated structures using an incremental method, the fast method, and d_l values, the average deviation of an experimental ¹³C NMR spectrum versus predicted chemical shifts, are calculated.

Fig. 2.13 The dialog window **Select Spectra to Calculate**




If ^{13}C chemical shift calculation is used during structure generation this step is skipped.

Second step Duplicate structures in the output file are deleted (**Tools\Remove Duplicate Structures...**, Fig. 2.12). Among the generated structures there are usually duplicates that differ from each other only in terms of the assignment of the chemical shifts to different carbon atoms. If this possibility is not appropriately considered when deleting isomorphous structures, then the structure with the correct assignment of the chemical shifts could conceivably be the deleted isomorphous structure. To avoid this eventuality, the system executes a special procedure for duplicate removal. For each duplicate family only the structure that has the minimum d_I value is retained in the file as “the best representative” of the family. After duplicates are removed, the structures are then ranked by the d_I value and sorted in ascending order (**Tools\Sort Structures by\C NMR Spectrum (Increments)**, Fig. 2.12). The smallest d_I value indicates the best match between the experimental and calculated spectra and this structure will therefore be the first in the list. Experience shows that the incremental calculation of ^{13}C NMR spectra and their subsequent ranking usually places the correct structure among the first several structures at the top of the list. Only in rare instances will the correct structure be listed below fifth place. Such a preliminary ranking of the big resulting files can help to reject thousands and tens of thousands of structures that are known to be unsuitable.

Third step ^{13}C chemical shift prediction is carried out using an NN algorithm and the structure file is ranked again with d_N deviations. If the resulting file is extremely large the calculations can be applied only to the first several thousand

structures (it will take several seconds). As a result of this step the preferable structure is selected with greater reliability.

If the initial structure file is not too large it is desirable to perform removal of duplicates after calculating d_I and d_N values: selection of the same best structure by both incremental and neural net approaches will raise the probability of obtaining the right structure with correct chemical shift assignment.

For the user's convenience the ranked file can be displayed in a **Tile** mode for which it is necessary to press the icon  to **Display Structure and Tile** in the Toolbar.



A right mouse click on the field where the list of structures is displayed leads to the appearance of a menu (Fig. 2.14). This menu contains intuitively understandable functions which allow the user to manage information associated with structures displayed in the **Tile** mode.

Fourth step During the fourth stage ^{13}C NMR spectra are usually calculated for the first 10–50 (sometimes up to 100) structures of the ranked file using a fragmental method based on the HOSE code approach (this procedure can take several minutes). The average deviation values between the experimental and calculated values (d_A) are found and the structures are again rank ordered. Subsequent ranking increases the probability of moving the correct structure to the first position in the list. For

Fig. 2.14 The menu for managing information related to the structures which are displayed in the **Tile** mode

Switch to Table View	
Copy to Editor	ENTER
Zoom In	
Zoom Out	
Select Current	Ins
Select All	Ctrl+A
Unselect All	
Invert Selection	
Set List from Selection	
Save Selection...	
Load Selection...	
Select Data...	Gray *
Select Font...	
SS Search Color Options...	

additional control over the correct choice of the output structure, the HOSE code-based proton chemical shifts can be predicted and displayed together with the corresponding deviation value d_H . A complex match factor $d_{\text{complex}} = d_N(\text{C}) + 10d_N(\text{H})$ can also be used for ranking the structures in the output file.

The position of the correct structure in the file determines its rank depending on the type of ranking parameter, i.e., d_N , d_A , d_I , d_H , or d_{complex} correspondingly. The “rates” of the correct structure in the ranked file are denoted as r_N , r_A , r_I , and r_H . If the correct structure is the first in the list ranked by d_A values, then $r_A = 1$. As a rule, the final structural ranking is carried out according to increasing d_A and d_N values, while magnitudes of the d_I and d_H parameters serve as secondary aids for estimating the reliability of the correct structure selection. The accuracy of chemical shift prediction for each carbon atom can be evaluated visually by pressing the toolbar button **Show\Hide Carbon Assignment**. The accuracy is marked by colored circles on the atoms, while the following colors are used: *green*—the difference Δ between the experimental and predicted chemical shifts is not higher than 3 ppm ($\Delta < 3$ ppm), *yellow*— $3 < \Delta < 15$ ppm, *red*— $\Delta > 15$ ppm. All kinds of information related to structures can be visualized using the following icons on the toolbar:  and . The first of them allows the experimental and predicted ^1H , ^{13}C , and ^{15}N NMR chemical shifts to be displayed as well as different representations of atom numbering (Fig. 2.15, left). The second button is used to display the kinds of connectivities that are selected by the user (Fig. 2.15, right).

The top structures or selected structures displayed in the **Tile** mode can be copied to the ChemSketch window by the command **File\Create Report\List of Structures** as shown in Fig. 2.16.

When the first and second ranked structures contain markedly differing structural elements, then the prediction of the MS match factor (m_i , where i is the position of a structure in the ranked file) may also be useful for confirmation of the preferable structure. For this purpose, it is necessary to activate the command **Mass Spectrum Assignment Value**, Fig. 2.12. The system utilizes a routine that is capable of calculating the percentage of peaks in the experimental MS spectrum that can be interpreted on the basis of a given structure. The calculation of the MS match factor is relatively time-consuming, so it is worth using it only in those cases when the difference $\Delta_{(2-1)} = d_A(2) - d_A(1)$ is small. Here $d_A(1)$ and $d_A(2)$ represent the deviations corresponding to the first and second structures in the ranked file.

In ambiguous cases it may be useful to display the calculated ^1H NMR spectra in graphical form. Also, to facilitate structure analysis in the output file, the *StrucEluc* system is supplied with a feature that calculates structural similarity coefficients (**Structure\Similarity Search in\Generated Molecules**). In this way if the investigator has an idea of the class of structure under investigation he can use this structure as an input to allow rank ordering relative to the structural similarity of the results file.

Numeration	
$A_{\#}$	Show Atom Numbers
^{13}C	Show Experimental Labels
^{13}C	Show Experimental Numbers
1H	Show Experimental Labels
1H	Show Experimental Numbers

13C NMR Peak Values	
<input checked="" type="checkbox"/>	^{13}C Experimental
<input type="checkbox"/>	^{13}C Calculated (HOSE)
<input type="checkbox"/>	^{13}C Calculated (Fast)
<input type="checkbox"/>	^{13}C Calculated (Increments)
<input type="checkbox"/>	^{13}C Calculated (Neural Net)
<input type="checkbox"/>	^{13}C Calculated (QM DFT)

1H NMR Peak Values	
<input type="checkbox"/>	1H Experimental
<input type="checkbox"/>	1H Calculated (HOSE)
<input type="checkbox"/>	1H Calculated (Increments)
<input type="checkbox"/>	1H Calculated (Neural Net)

15N NMR Peak Values	
<input type="checkbox"/>	^{15}N Experimental
<input type="checkbox"/>	^{15}N Calculated (HOSE)
<input type="checkbox"/>	^{15}N Calculated (Increments)
<input type="checkbox"/>	^{15}N Calculated (Neural Net)

17O NMR Peak Values	
<input type="checkbox"/>	^{17}O Experimental

Filter by Nuclei	
Show For Active Spectrum	
<input checked="" type="checkbox"/>	1H - 1H Correlations
<input checked="" type="checkbox"/>	13C - 1H Correlations
<input checked="" type="checkbox"/>	15N - 1H Correlations
<input checked="" type="checkbox"/>	Other Correlations

Filter by Bond Distance	
<input checked="" type="checkbox"/>	2J Correlations
<input checked="" type="checkbox"/>	3J Correlations
<input checked="" type="checkbox"/>	4J and More

Filter by Correlation Type	
<input checked="" type="checkbox"/>	Through Bonds
<input type="checkbox"/>	Through Space

Geminal Coupling	
<input type="checkbox"/>	Show Geminal 1H-1H Couplings

Ambiguous Correlations	
<input checked="" type="checkbox"/>	Show Outranged Ambiguous Correlations

Fig. 2.15 *Left* visualizing experimental and predicted 1H , ^{13}C , and ^{15}N NMR chemical shifts and different types of atom numbering. *Right* visualizing connectivities of different origin and different length

When the best structure is selected the following question can be posed: which structures would be generated if all theoretically possible HMBC and/or COSY correlations (types of 2D NMR spectra are selected by the user) were observed in the experimental spectra? The command **Structure\Create Project from Structure** is used to provide the answer to this question. The program creates a new

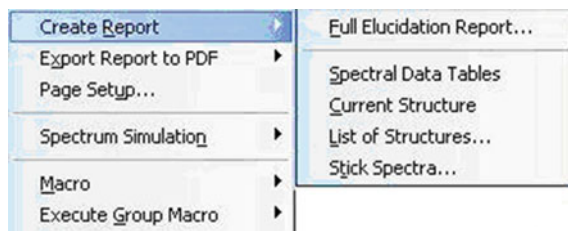


Fig. 2.16 A part of the **File** menu containing commands which are utilized for copying a List of Structures to ChemSketch

project with the MCD containing all theoretically possible connectivities from which structure generation can be performed.

2.2.2 Application of Fragments in Combination with 2D NMR Data

Computer-assisted structure elucidation using 2D NMR data is quite efficient for the elucidation of structures of complex organic molecules. However, if the structural restrictions imposed by the MCD are not sufficient for the generation of a reasonable number of possible structures within an appropriate time, it is to be expected that the utilization of molecular fragments can help facilitate the solving of the problem. Commonly appropriate fragments to aid in the solution of a problem can be found in the Structure Elucidator knowledge base. *The main advantage of these fragments is that all fragment carbon atoms are supplied with the ^{13}C NMR assignments obtained from the full structures that were used for creation of the fragment database-Fragment Library.*

The first step of the process is a fragment search of the Fragment Library which is initiated by the command **Structure Elucidation\Search Fragments by C NMR Spectrum** (Fig. 2.17). To change the default options the researcher can use the dialog window **Search Fragments by ^{13}C NMR Spectrum Options** (Fig. 2.17).

If the radio button **From Molecular Formula** is selected, all the Found Fragments will have molecular compositions not exceeding the Molecular Formula. Utilization of 2D NMR connectivities and usage of the filter during fragment search can optionally be activated by the user. When the fragment search process is completed the Found Fragments are displayed in the **Found Fragment** window (**View\Structures List\Found Fragments**, Fig. 2.18).

As a result of the fragment search a set of L Found Fragments is selected. The next step is the creation of MCDs using the found fragments (FF), for which the command **Structure Elucidation\Create MCDs Using Fragments** is activated (see Fig. 2.6), and then the dialog window **Create MCDs Options** is opened (Fig. 2.19). For the first run automatic determination of most options is allowed.

Fig. 2.17 Dialog window
Search Fragments by ^{13}C
NMR Spectrum Options

Search Options

Clear Found Fragments before Search

Spectral Data

Reject structures with Match Factor more than 5 ppm

Reject structures with number of signals less than 20 % in spectrum

Allow lack of signals in "full" structures: 5 signals

Allow excess of signals in structures: 2 signals

Allow excess only for quaternary carbons

Ignore peak intensity during Search

Composition

Composition

Check composition

From Molecular Formula From Defined Composition

Composition: C(0-12) H(0-7) O(0-2) Cl(0-1)

Monoisotopic Mass

Check Monoisotopic Mass

Monoisotopic Mass: 0.000-218.513 Tolerance (Da) 0.5

2D NMR Data

Use HSQC information (check chemical shifts of attached hydrogens)

Use HMBC and COSY information (check distance between chemical shifts)

Tolerances

	^{13}C	^1H	
Tolerance for "First Sphere" Atoms (no less than)	12	2	ppm
Tolerance for "Second Sphere" Atoms (no less than)	6	1	ppm
Minimum Possible Tolerance for All Atoms (no less than)	3	1	ppm
Maximum Possible Tolerance for All Atoms (no more than)	20	2	ppm

Filter

Allow Filter during Search Filter Options...

Search Databases

ACD Internal Full Structures Database

ACD Internal Fragments Database

Internal DBs

Add...

Remove

Up

Down

Spectral Data... OK Cancel Help

For MCD creation the selected number of FFs can be set either by the operator or by the program—automatically. The main idea of the algorithm that implements this procedure is as follows. The chemist defines the number of fragments, l ($l \leq L$), that will be used for MCD creation and sets an error, E , that defines the maximum difference allowed between the chemical shifts of the fragment carbons and the corresponding values observed in the experimental spectrum under study. The situation is common when several experimental chemical shifts are close to the

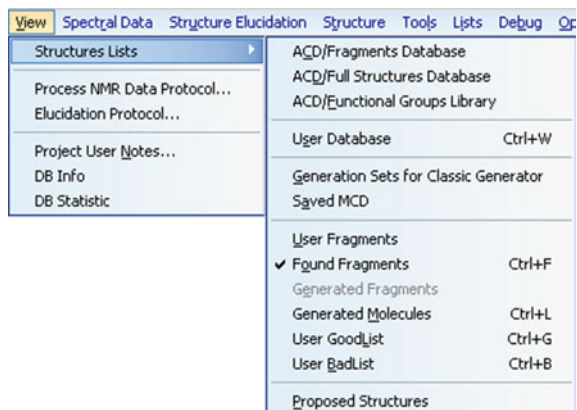


Fig. 2.18 A part of the **View** menu used for switching windows containing structural files common for Structure Elucidator

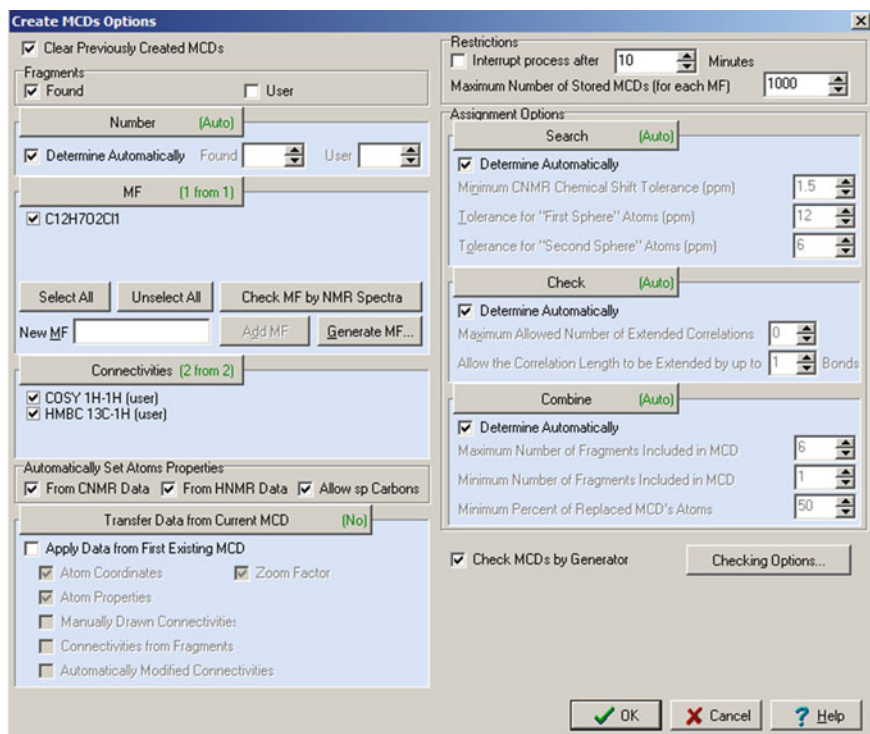


Fig. 2.19 The dialog box **Create MCDs Options**

chemical shift assigned to a given carbon atom of a fragment. It is important to note that both parameters, l and E , are closely interrelated and choosing the most efficient values may be a matter of trial and error.

The ^{13}C NMR subspectrum of each fragment is compared with all experimental chemical shifts. The number of hydrogen atoms attached to a carbon atom is taken into account during this process. Consider a fragment that contains n carbon atoms and an arbitrary atom C_i of the fragment has a chemical shift δ_i ($i = 1 \div n$) and multiplicity m_i . Suppose that the experimental chemical shifts $\delta_{i1}, \delta_{i2}, \dots, \delta_{iq}, \dots, \delta_{ip}$ meet the conditions $|\delta_i - \delta_{iq}| \leq E$ and $m_i = m_{iq}$. Then, all possibilities of substituting the δ_i values for the experimental values $\delta_{i1}, \delta_{i2}, \dots, \delta_{iq}, \dots, \delta_{ip}$ must be verified.

If the conditions $|\delta_i - \delta_{iq}| \leq E$ and $m_i = m_{iq}$ hold for all f carbon atoms, then the given fragment is recognized as a candidate for inclusion in the process of creating the MCD. If this condition does not hold then the fragment is excluded from consideration. The program also checks whether the carbon atom assignments correspond to the experimental chemical shift correlations comprising the skeletal atoms making up the fragment. The fragments that survive the test are then included in the set of *prospective* fragments.

The more the skeletal atoms “absorbed” by the fragments, the shorter is the process of structure elucidation. With this in mind an algorithm that combines the prospective fragments within one MCD was developed. To realize this procedure, all possible combinations of prospective fragments are searched and only combinations that are in agreement with the experimental 2D NMR correlations are chosen. The fragment combinations that pass this examination form a set of prospective fragment combinations. These fragments are then “projected” onto the MCDs together with any remaining free atoms. The user can then visually analyze these MCDs.

The total number of MCDs, n_{MCD} , depends on the following parameters which are defined by the user:

- L number of found fragments which will be used for the creation of MCDs ($l \leq L$);
- n_f the minimal number of fragments that must be present in each MCD;
- q the minimum percentage of all skeletal atoms that must be absorbed by the fragments present in each MCD.

In general the more the atoms that are “absorbed” by the fragments accepted by an MCD, the greater the likelihood that the process of structure generation from the given MCD will be more time efficient.

The speed of structure generation depends on the size of the molecular fragments. If the number of small fragments composing the MCD is large enough, then this will speed up the generation. Structure generation is much faster when the MCD comprises a small number of big fragments. Depending on the size of the molecule being analyzed and the size of fragments placed at the beginning of the ranked list of FFs, the n_f value is usually defined as a number from 1 to 4. The most efficient results are obtained if q is significant, generally 40–60 %. The dialog window which is used for setting the minimum and maximum numbers of fragments included into the created MCDs is presented in Fig. 2.20.

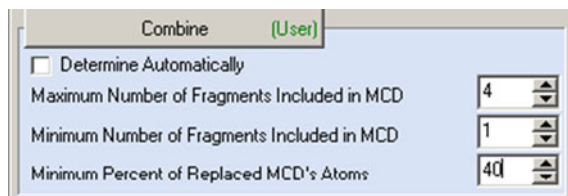


Fig. 2.20 The dialog window **Combine** which is used for setting the minimum and maximum number of fragments included into the created MCDs

The conclusion of all further verification procedures is a check of all produced MCDs for contradictions. The program offers an option that deletes all MCDs that are recognized as contradictory. The diagrams remaining after checking can be used in the structure generation process. The user has the opportunity to omit the connectivity verification because contradictory MCDs will be detected and rejected in the process of structure generation. Moreover, for the process of structure generation the user can select one or more MCDs that are attractive to the user who may have prior knowledge of a particular structure class or target structure. To alleviate having to choose a preferable MCD they are automatically ranked in order of the increasing number of free carbons. In this way it is possible to select a series of appropriate MCDs, starting from that ranked first.

The number of MCDs produced from a given set of fragments can be rather large (sometimes the n_{MCD} value is greater than one thousand). To provide for the possibility to edit a big set of MCDs a special procedure was elaborated, which allows one to transfer all changes made in the *first* MCD to all the MCD sets (**MCD Apply Properties to All MCDs**). In particular, it is possible to specify options which transfer atom coordinates, zoom factor, atom properties, manually drawn connectivities, and connectivities automatically modified during the MCD checking for presence of contradictions. This procedure essentially alleviates using a priori information in the fragment mode of structure elucidation.

In the process of analyzing a novel compound it is entirely possible that there will be no fragments in the database that will reduce the magnitude of the challenge. It is natural in such cases to expect that the introduction of user-defined fragments may help to form the MCDs. The main qualitative difference between a found fragment (FF), and a user fragment (UF), is that the FF already contains carbon atoms with assigned chemical shifts while the carbon atoms of the UF have no carbon chemical shift assignments. Two ways have been suggested to introduce UFs into the program:

- Calculate the carbon chemical shifts of the fragment using the HOSE code-based method (see Sect. 1.4.1.1);
- Search the KB for fragments that *comprise* the user fragment.

It is likely that fragments from at least one of the two sources would be available for use by the program.

2.2.2.1 Choice of E Value

In the process of MCD creation from fragments, the E value is of great importance since it markedly influences the result of applying the fragments. There are a number of principles governing the selection of the E value. As a rule, the smaller the value of E , the smaller the number of MCDs, n_{MCD} , created from FFs. The advantage of a small number of MCDs is of course that it can reduce the time for structure generation, t_g . At the same time, t_g is also a function of the *fragment dimensions*. Larger fragments generally shorten the structure generation process. However, if a fragment is large and correspondingly contains many assigned carbon atoms, then as a consequence it is not as likely that *all* carbon atoms, especially the terminal ones, of a large fragment will fit the experimental shifts thereby satisfying a narrow interval for $\pm E$. The program automatically sets the E value for terminal atoms equal to 12 ppm to account for this issue (Fig. 2.21).

Large fragments are the most useful but to utilize them in the structure elucidation process a large E value is frequently necessary. A large E value can correspondingly increase the n_{MCD} value. The optimal approach would be to set a large enough E value and select only those MCDs containing large fragments for the structure generation. This principle therefore justifies manual (user) or automatic rejection of MCDs containing small fragments. With testing it has been shown that the optimal program parameter controlling the minimum number of carbon atoms in the fragment used for MCD creating should be set to a value of five. Unfortunately, it is impossible to determine an optimal E that is valid for a diverse range of problems. The value of E should be optimized for each task by gradually increasing the E value starting from 1.5 ppm. This procedure can be performed manually or automatically during structure generation.

2.3 Nonstandard Correlations and Fuzzy Structure Generation

2.3.1 Challenge of Nonstandard Spectral Responses

CASE 2D NMR methodology can provide solutions for computer-assisted structure elucidation tasks in a reasonable time if the initial data (NMR, MS, IR, chemical

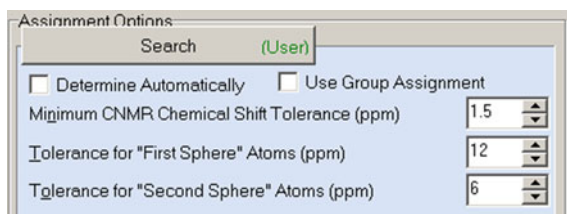


Fig. 2.21 Selection of ^{13}C chemical shift tolerances for creation of MCDs from FFs. Minimum C NMR chemical shift tolerance E is 1.5 ppm

assumptions, etc.) are *true, consistent, and complete*. The latter means that the number of observed 2D NMR correlations is large enough to sufficiently define the connectivities within a structure. If at least one of these conditions is violated the possibility to somehow find a correct solution to the problem decreases significantly. As mentioned earlier, methods to overcome the presence of contradictions in experimental data have been suggested. The corresponding algorithms have been developed and implemented in the StrucEluc system. In this chapter we will consider the different StrucEluc-based strategies for molecular structure elucidation in those situations when 2D NMR spectra contain nonstandard correlations (NSCs).

As a result of a series of computational experiments it has been shown that the program was capable of determining the *presence* of connectivities of nonstandard length in 90 % of all cases using the MCD checking procedure described in Sect. 2.1.4. These results are very encouraging since routine experimental methods guaranteeing the precise determination of COSY and HMBC connectivity lengths are not available. Knowledge of the presence of contradictions in 2D NMR data gives the investigator valuable information that can determine the strategy of structure elucidation with these data. An erroneous program report regarding the presence of nonstandard connectivities may appear if properties of at least one atom are assigned incorrectly (for instance, label “*fb*” is set instead of “*ob*”). A “false” message can also appear in those relatively rare cases when an unknown under study contains a pair of bonded heteroatoms, but the absence of such atomic pairs is set in the program options. In these situations a program message regarding the existence of contradictions can help the chemist to reveal the presence of bonded heteroatoms (see examples in Part III). There were no other cases of incorrect detection of nonstandard connectivities where contradictions in 2D NMR data were not present. The program frequently not only identifies the contradictions in the data correctly, but is able to successfully remove them automatically to allow determination of the correct structure. The program was unable to detect the presence of NSCs when 2D NMR data mainly contained only one HMBC nonstandard connectivity. This occurrence can be explained by the fact that if there are only one or two HMBC nonstandard connectivities in the data, the atoms in a conceivable structure may be arranged so that their arrangement complies with the standard length of all connectivities. If the number of NSCs is large, such an arrangement of atoms is unlikely. The presence of *implicit* nonstandard connectivities can become apparent as a result of structure generation and subsequent structure filtration with the use of spectral libraries: if all the generated structures obviously contradict the spectral data, the program produces an empty results file. Indirect evidence of the possibility that contradictions were not detected may not only be an empty result file but large values, more than 3.5–4.5 ppm, of the chemical shift deviations, d_A and d_N , calculated for the first ranked structure. Investigations have shown that nonstandard connectivities were detected by both direct and indirect methods for 95 % of the analyzed tasks containing contradictory data. If there are reasons to assume that the program did not detect contradictions in the initial data it would be highly likely that the problem could be solved with the use of Fuzzy Structure Generation as will be described in Chap. 5.

Since it is possible that 2D NMR data can contain implicit nonstandard connectivities, the most probable structure generally requires additional verification by independent methods. Particularly, incorrect structures can be rejected on the basis of predicted chemical shifts and multiplets in the ^1H NMR spectrum. However the most effective method, as we will see, is application of FSG. If the structure is generated after automated removal of contradictions then it is still desirable to check for the presence of nonstandard connectivities. The connectivities can be verified with appropriate experimental parameter optimization to probe the values of the spin couplings [4, 5].

Therefore, it is not always possible to find nonstandard connectivities and to automatically resolve the contradictions in 2D NMR data sets. In practice, the following difficult situations accounting for the presence of NSCs may typically arise:

1. The program detects the presence of NSCs and makes an attempt to remove the contradictions in the data but then reports that contradictions cannot be removed automatically. Frequently, FSG can help to solve the problem. Generally, additional experiments are required in an effort to detect NSCs.
2. The program fails to detect nonstandard connectivities and displays a message informing the researcher about the absence of contradictions. In this case strict structure generation is initiated. The following outcomes are possible: (a) no structure is generated and saved after filtering; (b) the wrong structure(s) is generated, which can generally be recognized because of the large values of the ^{13}C experimental versus predicted deviations. Again FSG can help in this situation.
3. The program detects the presence of nonstandard connectivities, makes an attempt to remove the contradictions in the data, and displays a message that the contradictions were removed, though in fact, some contradictions still remain. This is due to the fact that not all nonstandard connectivities are lengthened. Possible nondesirable consequences, and the methods to overcome them, are similar to those listed for Point 2 above.

It should be obvious that the most dangerous situations are when incorrect solutions are produced and these can occur for Points 2 and 3 above. Even in those cases when the program is not able to remove detected contradictions, specifically case 1, the fact that contradictions are detected is of great importance for selection of an optimal method of problem solving.

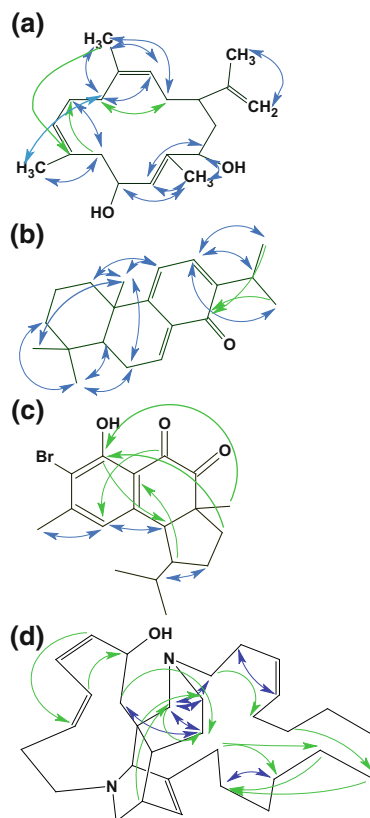
2.3.2 Solving Problems Using Fuzzy Structure Generation

In those cases when correlations are present in the 2D NMR data with nJ where $n > 4$, the method of automatic removal of contradictions, unfortunately, does not work. The augmentation of the path between two intervening nuclei by one bond obviously cannot lead to the generation of a correct structure in this case. Moreover,

due to a lack of constraints that are to be logically analyzed, even in those cases when $n = 4$ the algorithm gives no guarantee that all nonstandard correlations will be found and corrected. For example, the greater the number of carbon atoms with properly defined properties (in regard to the type of hybridization and different heteroatom neighborhoods) and/or the higher the total number of available 2D NMR connectivities, the higher the probability of successfully performing logical analysis to arrive at the correct structure. In contrast, severely proton-deficient molecules are among the most challenging. Obviously, the problem becomes more computationally complicated as the size and complexity of a molecule increases.

The number of NSCs contained within the 2D NMR data associated with a molecule, m , can be rather large—up to about 20 correlations. At the same time, to remove contradictions the augmentation of standard correlation lengths, a , could be 1–3. As an example of such situations several structures taken from the literature are used to demonstrate those examples with a large number of NSCs including 5J and 6J coupling constants (see Fig. 2.22).

Fig. 2.22 An illustration of a number of structures containing multiple nonstandard correlations.
a $m = 15$, $a = 3$ [16];
b $m = 13$, $a = 3$ [17]; **c** $m = 8$,
 $a = 3$ [18]; **d** $m = 18$, $a = 2$ [19]



The nonstandard COSY correlations are shown as blue arrows and the HMBC correlations by green arrows. In the legends for the structures m is the total number of nonstandard correlations, and a (augmentation) is the value of correlation lengthening allowed during the process of FSG.

To overcome the described difficulties, a computational approach was suggested that has been defined as Fuzzy Structure Generation.

2.3.3 Modes of Fuzzy Structure Generation

Numerous computational experiments have allowed us to conclude that if the program detects the presence of NSCs but fails to resolve contradictions in the 2D NMR data using algorithms described in Sect. 2.1.4, then FSG should be used to solve the problem. Moreover, it is quite probable that structure elucidation from 2D NMR data on the basis of FSG can be considered as a general CASE strategy because it is almost independent of the presence or absence of NSCs in the 2D NMR data.

FSG can easily be controlled by parameters that make up a set of options. The two main parameters are: m —number of nonstandard connectivities and a —the number of bonds by which some connectivity lengths should be augmented. Unfortunately, 2D NMR spectral data cannot deliver definitive information regarding the values of these variables and, as a matter of fact, both of them can be determined only during the process of structure elucidation. It has been concluded that in many cases the risk of choosing an erroneous value for a can be avoided and the solution of a problem can be considerably simplified if the lengthening of the m connectivities is replaced by their *deletion*. When set in the options the program can ignore by deletion connectivity responses that have to be augmented (by convention, the parameter a is set to a value of 16 in these cases). Such an approach can be successful in those cases when the number of 2D NMR connectivities is in some sense optimal. In this sense we mean that the total number of connectivities (structural constraints), N , must be large enough to facilitate description of the chemical structure. In many instances there are sufficient numbers of correlations in the ensemble of 2D NMR data acquired to essentially over determine the structure—in other words there is redundancy in some of the connectivity information. It can then be expected that deletion of m of the connectivities will not dramatically influence either the generation time or the size of the output file. On the other hand, the number of combinations of N connectivities taken m at a time can be very large. This can dramatically impede problem solving to a point that it is not feasible to solve the problem. Indeed, some researchers have commented that some of the ACCORDION-optimized long-range heteronuclear shift correlation experiments [15] actually provide too many long-range correlations of the type ${}^nJ_{\text{XH}}$ where $n \geq 4$.

If the total number of connectivities, N , is small then further decreasing N by m in a connectivity combination can lead to an excessive decrease in the number of structural constraints required for solving the problem. In such a case the problem

may be difficult to solve because the 2D NMR data structural constraints will only reduce the total number of possible isomers very slightly.

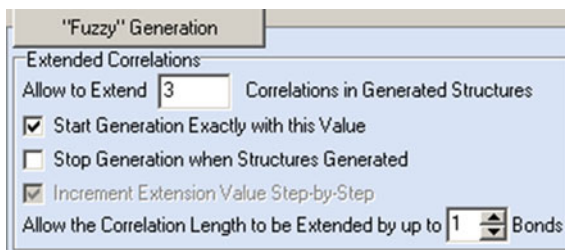
Independent of the use of augmentation or removal of connectivities, the crucial point in application of FSG is the number of connectivity combinations that should be checked during structure generation. For instance, if $N = 60$ and $m = 5$ then the number of connectivity combinations, $n_{\text{math}} = C_N^m$, is equal to ~ 5.5 million. Any attempt at structure generation has to be performed using each of these combinations. It is necessary to perform generation of structures from each of the C_N^m data sets and obtain the output file as a unification of all of the intermediate results. Although the StrucEluc structure generator is fast, the productivity is certainly insufficient in terms of coping with a combinatorial problem as outlined here.

To overcome this difficulty the system is delivered with an algorithm capable of reducing the number of combinations without the risk of losing the correct solution. The first step is to reduce the total number of connectivities N down to N_0 , where N_0 is the number of connectivities used to form the connectivity combinations. The data are preprocessed according to the following rules: (1) ambiguous connectivities are excluded from consideration; (2) if two connectivities C1 to C2 and C2 to C1 are present then only one of them is included in a data set. One of the two equivalent correlations is redundant and corresponds to overdetermination of the data needed for solution of the structure. The second and most important step is based on the results of logical analysis of the initial 2D NMR data. If connectivity sets containing NSCs are identified, then groups of these connectivities are utilized to produce connectivity combinations. As a consequence connectivities that are suspected to be nonstandard are included in all resulting combinations and the initial number of combinations reduces. In addition, the algorithm is capable of immediately detecting combinations of connectivities from which structure generation is impossible—a connectivity combination of this kind still contains at least one NSC. These combinations are skipped during the structure generation process. As a result FSG can be performed in a reasonable time even in those cases when n_{math} is very large. If the MCD checking process fails to detect nonstandard correlations in the 2D NMR data (the probability of failure is about 10 %) the program is forced to try all C_N^m connectivity combinations. This can drastically increase the time to solve the problem and the described approach is inefficient. In these cases, UFs and FFs can frequently be helpful. The ability of the program to calculate and display the real number of connectivity combinations to be validated during FSG allows the user to approximately evaluate the complexity of a given task even at the first stage of the structure elucidation process.

When option parameters are combined in a different way it is possible to initiate the following most appropriate modes of FSG:

Mode 1 Structures are generated such that the number of correlations that are extended is specified ($m = m_0$) and connectivity augmentation is also assigned ($a = a_0$). In this case for an HMBC correlation having a length of 1–2 skeletal bonds both the lower and upper length limits are updated and the connectivity length is extended to three bonds. Example: $m = 3$, $a = 1$, (Fig. 2.23).

Fig. 2.23 *Mode 1*: Example of fuzzy generation options



Mode 2 Structure generation is performed using the following options: it is assumed that the number of extendable (or ignored) connectivities can not exceed m_{\max} , ($m = 1, 2, \dots, m_{\max}$), while a is equal to a_0 . The m_{\max} value is defined as the *maximum* allowed number of nonstandard correlations in the 2D NMR data. Typically the m_{\max} value is set equal to 20 thereby covering a wide range of nonstandard connectivities (see Fig. 2.22). The program initially performs structure generation with a value of $m = 1$. If the attempt is unsuccessful then the m value is *automatically* incremented by 1 and a new run is made with $m = 2$ and so on. An iteration is declared unsuccessful if either no structure is stored after structure generation and spectral filtration or if an *unacceptable* solution was found. When m reaches the m_g value the program considers the 2D NMR data to be consistent, then FSG is initiated with $m = m_g$. The program stops after completing structure generation with $m = m_g$ if the output structure file is not empty and if an *acceptable* solution is provided. Example: $m = 1-20$, $a = 1$, (Fig. 2.24).

Mode 3 The number of connectivities m is allowed to vary between m_{\min} and m_{\max} values ($m_{\min} \leq m \leq m_{\max}$), while the fixed number of bonds a_0 is set. The minimum number m_{\min} is usually derived as a result of checking the 2D NMR data for consistency. The program stops when similar conditions as described for *Mode 2* are achieved. Example: $m = 1-20$, $a = 1$, (Fig. 2.25).

Mode 4 This mode is a generalization of *Mode 3* where the interval for m value variation is defined by the condition $m_{\min} \leq m \leq m_{\max}$ at $m_{\min} = 0$. The peculiarity of this mode is that it is a “generalized” mode of structure generation and can be initiated with $m = 0$. In this mode, the program starts by checking the hypothesis that NSCs are absent in a given 2D NMR dataset. If the dataset does not contain nonstandard connectivities then the program completes the process of structure generation and the further solution of the problem is carried out as described previously (Sect. 2.2).

Fig. 2.24 *Mode 2*: Example of fuzzy generation options

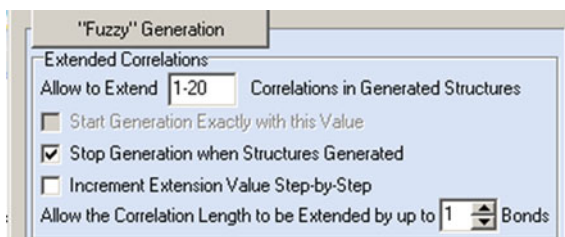
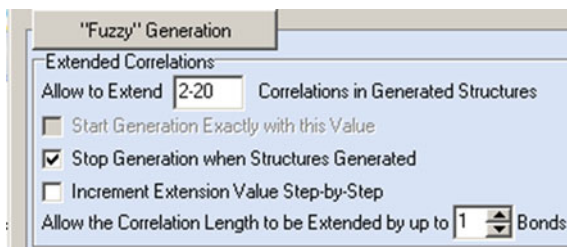


Fig. 2.25 *Mode 3*: Example of fuzzy generation options



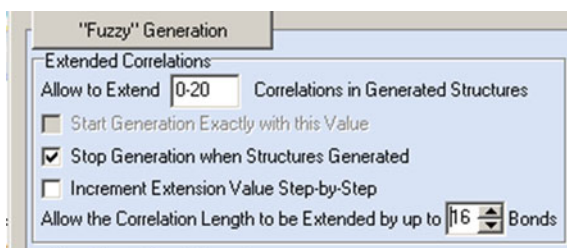
If an attempt with $m = 0$ proves to be unsuccessful then the program automatically performs FSG starting with $m = 1$, $a = 16$ and continues problem solving in the manner described earlier for *Mode 3*. The merit of such an approach is that no assumption regarding the a value is necessary. Example: $m = 0-20$, $a = 16$, (Fig. 2.26).

Mode 5 This mode is initiated if it is necessary to perform FSG iteratively covering all values of m starting from m_{\min} to m_{\max} without exclusion. For example, if structure generation is successful at $m = m_g$ then the program automatically switches to $m = m_g + 1$ and so on until it reaches $m = m_{\max}$. The structures generated at each step are added to those generated during the previous step. This mode is useful to check the solution for stability to make sure that the best structures found at steps $m = m_g$ and $m = m_g + 1$ or higher are equivalent. Examples: $m = 1-4$, $a = 16$:

- Generation is performed at $m = 1, 2, 3, 4$, and even if no structure is saved at some m value the generation will be continued at $m + 1$ and so on (Fig. 2.27).
- If no structure is saved at some m value the generation will be stopped (Fig. 2.28).

Mode 6 This mode resembles *Mode 5*, but the function of this mode is to generate all structures for which the number of nonstandard connectivities is less or equal to m at the given value of a . The corresponding options are denoted as $\{m \leq m_0, a = a_0\}$. The number of connectivity combinations from which FSG is performed depends only on the N_0 and m values. In contrast to the “step-by-step” modes some combinations of the connectivities are united by this approach and this in principle can speed up the calculations. When this procedure is performed only the maximal lengths of HMBC connectivities (i.e., two skeletal bond lengths) are

Fig. 2.26 *Mode 4*: Example of fuzzy generation options



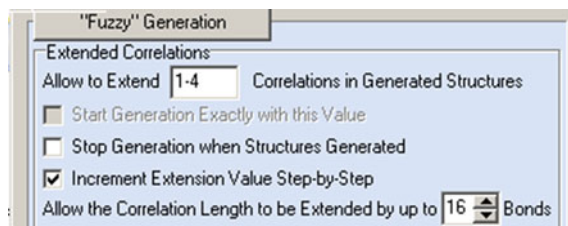
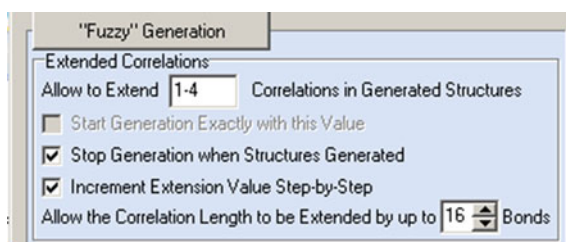


Fig. 2.27 *Mode 5*: Example of fuzzy generation options

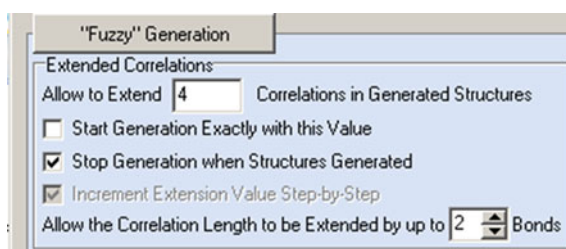
Fig. 2.28 *Mode 5*: Example of fuzzy generation options



enlarged. For example, consider an HMBC connectivity between C-1 and C-2 atoms whose “standard” length is varied from 1 to 2 skeletal bonds. In this mode the updated connectivity length varies from 1 to 3 skeletal bonds. It is important to note that the number of nonisomorphic structures generated in this mode is equal to the total number of nonisomorphic structures generated during all steps of *Mode 5*. However, the total time necessary for completion of FSG can be significantly different between these modes. Example: $m = 4$, $a = 2$, (Fig. 2.29).

Mode 7 This approach gives the researcher a chance to solve a problem in a fully automated mode. To initiate this mode the commands “**Allow Fuzzy Structure Generation**” and “**Determine Options Automatically**” are selected (see Fig. 2.10). The program analyzes the 2D NMR data and depending on the results makes a corresponding decision on the choice of the generation parameters and the

Fig. 2.29 *Mode 6*: Example of fuzzy generation options



strategy of their application. If a problem can be solved in the **Common Mode** (without using fragments) FSG with automatically determined parameters is very effective.

Note that *Mode 4* with the parameter *a* set equal to 16 (see an example below) can be considered as the most comprehensive mode since in principle it will solve a problem in which the 2D NMR data contain an *unknown* number of nonstandard connectivities of an *unknown* length.

If the problem is successfully solved with a given set of options then the real *m* and *a* values are reported by the program. Nonstandard connectivities are observed visually from the resulting structure which is displayed along with all COSY and HMBC connectivities. Nonstandard connectivities are easily recognized as they are marked in a red color.

In addition to the approaches mentioned for controlling FSG there is also a possibility to exclude the COSY data from the process of FSG as a user option (see Fig. 2.10, the group **Use Connectivities from...**). In some cases, especially those when the COSY data contain many NSCs and at the same time the HMBC data are rich enough, the exclusion of the COSY data accelerates the solution of the problem. Removal of weak peaks from COSY and HMBC spectra and elongation of all COSY connectivities up to three bonds (correlations of ${}^4J_{\text{HH}}$ type) can also be helpful. The presence of NSCs in the COSY data can sometimes be detected by repeated MCD checking—with COSY data switched *on* and *off*.

2.3.4 The Strategy of Applying Fuzzy Structure Generation

The possibility of employing several different modes of FSG proves to be a very flexible analytical tool. However, the diversity of modes available is also a source of complexity since the user has to choose the optimal mode when solving a specific problem. Before starting the calculations it is unclear which mode will lead to a solution in a reasonable time.

An attempt was made to answer the question of whether there is a general strategy of structure elucidation using FSG that works best. A set of more than 100 problems were selected where either the HMBC or COSY spectra, or both, contained a total of 1–18 nonstandard connectivities corresponding to a range of coupling constants ${}^nJ_{\text{HH}, \text{CH}}$ where $n = 4–6$. The structures under investigation were all natural products and the number of skeletal atoms in the molecules varied between 15 and 75 skeletal atoms.

For each problem the NMR spectral data were entered into the program and graphically represented as MCDs. The procedure for checking the 2D NMR data for contradictions was then applied to every problem. If the presence of NSCs was

revealed then the program displayed the minimum number of nonstandard connectivities and made an attempt to automatically resolve the contradictions as described above. In successful cases the updated MCDs were displayed with modified connectivities marked by violet color.

As a result of these studies all problems were classified into three sets as follows:

- (1) 53 problems were identified where NSCs were detected and the initial MCDs were updated;
- (2) 34 problems were identified where the program revealed the presence of NSCs but failed to update the MCDs;
- (3) 13 problems were identified where the program failed to detect NSCs.

This classification describes all conceivable results of checking the MCDs. Depending on the results of checking the MCD, various modes or combinations of modes can lead to a solution of the problem. Attempts to solve each problem were made using different FSG modes to investigate possible approaches. The problems for which valid solutions could not be found during the first attempt were eventually solved after utilizing different fuzzy generation options. Logical data pre-processing frequently allowed significant reduction of the number of connectivity combinations to be tested during FSG. Figure 2.30 shows the ratio $\rho = n_{\text{real}}/n_{\text{math}}$, where n_{real} is the number of tested connectivity combinations, $n_{\text{math}} = C_{N_0}^m$ is the theoretically calculated number of combinations. Figure 2.31 examines these combinations in greater detail.

The figures demonstrate that the theoretical number of combinations can be hundreds of billions but the real numbers reduce down to manageable dimensions. For instance, in 20 problems the theoretical number dropped by 10^4 – 10^6 times but the real numbers of combinations still remained rather large. Nevertheless, the speed of the structure generator algorithm was fast enough to solve almost all problems.

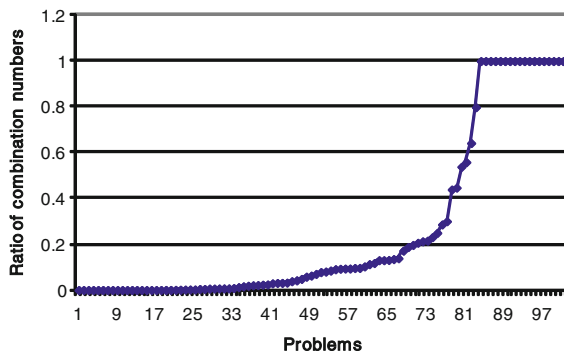
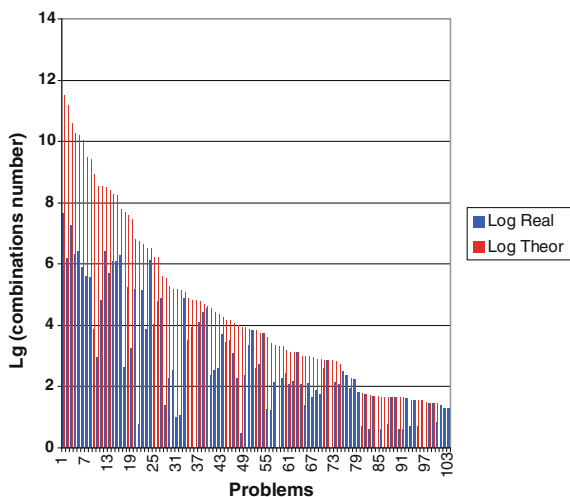


Fig. 2.30 The ratio of the numbers of real connectivity combinations to the numbers of theoretically possible combinations for the problems solved using FSG. The program failed to reduce the number of combinations mainly in those cases when nonstandard connectivities were not detected during checking of the MCD

Fig. 2.31 A plot of the logarithms of the theoretical (red) and real (blue) numbers of connectivity combinations



Fuzzy Structure Generation did, however, fail for the elucidation of structure **d** ($C_{32}H_{50}NO_2$) in Fig. 2.22. The 2D NMR data contain 18 nonstandard connectivities (12 HMBC and 6 COSY nonstandard connectivities; 5 connectivities are of type 5J). The theoretical number n_{math} of connectivity combinations is equal to $\sim 43 \times 10^{12}$ for this case. The difficulty could be circumvented by using the Fragment Mode, but no large appropriate fragment was found in the database during the ^{13}C NMR search. The application of a large UF led to an extremely large set of MCDs with each containing the UF with different distributions of the carbon chemical shifts. As a result these two combinatorial “explosions” hampered problem solving. The solution of such computationally difficult problems will hopefully be eased by further development of the algorithm providing fragment “implementation” in MCDs.

As a result of the studies described, general traits were identified that could help to find appropriate ways to solve a problem. These strategies, as applied to the three problem subsets mentioned above, are described in the following subsections.

2.3.4.1 NSCs Were Identified and the MCD Was Updated

Assuming that the MCD updating process was performed successfully (with the lengths of all NSCs increased) then *strict* structure generation is performed. If an acceptable solution is obtained then it should be checked for stability. FSG with the options $\{m = m_{\text{min}} \div 20$; stop at $m = m_g$, $a = 16\}$ is started from the initial MCD, not the updated MCD. The previously found solution will be confirmed if the first ranked structures for both strict and fuzzy solutions coincide. When an inequality $d_A^{\text{st}}(1) > d_A^{\text{fuz}}(1, m_g)$ is observed ($d_A^{\text{st}}(1)$ —the deviation calculated for the first ranked structure of the solution found by strict structure generation, $d_A^{\text{fuz}}(1, m_g)$ —

the same found by FSG at $m = m_g$), then it is concluded that not all NSCs were lengthened during updating of the MCD and Fuzzy Structure Generation should be repeated with $m_g + 1$ and so on until the minimum value of $d_A^{\text{fuz}}(1, m_g + v)$ and a valid solution is achieved at $m = m_g + v$. The corresponding structure is then considered as the most probable.

An unacceptable solution can be obtained as a result of strict structure generation from the updated MCD, i.e., a solution will be found for either $d_A^{\text{st}}(1) > D_A$, where D_A is a threshold value or an empty structural file is obtained ($k = 0$). In both cases the program is automatically switched to the mode where $\{m = m_{\text{min}} \div 20, \text{ stop at } m = m_g, a = 16\}$. Depending on the m_g values and the complexity of the problem (the size of n_{real} and the calculation time) evaluated during the first stages of solving the problem, the user can initiate FSG with the options $\{m \leq m_0, a = 16\}$, $m_0 = 5, 10$ or 15 to obtain the most reliable solution.

2.3.4.2 NSCs Were Identified but the MCD Failed to Be Updated

If the program identified NSCs but failed to update the MCD, then FSG is one manner by which to solve such a problem. Since the program only displays the minimum number of NSCs while their associated lengths remain unknown, FSG with the options $\{m = m_{\text{min}} - 20, \text{ stop at } m = m_g, a = 16\}$ should be used. The real numbers of the connectivity combinations, n_{real} , are displayed, as well as the number of combinations for a given $m = m_g$, and the approximately predicted time for structure generation allows the user to easily evaluate the complexity of the problem and the suggested time for execution. If *Mode 4* can be applied based on acceptable time estimates then it should be used.

2.3.4.3 NSCs Were Not Detected

If nonstandard connectivities were not revealed by checking the MCDs then there are two ways to interpret this result: either the 2D NMR data is free of nonstandard connectivities or the implicit NSCs are present but the program failed to detect them. Both of these situations are covered by FSG with the options $\{m = 0 \div 20, \text{ stop at } m = m_g, a = 16\}$. If NSCs are indeed absent from the 2D NMR data then structure generation is performed with $m = 0$ with a nonzero output file and the deviation values allow the user to determine whether the solution determined is acceptable. Obtaining deviation values that exceed the threshold for d_A , or deriving an empty output file after spectral filtering, both serve as hints to the presence of latent nonstandard connectivities.

When NSCs are not detected by logical data analysis then the number of connectivity combinations that must be tested during FSG cannot be reduced and it is equal to $C_{N_0}^m$, $m = 1, 2, 3, \dots$ at each m th step of the FSG process. This situation can cause significant difficulties due to an unmanageable number of connectivity

combinations needing to be processed; as discussed previously, both FF and UF can assist in this situation.

It is hardly possible to describe all of the nuances associated with FSG since these depend on each 2D NMR data set associated with a given problem. A series of examples illustrating the strategies leading to valid solutions with the minimum number of user assumptions will be presented in Chap. 5.

2.3.5 Is There an Alternative to Fuzzy Structure Generation?

Some researchers suggested that it was possible to overcome the problem of NSCs by setting default values for ${}^4J_{\text{CH}, \text{HH}}$ for *all* COSY and HMBC correlations observed in the 2D NMR spectra. It was important to answer the question: To what extent can the lengthening of *all* 2D NMR correlations act as a method for contradiction resolution in 2D NMR data? A study was undertaken to answer this question [7]. In this study an attempt was made to identify, in a quantitative manner, how the structure generation time increases and the amount of structural information obtained decreases if only correlations in the ${}^{2-4}J$ range were allowed.

Analysis of the results of this study showed that even in the case of small molecules the output file size increases considerably when the ${}^{2-4}J$ couplings are set as default. The generation time increases by many times to hundreds or even tens of thousand times greater. For one of the studied problems the size of the output file increased from 2 to ca. 3,000 structures, while the generation time increased by 6.5 million times! The main conclusion is that the lengthening of *all* correlations should be rejected as a general method of solving problems arising from the presence of nonstandard correlations in 2D NMR data.

References

1. Thongbai B, Surup F, Mohr K, Kuhnert E, Hyde KD, Stadler M (2013) Gymnopalynes A and B, chloropropynyl-isocoumarin antibiotics from cultures of the basidiomycete *Gymnopus sp.* J Nat Prod 76(11):2141–2144. doi:10.1021/np400609f
2. Benie AJ, Sørensen OW (2007) HAT HMBC: a hybrid of H2BC and HMBC overcoming shortcomings of both. J Magn Reson 184(2):315–321
3. Krishnamurthy V, Russell D, Hadden C, Martin GE (2000) 2J, (3)J-HMBC: a new long-range heteronuclear shift correlation technique capable of differentiating (2)J(CH) from (3)J(CH) correlations to protonated carbons. J Magn Reson 146(1):232–239
4. Nyberg NT, Duus JØ, Sørensen OW (2005) Heteronuclear two-bond correlation: suppressing heteronuclear three-bond or higher NMR correlations while enhancing two-bond correlations even for vanishing 2J(CH). J Am Chem Soc 127(17):6154–6155
5. Sprang T, Bigler P (2003) A new technique for differentiating between 2J(C, H) and 3/4J(C, H) connectivities. Magn Reson Chem 41(3):177–182

6. Molodtsov SG, Elyashberg ME, Blinov KA, Williams AJ, Martin GM, Lefebvre B (2004) Structure elucidation from 2D NMR spectra using the StrucEluc expert system: detection and removal of contradictions in the data. *J Chem Inf Comput Sci* 44:1737–1751
7. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ, Martin GE (2007) Fuzzy structure generation: a new efficient tool for computer-aided structure elucidation (CASE). *J Chem Inf Model* 47(3):1053–1066
8. Elyashberg ME, Blinov KA, Williams AJ (2009) A systematic approach for the generation and verification of structural hypotheses. *Magn Reson Chem* 47(5):371–389. doi:10.1002/mrc.2397
9. Blinov KA, Elyashberg ME, Martirosian ER, Molodtsov SG, Williams AJ, Sharaf MMH, Schiff PLJ, Crouch RC, Martin GE, Hadden CE, Guido JE, Mills KA (2003) Quindolinocryptotackeine: the elucidation of a novel indoloquinoline alkaloid structure through the use of computer-assisted structure elucidation and 2D NMR. *Magn Reson Chem* 41:577–584
10. Martin GE, Hadden BD, Russell CE, Kaluzny DJ, Guido JE, Duholke WK, Stiemsma BA, Thamann TJ, Crouch RC, Blinov KA, Elyashberg ME, Martirosian ER, Molodtsov SG, Williams AJ, Schiff PLJ (2002) Identification of degradants of a complex alkaloid using NMR cryoprobe technology and ACD/structure Elucidator. *J Heterocycl Chem* 39:1241–1250
11. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ (2013) Structure revision of asperjinone using computer-assisted structure elucidation methods. *J Nat Prod* 76:113–116
12. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ (2012) Elucidating “undecipherable” chemical structures using computer assisted structure elucidation approaches. *Magn Reson Chem* 50:22–27
13. Elyashberg ME, Williams AJ, Martin GE (2008) Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation. *Prog Nucl Magn Reson Spectrosc* 53(1/2): 1–104
14. Elyashberg ME, Williams AJ, Blinov KA (2012) Contemporary computer-assisted approaches to molecular structure elucidation, vol 1. *New Developments in NMR*. RSC Publishing, Cambridge
15. Berger S, Braun S (2004) 200 and more NMR experiments: a practical course. Wiley, New York
16. Collins DO, Reynolds WF, Reese PB (2004) New cembranes from *Cleome spinosa*. *J Nat Prod* 67:179–183
17. Mensah AY, Houghton PJ, Bloomfield S, Vlietinck A, Berghe DV (2000) Known and novel terpenes from *buddleja globosa* displaying selective antifungal activity against dermatophytes. *J Nat Prod* 63:1210–1213
18. Wellington KD, Cambie RC, Rutledge PS, Bergquist PR (2000) Chemistry of sponges. 19. Novel bioactive metabolites from *Hamigera tarangaensis*. *J Nat Prod* 63:79–85
19. Oliveira JHHL, Grube A, Köck M, Berlinck RGS, Macedo ML, Ferreira AG, Hajdu E (2004) Ingenamine G and cyclostelletamines G-I, K, and L from the new Brazilian species of marine sponge *Pachychalina sp.* *J Nat Prod* 67:1685–1689

Part II
Getting Started with *Structure Elucidator*

Chapter 3

Simple Examples of Structure Elucidation

Abstract This chapter can be considered as an introduction to practical approaches used for structure elucidation based on the application of Structure Elucidator. 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 structure elucidation from 1D and 2D NMR, MS, 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 Structure Elucidator. A detailed description of the manual solutions is available on 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 the previous chapters. For this purpose the student can use a limited version of Structure Elucidator which can be downloaded for free (www.acdlabs.com/TeachingSE). All spectroscopic data required for each task is already presented in electronic formats appropriate for use with the program. The student may vary the axioms formulated for solving a given task and follow how the solution changes depending on the user assumptions. We believe that the reader, who works through this chapter, in combination with solving all challenges, would acquire the knowledge and skill necessary to solve the complicated real-world problems given in Chaps. 4 and 5.

In this chapter we will consider the application of Structure Elucidator for solving 22 challenges offered to students in Chap. 5 of the textbook by Reichenbächer and Popp [1]. A complete set of spectroscopic data including MS, IR, UV, as well as NMR (^1H , ^{13}C , HSQC, HMBC, and COSY spectra) are presented in the book body, while the reader is given a unique opportunity to download a detailed description of the problem solutions from the “Extra Materials” [2] stored on the server of the publisher. This possibility makes the textbook [1] extremely valuable for those who want to gain insight into the art of molecular structure elucidation from spectroscopic data.

The problems described in this chapter are presented in the folder “Challenges_Chapter_3” (see www.acdlabs.com/TeachingSE) as projects stored in Structure Elucidator format (.gnr). Each file contains the molecular formula of an “unknown” in the **PM** window and electronic tables of spectroscopic data used as the basis of problem solving. A Molecular Connectivity Diagram (**User MCD** window), a proposed molecule (**PM** window) and structures included into the output file (**Mol** window) are saved for part of the challenges to ease the user in becoming familiar with the program interface. All challenges were solved at $tol = 0.005$ ppm set for F1 and F2, where tol is the value of **User Defined Tolerance** shown in the dialog window **Spectrum Parameters**. The influence of this parameter on the MCD properties and the problem solution can be investigated by the user. After installation of the restricted version of Structure Elucidator, a challenge can be downloaded into the program either by dragging the corresponding file to the main program window or using the command **File\Open**. The chemical editor ACD\ChemSketch [3], an integral part of Structure Elucidator, will also be automatically installed. In addition to the various tools which provide for the creation and processing of chemical structures, ChemSketch allows for the searching of a drawn structure against rich online sources of chemical information such as PubChem [4], ChemSpider [5], and eMolecules [6]. Some physicochemical parameters of the analyzed molecule can also be predicted (the **Properties** command in the right lower corner of the ChemSketch window).

With these aids, the user will have an opportunity to repeat all steps of problem solving described in this chapter. All dialog windows are supplied with **Help** buttons, which significantly helps in navigating the program if needed. The main restriction intrinsic to the available limited program version is that the spectroscopic data presented in electronic tables for a given problem cannot be edited, but the data can be freely edited in the MCD created from the initial data. Therefore, the reader may suggest his own axioms, introduce them into the MCD, and check how the new axioms influence the solution of the problem. We hope that such computational experiments will allow the reader to understand more fully the essence of the CASE approach and realize its advantages. If after solving several problems with the assistance of the book the user feels that he is ready to fully test his experience with computer-assisted structure elucidation, then he may start solving the next problems before reading the corresponding sections in this chapter.

Some menus and dialog windows of the restricted version of the program provided differ somewhat from those presented in Chap. 2 for the full version of the program. We believe the difference is not significant and will not create any confusion.

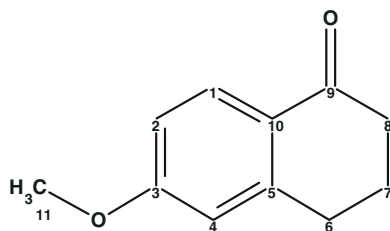
As emphasized earlier the application of CASE does not save the chemist from the necessity of needing to develop skills in the manual interpretation of molecular spectra. Moreover, the more experienced the spectroscopist, the more efficient and successful are his efforts in computer-assisted structure elucidation. The reason is that all data input into the program must be carefully checked and edited

(if necessary) by the researcher to ensure the utilization of all evident information and avoid penetrating erroneous axioms in the data set.

The approach to “computer reasoning” differs from that intrinsic to a human expert. We believe that the combined usage of Structure Elucidator with the detailed description of manual problem solving will allow the reader to sense the difference. We also believe that in the process of unassisted solving of simple “educational” problems presented in this chapter and using Structure Elucidator, the reader will develop the skills necessary to apply computer-aided approaches to complex real problems which are given in Part III. As mentioned earlier, many of these problems are related to the structure identification of new natural products isolated and published recently (2011–2013). It is worth noting here that the spectral data for most of them were adopted from papers where the structures were characterized as *unprecedented* or *unique*.

3.1 Challenge 1

The structure of an “unknown”:



3.1

The first step in molecular structure elucidation is determination of the molecular formula of the unknown. Analysis of the MS spectrum carried out in the textbook [1] showed clearly that $m/z = 176$ amu was the molecular ion. Because the value of M^+ is even it can be expected that the molecule either does not contain a nitrogen atom or it contains an even number of nitrogen atoms ($n_N = 0$ or 2 or 4...). The IR spectrum of the “unknown” gives no explicit information about the presence or absence of nitrogen atoms: the absorption band at $1,673\text{ cm}^{-1}$ can be accounted for both by a conjugated ketone and a tertiary amide (no absorptions characteristic for NH or OH were observed in the area above $3,000\text{ cm}^{-1}$). No hints regarding the presence of nitrile group were observed ($2,200\text{--}2,600\text{ cm}^{-1}$). IR absorption bands at $1,594$ and at $1,497\text{ cm}^{-1}$ suggested the presence of a benzene ring in the molecule, which is confirmed by the characteristic peak at $m/z 91$ amu in the MS spectrum. The ^{13}C NMR spectrum contains 11 signals so the minimum number of carbon atoms is equal to 11. Integrals measured in the ^1H NMR spectrum as well as the ^{13}C DEPT 135 pattern provide evidence that the minimum number of hydrogen

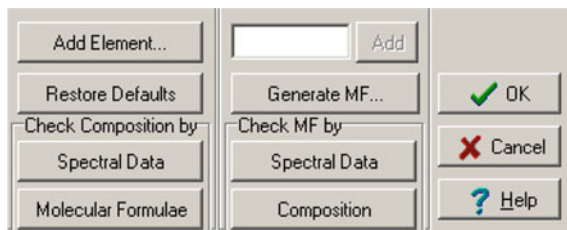


Fig. 3.1 Challenge 1: Part of the **Edit Spectrum Query** window

atoms is 12. The MS spectrum contains no isotopic patterns related to the presence of Cl, Br, and F (no $\Delta m = 19$ or 20, no multiplets in ^{13}C). The absence of a $M+2$ peak with $I_{\text{rel}} \geq 4.4\%$ excludes the possibility of sulfur atoms. The IR band at $1,673\text{ cm}^{-1}$ ($\text{C}=\text{O}$) and a CH_3 singlet at 54.9 ppm is characteristic for a $\text{O}-\text{CH}_3$ in ^{13}C NMR and this suggests that the presence of at least two oxygen atoms is possible.

We conclude that the molecule can belong to the class CHNO . Now we have enough information to infer the conceivable molecular formula(s) of the unknown using the StrucEluc command **Edit Spectrum Query/Generate MF...** (Fig. 3.1).

Figure 3.2 shows the **Molecular Formula Generator** window where the options and parameters of the molecular formula generation are shown.

In the upper left corner, the admissible numbers of the possible chemical elements are shown. It is supposed that 1 or 2 quaternary carbon atoms characterized by low signal intensity in the ^{13}C spectrum could be missed, C (11–13). In the general case the ^1H signals of one or two hydrogen atoms could fall in a dark area caused by absorption of a solvent or would not be observable if they belong to exchangeable hydrogens. In the current problem, dark areas are absent in the ^1H spectrum and IR spectrum and provides evidence of the absence of exchangeable protons. An interval H (12–14) was set to illustrate the principles. Though the minimum number of oxygen atoms is two, the numbers of N (0–6) and O (0–10) atoms are allowed.

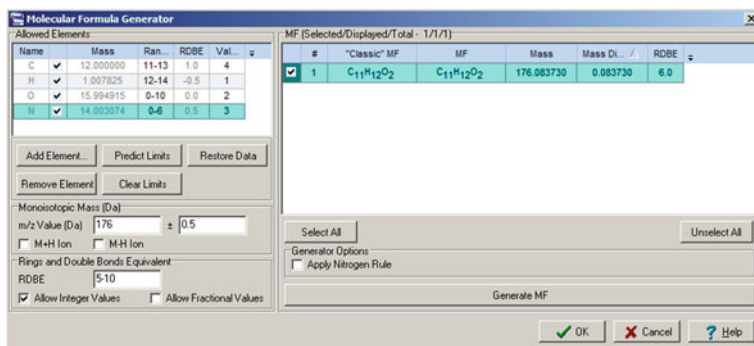


Fig. 3.2 Challenge 1: **Molecular Formula Generator** window

When real-world problems are solved, an accurate value of a molecular ion is usually measured and compared with values calculated for all suggested molecular formulae. For structure **1** $M_{\text{calcd}}^+ = 176.083725$ amu. As we have only a nominal molecular mass of the “unknown” (176 amu), the tolerance was set as ± 0.5 amu. As the presence of a carbonyl and benzene ring was suggested the interval for the rings and double bonds equivalent (RDBE) was postulated as 5–10.

When the command **Generate MF** (Fig. 3.2) was executed a unique molecular formula **C₁₁H₁₂O₂** appeared in the upper right-hand corner. One or more selected molecular formulae can be input into the program by clicking on **OK**. Therefore, the logical inference of a molecular formula (as explained in the textbook [1]) was replaced by combinatorial calculations performed by the program.

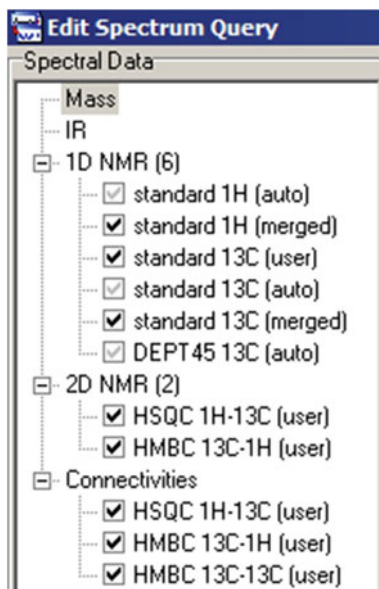
If a traditional method of molecular structure elucidation is used (as described in Sect. 5.4.1 of the textbook [1]), the further steps consist of successive determination of possible molecular fragments by deducing structural suggestions and their verification using all available spectroscopic data. When structure elucidation is carried out with the aid of StrucEluc, during the next step of data processing the main attention of the researcher is placed on the peak picking of signals in the available 1D and 2D NMR spectra. This procedure can be carried out by StrucEluc automatically, but the best way to avoid inputting some erroneous “axioms” is to carry out peak picking in an interactive mode [3]. As a result a table of spectroscopic data is produced which should be checked by the user thoroughly, and corrections have to be made if necessary. The spectrum-structure information that was extracted from the ¹³C, ¹H, HSQC, and HMBC data as a result of spectrum analysis is presented in Table 3.1.

In reality, Table 3.1 was automatically formed by the program when the solution to the problem was found (i.e., a preferable structure with assigned ¹H and ¹³C chemical shifts was selected as the preferred solution). This is why the table contains a column δC_{calc} where ¹³C chemical shifts calculated using HOSE code-based

Table 3.1 Challenge 1: Spectroscopic NMR data used for structure elucidation using StrucEluc

Labels	δC	δC_{calc}	CH _n	δH	M(J)	C HMBC
C1	129	129.73	CH	7.86	d(8.7)	C9, C5, C3
C2	112.6	113.07	CH	6.68	dd(8.7, 2.5)	C10, C4
C3	163.1	163.42	C	–	–	–
C4	112.1	112.39	CH	6.56	d(2.5)	C2, C10
C5	145.4	147.15	C	–	–	–
C6	29.7	30.21	CH ₂	2.78	t(6.1)	C8, C10, C7
C7	22.9	23.48	CH ₂	1.97	quint. (6.3)	C8, C9, C5, C6
C8	38.4	39.1	CH ₂	2.46	t(6.5)	C9, C7
C9	196.4	197.13	C	–	–	–
C10	125.8	126.53	C	–	–	–
C11	54.9	55.49	CH ₃	3.72	S	C3

Fig. 3.3 Challenge 1:
Experimental 1D and 2D
NMR spectra used for solving
the problem



approach are shown and can be easily compared with experimental shifts. Column M(J) contains the multiplicities and J_{HH} coupling constants (in Hz) that were determined from the 1D and 2D NMR spectra. Table 3.1 is designed for representing the initial spectroscopic data in a format which is used for publications.

The list of NMR spectra used for solving the problem is presented in the left upper corner of the window **Edit Spectrum Query** (Fig. 3.3). Figure 3.4 shows the HSQC and HMBC spectra for Challenge 1 as they are displayed in the dialog window.

The data presented in Table 3.1 and in Fig. 3.4 were used for creating the MCD (see Sect. 2.1.3) presented in Fig. 3.5. For this goal the following command is used: **Structure Elucidation\Create Molecular Connectivity Diagram (MCD)...** Fig. 3.5 displays the structural blocks C, CH, CH₂, CH₃ with their chemical shifts and atom properties and the heteroatoms presented in the molecular formula. Carbon atoms are colored in specific colors which mark the states of hybridization of the various carbons: blue— sp^3 , violet— sp^2 , light blue—*not sp* (sp^2 or sp^3). The label “*fb*” shows that the connection of atom C to a heteroatom is *forbidden*, while the label “*ob*” means that the connection to a heteroatom is *obligatory*. The labels *fb* and *ob* are assigned by the program on the basis of the **Atom Property Correlation Table** (see Sect. 1.3.2.4). Green arrows drawn between the carbon atoms map to the corresponding HMBC connectivities.

Figure 3.5 shows that the hybridization states of all carbon atoms except three of them (C 112.10, C 112.6, and 125.8, light blue colored) were determined by the program unambiguously. “Fuzzy” hybridization *not sp* was assigned to the mentioned three atoms because the algorithm took into account the fact that the

(a)

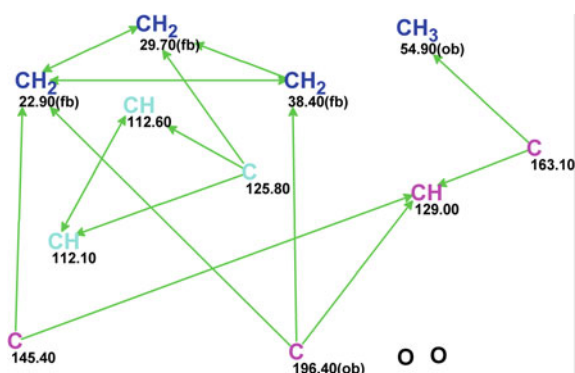
#	ID Δ	Label Δ		F2(ppm)	F1(ppm)	F2 Labels	F1 Labels
1	1	1	✓	129.000	7.860	C 1 (129.000)	H 7 (7.860)
2	2	2	✓	112.600	6.680	C 2 (112.600)	H 6 (6.680)
3	3	3	✓	112.100	6.560	C 4 (112.100)	H 5 (6.560)
4	4	4	✓	54.900	3.720	C 11 (54.900)	H 4 (3.720)
5	5	5	✓	29.700	2.780	C 6 (29.700)	H 3 (2.780)
6	6	6	✓	38.400	2.460	C 8 (38.400)	H 2 (2.460)
7	7	7	✓	22.900	1.970	C 7 (22.900)	H 1 (1.970)

(b)

#	ID Δ	Label Δ		F2(ppm)	F1(ppm)	F2 Labels	F1 Labels
1	1	1	✓	1.970	29.700	H 1 (1.970)	C 6 (29.700)
2	2	2	✓	1.970	38.400	H 1 (1.970)	C 8 (38.400)
3	3	3	✓	1.970	145.400	H 1 (1.970)	C 5 (145.400)
4	4	4	✓	1.970	196.400	H 1 (1.970)	C 9 (196.400)
5	5	5	✓	2.460	22.900	H 2 (2.460)	C 7 (22.900)
6	6	6	✓	2.460	196.400	H 2 (2.460)	C 9 (196.400)
7	7	7	✓	2.780	22.900	H 3 (2.780)	C 7 (22.900)
8	8	8	✓	2.780	38.400	H 3 (2.780)	C 8 (38.400)
9	9	9	✓	2.780	125.800	H 3 (2.780)	C 10 (125.800)
10	10	10	✓	3.720	163.100	H 4 (3.720)	C 3 (163.100)
11	11	11	✓	6.560	112.600	H 5 (6.560)	C 2 (112.600)
12	12	12	✓	6.560	125.800	H 5 (6.560)	C 10 (125.800)
13	13	13	✓	6.680	112.100	H 6 (6.680)	C 4 (112.100)
14	14	14	✓	6.680	125.800	H 6 (6.680)	C 10 (125.800)
15	15	15	✓	7.860	145.400	H 7 (7.860)	C 5 (145.400)
16	16	16	✓	7.860	163.100	H 7 (7.860)	C 3 (163.100)
17	17	17	✓	7.860	196.400	H 7 (7.860)	C 9 (196.400)

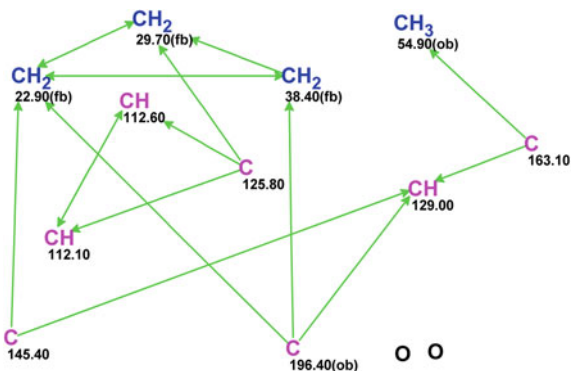
Fig. 3.4 Challenge 1: HSQC (a) and HMBC (b) NMR data used for solving the problem

Fig. 3.5 Challenge 1: Initial MCD



molecule contains two oxygen atoms. Consequently, in accordance with the spectrum-structure correlations, each of the light blue carbons may be included either into the fragment O–C–O (sp^3) or in a carbon double bond (C=C, sp^2) isolated from a heteroatom.

Fig. 3.6 Challenge 1: Edited MCD



After MCD creation, it should be inspected by the researcher. StrucEluc provides aids which allow the researcher to edit the MCD using all reliable additional information and the expert's own conclusions based on preliminary analysis of the spectroscopic data. In the current case there is no doubt about the presence of both C=O and O-CH₃ groups (C 196.4 and C 54.9 got "ob" label automatically), therefore formation of an O-C-O fragment is impossible, and *sp*² hybridization should be assigned to the light blue carbon atoms (by pressing the icon **Edit Atom Properties** and clicking on the corresponding atom). As a result of editing the MCD we obtain the pattern presented in Fig. 3.6

Checking the MCD (**Structure Elucidation\Check MCDs...**) revealed that there were no contradictions in the HMBC (the MCD passed the check for the presence of contradictions). Therefore, Strict Structure Generation was initiated (**Structure Elucidation\Run SCB Generator...**) using common **SCB Generator Options** (see Figs. 2.6 and 2.10).

Structure generation was completed with the following results: 26 molecules have been generated and 20 molecules have been stored after filtering in a generation time of 0.014 s. Then ¹³C chemical shift prediction was performed using an incremental method (**Tools\Calculate\NMR Spectrum\Increments**) and identical structures were removed (**Tools\Remove Duplicated Structures**). Finally, 10 nonidentical candidate structures were obtained. We present the result of structure generation in the following concise form: $k = 26 \rightarrow 20 \rightarrow 10$, $t_g = 0.014$ s. This form will be used as a template in this book for summarizing the results of problem solving.

According to the general CASE strategy (see Sect. 2.2.1), the output file was ranked in ascending order of average deviations and then ¹³C chemical shift prediction was performed using both neural net and HOSE code-based approaches (**Tool\Calculate\NMR (Neural Net)**) or **Tool\Calculate\All Spectra...**). Structures were ordered again by the average deviations d_A found for HOSE code-based ¹³C chemical shift calculations. To ease analysis of the solution obtained the structures are displayed in the form of a Structure List (click on icon **Structure List**). To see the average deviations calculated by the different methods for different

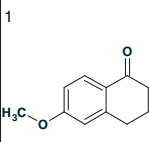
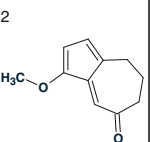
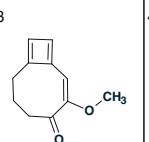
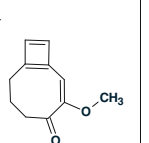
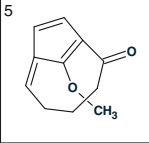
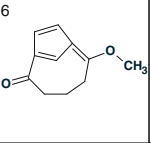
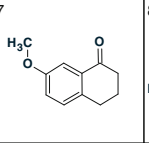
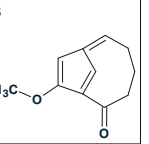
1 	2 	3 	4 
$d_A(^{13}\text{C})$: 0.641 $d_N(^{13}\text{C})$: 0.554 $d_I(^{13}\text{C})$: 0.961	$d_A(^{13}\text{C})$: 8.195 $d_N(^{13}\text{C})$: 7.826 $d_I(^{13}\text{C})$: 6.220	$d_A(^{13}\text{C})$: 8.920 $d_N(^{13}\text{C})$: 10.207 $d_I(^{13}\text{C})$: 8.954	$d_A(^{13}\text{C})$: 8.936 $d_N(^{13}\text{C})$: 9.211 $d_I(^{13}\text{C})$: 9.624
5 	6 	7 	8 
$d_A(^{13}\text{C})$: 9.712 $d_N(^{13}\text{C})$: 9.769 $d_I(^{13}\text{C})$: 9.082	$d_A(^{13}\text{C})$: 9.745 $d_N(^{13}\text{C})$: 9.956 $d_I(^{13}\text{C})$: 7.722	$d_A(^{13}\text{C})$: 9.789 $d_N(^{13}\text{C})$: 9.498 $d_I(^{13}\text{C})$: 8.877	$d_A(^{13}\text{C})$: 9.799 $d_N(^{13}\text{C})$: 11.082 $d_I(^{13}\text{C})$: 11.117

Fig. 3.7 Challenge 1: The eight top structures in the output file. Designation of average deviations: d_A —HOSE code, d_N —Neural Nets, d_I —Increments

nuclei it is necessary to place a cursor on the **Structure List** field and select a command **Select Data...** in the context menu. When the ranked output file is ready for the user to review it is worth producing a ChemSketch document amenable for printing. For this goal, the following command should be executed: **File>Create Report/List of Structures...** In the **Create Report** options, the number of structures that have to be included into the report should be typed (records 1–8 in this case).

Following this process a List of Structures containing the eight items presented in Fig. 3.7 was obtained.

Figure 3.7 shows that all three methods of ^{13}C chemical shift prediction select the first structure (#1) as the most probable one. The difference between deviations calculated for structures #2 and #1 is very large ($\Delta = d(2) - d(1) = 5\text{--}7$ ppm), which is reliable evidence of the correctness of the structure. The priority of the selected structure can be demonstrated visually if one clicks on the icon “ ^{13}C ” marked by a yellow spot on the toolbar. This command colors the carbon atoms in different specific colors depending on the difference between the experimental and calculated chemical shifts for a given atom. If $d \leq 3$ ppm the circle is green, $d \leq 15$ ppm is yellow, and $d \geq 15$ ppm is red.

The ranked structures of the output file supplied with the corresponding colored circles on the atoms and the values of the maximum deviations are presented in Fig. 3.8.

Figure 3.8 convincingly demonstrates the priority of structure #1 (all carbon atoms are highlighted in green). Note that the most similar isomer of structure #1

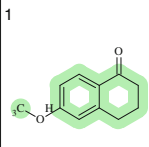
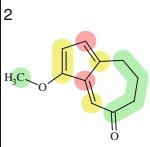
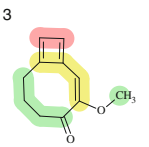
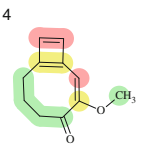
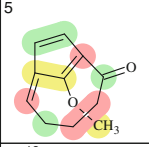
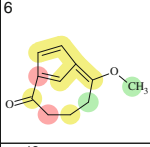
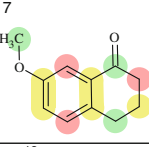
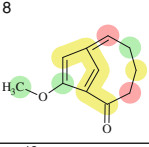
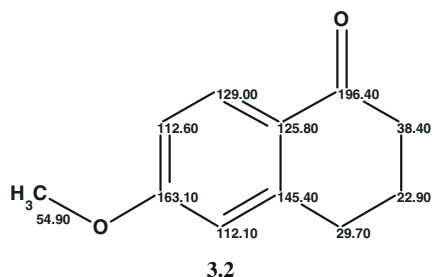
			
$d_A(^{13}\text{C}): 0.641$ $d_N(^{13}\text{C}): 0.554$ $d_I(^{13}\text{C}): 0.961$	$d_A(^{13}\text{C}): 8.195$ $d_N(^{13}\text{C}): 7.826$ $d_I(^{13}\text{C}): 6.220$	$d_A(^{13}\text{C}): 8.920$ $d_N(^{13}\text{C}): 10.207$ $d_I(^{13}\text{C}): 8.954$	$d_A(^{13}\text{C}): 8.936$ $d_N(^{13}\text{C}): 9.211$ $d_I(^{13}\text{C}): 9.624$
			
$d_A(^{13}\text{C}): 9.712$ $d_N(^{13}\text{C}): 9.769$ $d_I(^{13}\text{C}): 9.082$	$d_A(^{13}\text{C}): 9.745$ $d_N(^{13}\text{C}): 9.956$ $d_I(^{13}\text{C}): 7.722$	$d_A(^{13}\text{C}): 9.789$ $d_N(^{13}\text{C}): 9.498$ $d_I(^{13}\text{C}): 8.877$	$d_A(^{13}\text{C}): 9.799$ $d_N(^{13}\text{C}): 11.082$ $d_I(^{13}\text{C}): 11.117$

Fig. 3.8 Challenge 1: The top structures of the output file with *colored circles* illustrating the accuracy of the ^{13}C chemical shift predictions: $d \leq 3$ ppm—green, $d \leq 15$ ppm—yellow, $d \geq 15$ ppm—red

(a competing structure 5.2 considered in the textbook [1, 2]) is characterized by huge deviations and is placed in seventh position by the ranking procedure.

In the first run not all spectroscopic information presented in Table 3.1 is used. Column M(J) in Table 3.1 contains the multiplicities of the signals and the corresponding coupling constants measured by ^1H NMR. The multiplicity values indicate the total number of hydrogen atoms $n(\text{H})$ attached to other carbon atoms existing in the nearest environment of a given carbon atom, $n(\text{H}) = (M - 1)$. For instance, if the multiplicity $M = 1$ (singlet) the value of $n(\text{H})$ equals zero and if $M = 2$ (doublet) then is equal to 1, etc. This information can be used as a set of additional constraints during structure generation. To introduce the $n(\text{H})$ values one has to activate the icon **Edit Atom Properties** on the toolbar in the **MCD** window and type the corresponding numbers in the field **Number of Hydrogens on Neighbor Atoms**. The repeated structure generation utilizing the numbers of hydrogen atoms gave the following result: $k = 2 \rightarrow 1$, $t_g = 0.004$ s. Therefore, the additional structural information carried by the ^1H signal multiplicities generated a single structure with the generation time t_g reduced by a factor of 4.

For the sake of completeness structure generation was repeated also from the initial MCD (Fig. 3.5), without any user edits. The result: $k = 26 \rightarrow 20 \rightarrow 10$, $t_g = 0.070$ s and the structures are the same as shown in Figs. 3.7 and 3.8 but the generation time is five times longer in this case. Of course, when the generation time is measured in milliseconds the difference cannot be noticed, but it becomes significant for problems requiring a long structure generation time. The elucidated structure 3.2 with the associated ^{13}C chemical shift assignment is shown below.

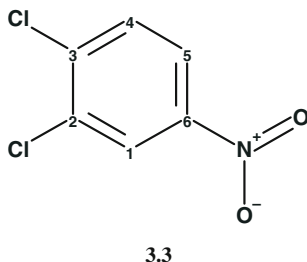


Thus, a problem requiring significant time and intellectual effort to be solved in the traditional manner was solved unambiguously, reliably, and instantaneously with the aid of the StrucEluc expert system.

Since the molecule under investigation is small, the number of isomers corresponding to the molecular formula $C_{11}H_{12}O_2$ (N) was calculated using the Classic Generator. This gave the number to be $N = 555,173,961$. As the number of found candidate structures, k , is equal to 10, we can establish a term μ to represent the moiety of total structural information extracted as a result of automated problem solving. Substituting the variable values N and k in the formula $\mu = (1 - \log k / \log N)$ we will obtain that $\mu = 0.88$ (see Sect. 1.1.1). Therefore, the remaining moiety μ ($^{13}C_{\text{calc}}$) of the structural information necessary for selecting a *single* molecule was extracted using the ^{13}C chemical shift prediction. We see that μ ($^{13}C_{\text{calc}}$) = 0.12 for this case. Since employing the 1H signal multiplicities led to generation of a single and correct structure we can say that the information moiety carried by the 1H multiplicities is also equal to 0.12.

3.2 Challenge 2

The structure of the “unknown”:



The molecular ion m/z 191 suggests the presence of an odd number of nitrogen atoms in the unknown. Along with the molecular ion, $M+2$ and $M+4$ peaks are observed with the intensity ratio $(I(M):I(M+2):M+4) = 100:65:10$, which is a hint to

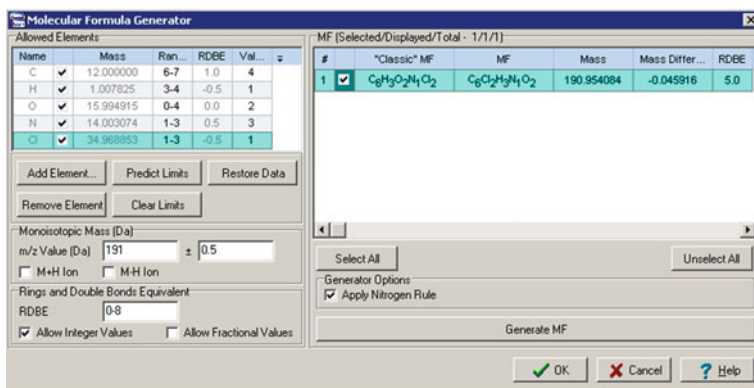


Fig. 3.9 Challenge 2: The **Molecular Formula Generator** window

the presence of two atoms of chlorine. Losses of 46 (NO_2) and 30 (NO) amu in the MS spectrum and two strong IR bands at 1,521 and 1,350 cm^{-1} suggest the presence of an NO_2 group in the molecule. IR bands observed at 1,600 and 1,575 cm^{-1} most probably belong to stretching vibrations of the benzene ring. According to the ^{13}C and ^1H NMR spectra the minimum numbers of carbon and hydrogen atoms are equal to 6 and 3 respectively.

Information regarding the elemental composition of the molecule with the corresponding tolerances postulated for different kinds of chemical elements was input into the **Molecular Formula Generator**. As a result only one molecular formula $\text{C}_6\text{H}_3\text{NO}_2\text{Cl}_2$ was generated (Fig. 3.9).

Note that the presence of an NO_2 group was suggested in the analyzed molecule on the basis of the MS and IR spectra. The ^1H and ^{13}C NMR spectra cannot be used directly for this aim but the StrucEluc system provides a “generalized portrait” procedure which gives hints (not evidence!) regarding the presence of functional groups included in a Typical Functional Group Library (see Sect. 1.3.2.6). This procedure can be demonstrated using the current example.

First a fragment search against the **Fragment Library** was performed using the ^{13}C spectrum of the molecule (**Structure Elucidation\Search Fragments by CNMR spectrum...**). As a result 52 fragments were found (click on **View/Structures Lists/Found Fragments** to see them). Then the **Functional Groups Library** was opened (**Structure Lists\ACD Functional Groups Library**) and the command **Structure Elucidation/Advanced/Search Functional Groups** was initiated. Functional groups were ranked by the frequency of their presence in the 52 Found Fragments appearing in the active window, the functional groups ordered in descending order of the frequency. The top structures of the obtained list of functional groups are presented in Fig. 3.10 that shows that a benzene ring and nitro group are those functionalities that most frequently occur in the Found Fragments. It also turns out that the probable pattern of benzene ring substitution is 1,2,4. At the same time groups are listed which were not found in any of the 52 fragments.

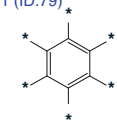
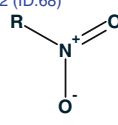
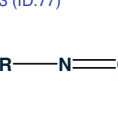

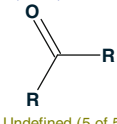
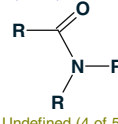
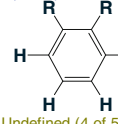
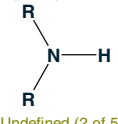
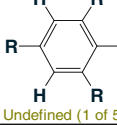
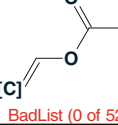
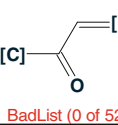
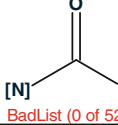

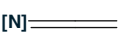
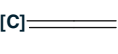
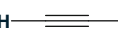
1 (ID:79)  Undefined (13 of 52)	2 (ID:68)  Undefined (10 of 52)	3 (ID:77)  Undefined (10 of 52)	4 (ID:38)  Undefined (8 of 52)
5 (ID:80)  Undefined (5 of 52)	6 (ID:21)  Undefined (4 of 52)	7 (ID:37)  Undefined (4 of 52)	8 (ID:25)  Undefined (2 of 52)
9 (ID:39)  Undefined (1 of 52)	10 (ID:1)  BadList (0 of 52)	11 (ID:3)  BadList (0 of 52)	12 (ID:4)  BadList (0 of 52)
13 (ID:5)  BadList (0 of 52)	14 (ID:6)  BadList (0 of 52)	15 (ID:7)  BadList (0 of 52)	16 (ID:9)  BadList (0 of 52)

Fig. 3.10 Challenge 2: The *top* of the list of Functional Groups with an indication of the frequencies of their occurrence in the **Found Fragments** set

These groups are labeled as belonging to a BadList and, in principle, could be used during the structure generation process if they were included into the **User Badlist**. It is worth noting that conclusions following from the **Functional Group Search** are of a probabilistic character and, in general, they should be carefully checked. Nevertheless, this information helps in the recognition of which functionalities can be included in the structure of the unknown and which are not.

Table 3.2 presents the spectroscopic data extracted from the ^1H , ^{13}C , HSQC, and HMBC spectra of the analyzed compound.

Table 3.2 Challenge 2:
Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M (J) ^a	C HMBC
C1	126.2	125.34	CH	8.39	u	139.3
C2	133.2	134.42	C	–	–	–
C3	139.3	140.01	C	–	–	–
C4	132.5	130.76	CH	7.88	u	147.60, 133.20
C5	124.2	123.34	CH	8.15	u	139.3
C6	147.6	146.79	C	–	–	–

^a Here the symbol “u” denotes that the J_{HH} values were either not measured (undefined) or unused for the structure elucidation by StrucEluc system

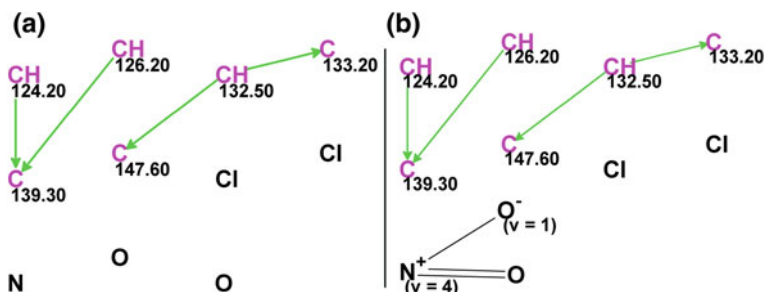


Fig. 3.11 Challenge 2: Initial MCD (a) and edited MCD (b)

To elucidate the structure of the unknown an MCD was created (Fig. 3.11a). As the presence of an NO₂ group is evident from the MS and IR spectra and was confirmed by the **Functional Group Search** this group was drawn on the MCD by hand (an icon **Draw Chemical Bond**) as shown in Fig. 3.11b.

To draw the NO₂ group the properties of the nitrogen and one of the oxygen atoms were edited (**Edit Atom Properties** icon on the toolbar in MCD window) as shown in Fig. 3.12a, b.

No contradictions were detected in the HMBC data using MCD checking and Strict Structure Generation was initiated. It gave the following results: $k = 452 \rightarrow 310 \rightarrow 27$, $t_g = 0.145$ s.

The six top structures of the ranked output file are presented in Fig. 3.13.

Figure 3.13 shows that 1,2-dichloro-4-nitrobenzene (**3.3**) was assigned by the program as the best structure. The next four competing structures contain different substitutions of the benzene ring by the nitro group and the two chlorine atoms.

As the molecular formula of the unknown is rather small the problem can be easily solved by StrucEluc using ¹H, ¹³C, and HSQC spectra only (without HMBC). To switch off the HMBC spectrum all connectivities shown in the MCD (Fig. 3.11b) should be deleted using the **Delete** tool (or the HMBC Spectrum can be disabled in the **Edit Spectrum Query** window and the MCD created anew). In this case the following result was obtained: $k = 1,476 \rightarrow 972 \rightarrow 29$, $t_g = 0.482$ s. In spite of the fact that the initial number of generated structures became significantly larger, the final number of candidates is almost the same due to the high

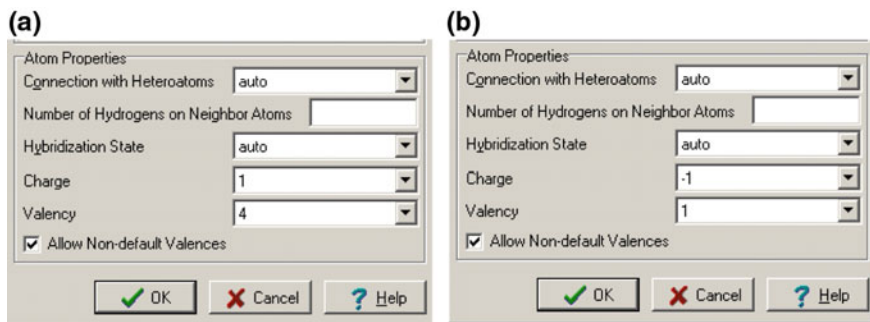


Fig. 3.12 Challenge 2: Edits of nitrogen (a) and oxygen (b) atom properties

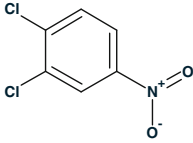
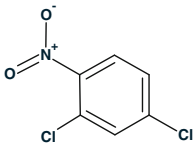
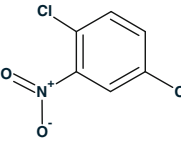
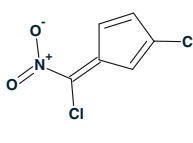
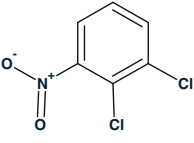
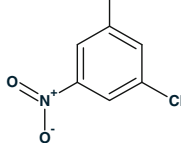
		
$d_A(^{13}\text{C}): 1.033$ $d_N(^{13}\text{C}): 1.035$ $d_I(^{13}\text{C}): 1.291$	$d_A(^{13}\text{C}): 1.933$ $d_N(^{13}\text{C}): 1.969$ $d_I(^{13}\text{C}): 2.699$	$d_A(^{13}\text{C}): 3.588$ $d_N(^{13}\text{C}): 3.868$ $d_I(^{13}\text{C}): 3.856$
		
$d_A(^{13}\text{C}): 4.132$ $d_N(^{13}\text{C}): 5.714$ $d_I(^{13}\text{C}): 5.361$	$d_A(^{13}\text{C}): 4.275$ $d_N(^{13}\text{C}): 4.909$ $d_I(^{13}\text{C}): 3.993$	$d_A(^{13}\text{C}): 4.483$ $d_N(^{13}\text{C}): 4.859$ $d_I(^{13}\text{C}): 4.661$

Fig. 3.13 Challenge 2: The six top ranked structures of the output file

performance of the filtering procedure. A threefold increase in generation time is irrelevant as the t_g value is only fractions of a second.

We have shown above how the task was solved since the presence of an NO_2 group was proven using MS and IR spectra. This structural information, in addition to the NMR data, is desirable to ease the computational calculations, but it is not obligatory to include it. To solve the task without any further involvement of the user we ran the structure generation from the initial MCD (Fig. 3.11a) using the options presented in Fig. 3.14. The following results were obtained: $k = 18,996 \rightarrow 10,109 \rightarrow 1,508$, $t_g = 6$ s, and the two top structures of the ranked output file coincided with those shown in Fig. 3.14.

Figure 3.14 Challenge 2. Options of structure generation from the initial MCD.

The number of generated structures is rather high due to both enumerating all possible valences of a nitrogen atom and producing a great number of improbable structures.

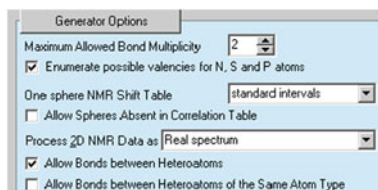
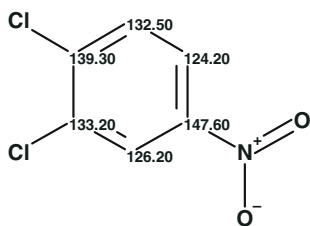


Fig. 3.14 Challenge 2: Options of structure generation from the initial MCD

The answer structure **3.4** with automatic ^{13}C chemical shift assignment is shown below:

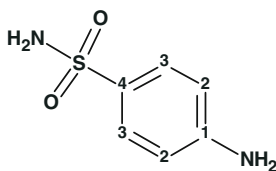


3.4

For the molecular formula $\text{C}_6\text{H}_3\text{NO}_2\text{Cl}_2$, the total number of mathematically conceivable isomers is equal to 2,858,492. We cannot deliberately determine the number of stable isomers. Nevertheless, we found at least the number, N_s , of those isomers which do not contain N–Cl and O–Cl bonds, $N_s = 980,116$. For this goal N–Cl and O–Cl fragments were placed into the **User BadList** during the classic structure generation from the molecular formula. At $k = 29$, the calculated moieties of the extracted structural information are equal to $\mu = 0.76$ (a contribution from both CASE logic and combinatorics) and $\mu (^{13}\text{C}_{\text{calc}}) = 0.24$ (a contribution from the ^{13}C chemical shift prediction).

3.3 Challenge 3

The structure of the “unknown”:



3.5

The general methodology of the StrucEluc-based molecular formula determination from a nominal molecular mass, characteristic ions, NMR, and IR data was demonstrated in the previous challenges. We will now demonstrate utilizing a collection of factual data and the corresponding conclusions to allow for the inference of the molecular formula in a table (see Table 3.3).

To determine the structure of the unknown, the ^1H , ^{13}C , HSQC, and HMBC data that are presented in Table 3.4 were used. The number of equivalent carbon atoms is shown in brackets, e.g., C2 (2). Carbon atoms are considered equivalent if they

Table 3.3 Challenge 3: Inferring a molecular formula

Factual data	Conclusions
MS: $[M]^+ = 172$	Number of nitrogen atoms is even, $n_N = 0$ or 2 or 4...
MS: $[M+2] = 174$ (5.1 %)	Presence of S is possible
MS: $\Delta = 172 - 108 = 64$	SO ₂ is highly probable
MS: $\Delta = 172 - 156 = 16$	NH ₂ is possible
IR: 3478, 3376, 3269 cm ⁻¹	NH ₂ , NH, OH are possible
IR: no absorption characteristics for CN or NO ₂	CN, NO ₂ are not probable
IR: 1597, 1505 cm ⁻¹	Benzene ring is possible
¹³ C: 4 signals are observed	Minimum number of carbon atoms is 4
¹³ C and ¹ H	Minimum number of hydrogen atoms is 8
Ranges of atom numbers	C (4–6), H (8–10), O (2–5), N (0–4), S (1)
Molecular Formula Generated	C₆H₈N₂O₂S

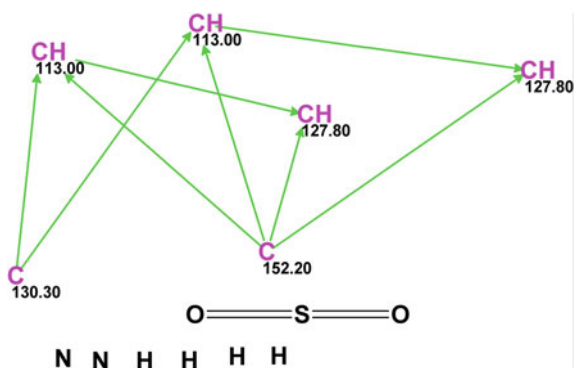
Table 3.4 Challenge 3: Spectroscopic NMR data

Label	δC	δC_{calc}	CH _n	δH	M(J)	C HMBC
C1	152.2	152.01	C	–	–	–
C2(2)	113	115.07	CH	6.6	u	C1, C4
C3(2)	127.8	128.59	CH	7.45	u	C1, C2
C4	130.3	133.87	C	–	–	–

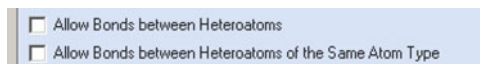
have identical ¹³C chemical shifts and multiplicities and the protons attached to these carbons have identical chemical shifts.

An MCD was created and an SO₂ group whose presence is highly probable was drawn by hand on the MCD (see Fig. 3.15).

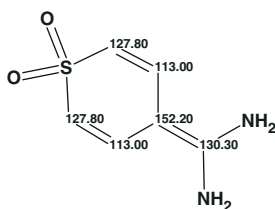
A first run using Strict Structure generation was initiated under the most common condition that chemical bonding between heteroatoms is forbidden. This constraint is used frequently for the first run because (i) bonds between heteroatoms are present in a relatively restricted number of chemical classes of organic

Fig. 3.15 Challenge 3: Molecular connectivity diagram. An SO₂ functional group was drawn by hand

compounds, (ii) when these bonds are allowed structure generation usually produces a huge number of structures having no chemical sense, and processing time can be large, (iii) if the ^{13}C NMR chemical shift prediction followed by structure file ranking shows that the solution obtained is likely incorrect (because of its very large average deviations), this hints at the fact that bonds between heteroatoms were mistakenly forbidden. To control these constraints the following check boxes are selected or deselected in the **CSB Generator Options** window:

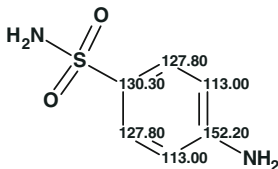


With the options as shown above the following result was obtained: $k = 1$, $t_g = 0.005$ s. The unique structure **3.6** had received deviations in the range 20–24 ppm and maximum deviations of 55 ppm.



3.6

It is evident that structure **3.6** is wrong as the average and maximum deviations are huge. The check box “**Allow Bonds between Heteroatoms**” was selected, and structure generation was run again with the result: $k = 6 \rightarrow 6 \rightarrow 3$, $t_g = 0.005$ s. The ranked output file is presented in Fig. 3.16, which shows that the correct structure (**3.7**) was convincingly established (compare the deviations calculated for structure #1 and #2).



3.7

It is worth noting that when the HMBC spectrum was ignored the same final solution was obtained: $k = 12 \rightarrow 12 \rightarrow 3$, $t_g = 0.01$ s. This is a consequence of the small size of the molecule.

In this example we were fortunate to determine the $\text{O}=\text{S}=\text{O}$ group from the MS spectrum. Unfortunately, this group cannot be determined directly from NMR spectra but, nevertheless, the CASE methodology in principle allows for the

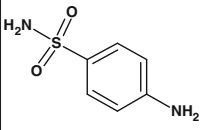
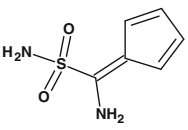
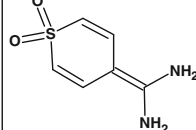
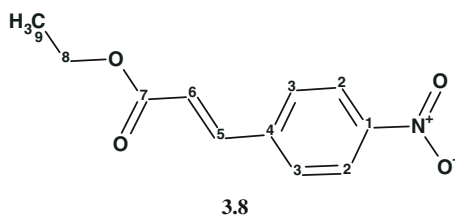
1	2	3
		
$d_A(^{13}\text{C}): 1.580$ $d_N(^{13}\text{C}): 1.838$ $d_I(^{13}\text{C}): 1.612$	$d_A(^{13}\text{C}): 10.775$ $d_N(^{13}\text{C}): 11.035$ $d_I(^{13}\text{C}): 11.317$	$d_A(^{13}\text{C}): 23.818$ $d_N(^{13}\text{C}): 28.445$ $d_I(^{13}\text{C}): 20.182$

Fig. 3.16 Challenge 3: The ranked output file obtained when bonds between heteroatoms were allowed

elucidation of molecules containing $\text{O}=\text{S}=\text{O}$ without any prompts coming from additional sources of structural information. The initial MCD was created repeatedly and structure generation was run with permission to enumerate the S and N valences. To exclude deliberately unstable structures, the substructures $\text{O}-\text{N}-\text{O}$, $\text{O}-\text{O}-\text{N}$, and $\text{O}-\text{O}$ were forbidden, for which the “molecules” $\text{OH}-\text{NH}-\text{OH}$, $\text{OH}-\text{O}-\text{NH}_2$ and, $\text{OH}-\text{OH}$ were placed in the User BadList. Results: $k = 11,057 \rightarrow 4,567 \rightarrow 2,330$, $t_g = 8$ s. ^{13}C NMR chemical shift prediction and structure ranking allowed us again to select structure **1** as the most probable one. It should be noted that the majority of generated structures contained atom combinations which make these structures senseless.

3.4 Challenge 4

The structure of the “unknown”:



In this challenge only six signals are confidently observed in the ^{13}C NMR spectrum. As 11 carbon atoms are present in the generated molecular formula careful inspection of the ^{13}C NMR spectrum allows three additional very weak signals to be identified at 140.5, 148.4, and 165.9 ppm from quaternary carbons (Table 3.5).

Utilization of the Generalized Portrait procedure also indicated likely issues which can exist in the structure of the molecule analyzed. 210 fragments were

Table 3.5 Challenge 4: Inferring the molecular formula

Factual data	Conclusions
MS: $[M]^+ = 221$	Number of nitrogen atoms is odd, $n_N = 1$ or 3 or 5...
MS: no $[M+2]$	Presence of S is impossible
MS: no isotope pattern for Cl and Br	Cl and Br are impossible
^{13}C : no multiplets from the presence of F	F is impossible
IR: 1,518 and 1,342 cm^{-1} , very strong	NO_2 is possible
IR: no absorption from CN, NH, NH_2 , OH	CN, NH, NH_2 , OH are not probable
IR: 3108, 3080, 1599, 1505 cm^{-1} , 1714 (strong), 1645 cm^{-1} (medium)	Benzene ring, C=O, C=C (alkene)
^{13}C : 6 signals are observed, 2 CH are equivalent (1,4-Ar)	Minimum number of carbon atoms is 8
^{13}C and ^1H : Hydrogen count identified	Minimum number of hydrogen atoms is 11
Ranges of atom numbers	C (8–12); H (11–12); O (3–5), N (1)
Molecular Formula Generated	$\text{C}_{11}\text{H}_{11}\text{NO}_4$

found as a result of the **Fragment Search** and functional groups were “sifted” through the **Found Fragments** set. The most probable functional groups from the **ACD Functional Group Library** are presented in Fig. 3.17.

Comparison of structure **3.8** with Fig. 3.1 demonstrates that the top ranked functional groups (items #1, #2, #3, #5, and #6) really exist in the structure of the unknown. It should be emphasized again that the described methodology gives only a hint to the possibility of selected functional groups being present but it does not prove their existence in the molecule.

To elucidate the structure 1D, HSQC, HMBC, and COSY data collected in [1] were used (Table 3.6).

The MCD is presented in Fig. 3.18.

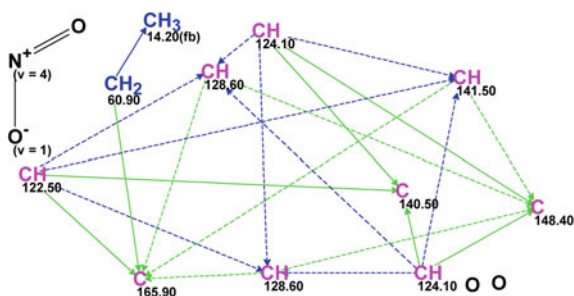
The NO_2 group was drawn by hand on the MCD as its presence in the molecule seems fairly probable. Note that all COSY connectivities except one (between CH_3

1 (ID:80) GoodList (118 of 210) Formula: CO?	2 (ID:63) Undefined (32 of 210) Formula: C3H5O	3 (ID:79) Undefined (28 of 210) Formula: C6?	4 (ID:77) Undefined (20 of 210) Formula: NO?	5 (ID:42) Undefined (18 of 210) Formula: C6H4?
6 (ID:68) Undefined (18 of 210) Formula: NO2?	7 (ID:28) Undefined (17 of 210) Formula: CH3O?	8 (ID:25) Undefined (10 of 210) Formula: HN?	9 (ID:43) Undefined (7 of 210) Formula: C6H5?	10 (ID:21) Undefined (6 of 210) Formula: CNO?

Fig. 3.17 Challenge 4: Most probable functionalities selected using the **Generalized Portrait** procedure

Table 3.6 Challenge 4: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	C HMBC
C1	148.4	147.9	C	–	–	–	–
C2(2)	124.1	124.03	CH	8.21	u	7.66	C4, C1
C3(2)	128.6	129.6	CH	7.66	u	–	–
C4	140.5	139.84	C	–	–	–	–
C5	141.5	140.61	CH	7.66	u	8.21, 6.53	C7, C1
C6	122.5	121.88	CH	6.53	u	7.66	C7, C4
C7	165.9	167.72	C	–	–	–	–
C8	60.9	60.9	CH_2	4.26	u	1.32	C7
C9	14.2	14.89	CH_3	1.32	u	4.26	–

Fig. 3.18 Challenge 4: Molecular connectivity diagram. The NO_2 group is drawn by hand

and CH_2) and several HMBC connectivities are marked by dotted lines. Dotted lines are used to denote ambiguous connectivities which arise due to signal overlap in the ^{13}C and ^1H NMR spectra. In the current case the nonequivalent carbon atoms C_3 and C_5 have attached hydrogens resonating at the same chemical shifts (7.66 ppm, see Table 3.6). On the basis of general CASE principles we can expect that the number of generated structures and the generation time are greater in the presence of ambiguous connectivities because ambiguous constraints admit some atom combinations that could be forbidden if the constraints were strict. This difference becomes significant when structures are large (we will see many examples in Part III), but for the molecule given here the difference can hardly be noticed. Strict Structure Generation was run using the MCD shown in Fig. 3.18 and the following results were obtained: $k = 14 \rightarrow 7 \rightarrow 2$, $t_g = 0.1$ s. The ranked structures are presented in Fig. 3.19.

The correct structure was reliably selected by deviations calculated using all three methods. To shed light on the role of COSY and HMBC correlations used for structure elucidation the second run was initiated from an MCD created without the COSY and HMBC spectra (the spectra were deselected in the **Edit Spectrum Query** window for creating MCD). Results: $k = 4,592 \rightarrow 1,688 \rightarrow 163$, $t_g = 1$ m

1	2
$d_A(^{13}\text{C}): 0.665$ $d_N(^{13}\text{C}): 0.669$ $d_I(^{13}\text{C}): 0.840$ $\text{max}_dA(^{13}\text{C}): 1.820$	$d_A(^{13}\text{C}): 4.116$ $d_N(^{13}\text{C}): 4.474$ $d_I(^{13}\text{C}): 5.333$ $\text{max}_dA(^{13}\text{C}): 14.980$

Fig. 3.19 Challenge 4: Ranked structures obtained as a result of structure generation from 1D, HSQC, HMBC, and COSY data

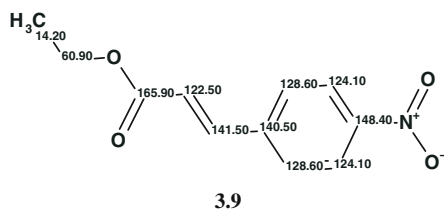
18 s. The result visually shows the great importance of the structural information carried even by ambiguous COSY and HMBC correlations: the number of non-identical structures increased by a factor of 80 while the number of initially generated structures—by 300. The generation time increased 800 times. The four top ranked structures of the output file are shown in Fig. 3.20. It is interesting to note that the average deviations characterizing structures #2–#4 are less than the deviations calculated for the second structure presented in Fig. 3.19. This is explained by the fact that structures #2–#4 could be generated only due to the absence of constraints imposed by the COSY and HMBC data. For instance, the CH_3 and CH_2 groups are connected by an unambiguous COSY connectivity (see Fig. 3.18) and therefore can be connected by a chemical bond, but as seen on structures #2 and #3, the distance between CH_2 and CH_3 groups is of 3 and 4 bonds for #2 and #3 correspondingly.

In this task the presence of an NO_2 group was evident from the IR spectrum which was confirmed by the Functional Library Search in the set of Found Fragments. Nevertheless, the right solution to the problem can also be obtained in an “ab initio” mode without drawing a nitro group in the MCD window. For this purpose, an MCD was created again and structure generation was repeated with the ticked boxes “Enumerate Possible Valences” and “Allow Bonds between Heteroatoms” in the generation options. Results: $k = 136 \rightarrow 74 \rightarrow 44$, $t_g = 1.7$ s. The top

1	2	3	4
$d_A(^{13}\text{C}): 0.665$ $d_N(^{13}\text{C}): 0.669$ $d_I(^{13}\text{C}): 0.840$	$d_A(^{13}\text{C}): 2.969$ $d_N(^{13}\text{C}): 3.312$ $d_I(^{13}\text{C}): 3.509$	$d_A(^{13}\text{C}): 3.365$ $d_N(^{13}\text{C}): 3.742$ $d_I(^{13}\text{C}): 2.120$	$d_A(^{13}\text{C}): 3.653$ $d_N(^{13}\text{C}): 4.097$ $d_I(^{13}\text{C}): 3.899$

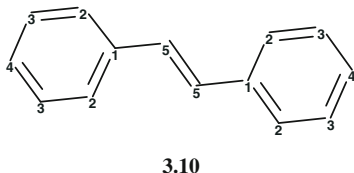
Fig. 3.20 Challenge 4: Ranked structures obtained as a result of structure generation from 1D and HSQC data

two structures of the ranked output file coincided with those shown in Fig. 3.19. Structure 3.9 maps the elucidated structure with the ^{13}C chemical shift assignment performed by the program:



3.5 Challenge 5

The structure of the “unknown”:



The presence of only four signals in the ^{13}C spectrum while the molecular formula contains 14 carbon atoms indicates high symmetry within the molecule which indeed has a center of symmetry. One can therefore expect that the stretching vibrations of the alkene double bond will be inactive in infrared spectroscopy and they will be fairly intense in the Raman spectrum. The data displayed in Table 3.7 confirm this expectation: the IR spectrum clearly shows the presence of aromatic rings while the double bond situated in the center of molecular symmetry was detected from the Raman spectrum.

Table 3.7 Challenge 5: Molecular Formula Inference

Factual data	Conclusions
MS: $[\text{M}]^+ = 180$	No hints regarding heteroatoms
^{13}C : 4 signals	Symmetry. Min. $n_{\text{C}} = 4$
^1H : 3 signals, $n_{\text{rel}}(\text{H}) = 2:3:1$	Min. n_{H} is 6
Raman: 1642, 1598 cm^{-1} IR: 1598, 1539, 1496	C=C alkene, Aromatics Aromatics
Ranges of atom numbers	C (4–20), H (6–18), RDB (0–10)
Molecular Formula Generated	C₁₄H₁₂

Table 3.8 Challenge 5: Spectroscopic NMR data

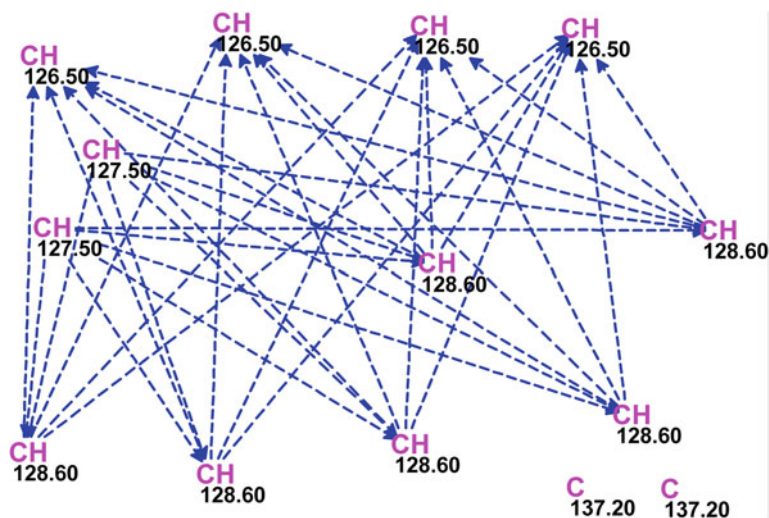
Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY
C1(2)	137.2	137.72	C	–	–	–
C2(4)	126.5	126.54	CH	7.66	u	7.5
C3(4)	128.6	128.87	CH	7.5	u	7.66, 7.40
C4(2)	127.5	127.53	CH	7.4	u	7.5
C5(2)	128.6	127.67	CH	7.25	u	–

To perform structure elucidation the ^{13}C , ^1H , HSQC, and COSY NMR spectra (Table 3.8) obtained from textbook [1, 2] were used.

A peculiarity of the data presented in Table 3.8 is that four carbons belong to one group of equivalent atoms (C3) and a pair of carbons belonging to another group (C5) have identical ^{13}C chemical shifts (128.6), while their attached hydrogens, as determined from the HSQC spectrum are characterized by different ^1H chemical shifts (7.5 and 7.25 ppm correspondingly). The coincidence of the ^{13}C chemical shifts of the carbons belonging to different groups of equivalent atoms is accidental, but as a result 6 of the 14 carbons have identical chemical shifts.

The MCD is presented in Fig. 3.21.

We see that *all* COSY connectivities are ambiguous and this means that constraints imposed by the COSY data cannot play a significant role in structure generation. Structure generation was completed with the following result: $k = 23 \rightarrow 19 \rightarrow 4$, $t_g = 0.464$ s. It should be noted that structure generation of symmetric molecules from 2D NMR data is a challenge for expert systems and as far as we know only Structure Elucidator is capable of dealing with those molecules. The ranked output file is presented in Fig. 3.22 and shows that the correct

**Fig. 3.21** Challenge 5: Molecular connectivity diagram

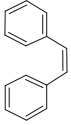
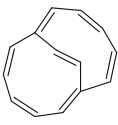
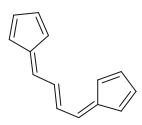
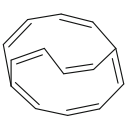
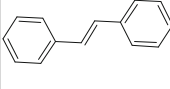
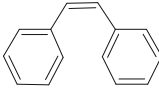
1 	2 	3 	4 
$d_A(^{13}\text{C}): 1.124$ (v.14.) $d_N(^{13}\text{C}): 0.503$ $d_I(^{13}\text{C}): 0.615$	$d_A(^{13}\text{C}): 3.349$ (v.14.) $d_N(^{13}\text{C}): 2.417$ $d_I(^{13}\text{C}): 4.417$	$d_A(^{13}\text{C}): 5.571$ (v.14.) $d_N(^{13}\text{C}): 5.705$ $d_I(^{13}\text{C}): 2.506$	$d_A(^{13}\text{C}): 27.886$ (v.14.) $d_N(^{13}\text{C}): 9.948$ $d_I(^{13}\text{C}): 7.233$

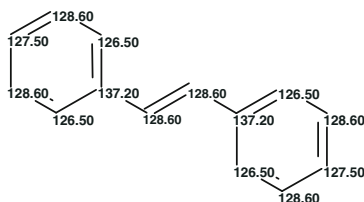
Fig. 3.22 Challenge 5: The ranked output file where structure #1 is in the *cis* form

Fig. 3.23 Challenge 5: The *cis*- and *trans*isomers of the correct structure with their calculated average and maximum ^{13}C and ^1H chemical shift deviations

1 	2 
$d_A(^{13}\text{C}): 0.300$ $d_I(^{13}\text{C}): 0.520$ $d_N(^{13}\text{C}): 0.531$ $\text{max_}d_A(^{13}\text{C}): 0.930$ $\text{max_}d_I(^{13}\text{C}): 1.285$ $\text{max_}d_N(^{13}\text{C}): 1.562$	$d_A(^{13}\text{C}): 1.124$ $d_I(^{13}\text{C}): 0.615$ $d_N(^{13}\text{C}): 0.503$ $\text{max_}d_A(^{13}\text{C}): 2.910$ $\text{max_}d_I(^{13}\text{C}): 1.285$ $\text{max_}d_N(^{13}\text{C}): 1.562$

structure is reliably selected by its average deviations. However, it would be desirable to establish which kind of isomer—*cis* or *trans*—is the most probable. We can expect that the real configuration would be selected on the basis of comparison of ^{13}C chemical shift deviations calculated for the *cis* and *trans* isomers. For this goal structure #1 was transformed into the *trans* configuration and the ^{13}C chemical shift calculation was repeated. Both isomers with their average and maximum deviations calculated for the ^{13}C and ^1H chemical shifts are shown in Fig. 3.23 and this leads us to conclude that a *trans* configuration is the most probable. This conclusion is in accordance with a choice made in the textbook [1] where a *trans* configuration was deduced from a comprehensive analysis of the UV spectrum.

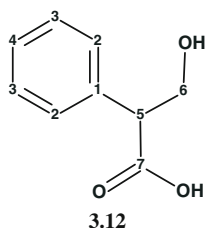
For the sake of completeness structure generation was also run from the MCD for which the COSY correlations were switched off. Result: $k = 29 \rightarrow 25 \rightarrow 4$, $t_g = 0.497$ s. The same four structures were obtained in almost the same time on the basis of the 1D NMR and HMQC data. This is a consequence of a very high degree of ambiguity intrinsic for COSY connectivities. The resulting structure 3.11 along with the assigned ^{13}C chemical shifts is shown below:



3.11

3.6 Challenge 6

The structure of the “unknown”:



3.12

The conclusions obtained regarding the functional groups listed in the table were additionally verified using the **Generalized Portrait Procedure**. As a result of a **Fragment Search**, 30 fragments were selected and the top functional groups of the **Generalized Portrait** are presented in Fig. 3.24. We see that the presence of OH, C=O groups, and a benzene ring are also confirmed by the procedure utilized (Table 3.9).

The 1D, HMQC, and HMBC NMR data available are presented in Table 3.10. Figure 3.25 shows the MCD.

No MCD edits were made and structure generation delivered the following result: $k = 2 \rightarrow 1$, $t_g = 0.001$ s. We see that the single structure 3.13 is identical to

1 (ID:28) GoodList (14 of 30)	2 (ID:80) GoodList (10 of 30)	3 (ID:43) Undefined (8 of 30)	4 (ID:79) Undefined (8 of 30)	5 (ID:41) Undefined (2 of 30)
6 (ID:42) Undefined (2 of 30)	7 (ID:37) Undefined (1 of 30)	8 (ID:40) Undefined (1 of 30)	9 (ID:1) BadList (0 of 30)	10 (ID:3) BadList (0 of 30)

Fig. 3.24 Challenge 6: The top of the **Functional Group** list with the numbers of fragments containing the corresponding functional groups (in *brackets*)

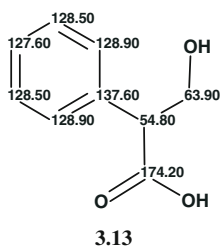
Table 3.9 Challenge 6: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 166, even number	$n_N = 0$ or 2 or 4
^{13}C : 4 CH groups, CH_2	Min. $n_H = 6$
^{13}C : 7 signals	Min. $n_C = 7$, Aromatics are possible.
^1H : at least 9 H atoms are suggested	Min. $n_H = 9$
IR: 3400, 2600–3100 (broad), 1711, 1584, 1497 cm^{-1}	OH (alc.), NH (0 or 2), C=O, COOH, min. $n_O = 3$, Aromatics
Ranges of atom numbers	C (7–12), H (9–11), O (3–5), N (0–2), DBE (5–6)
Molecular Formula Generated	$\text{C}_9\text{H}_{10}\text{O}_3$

Table 3.10 Challenge 6: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	137.6	136.47	C	–	–	–
C2(2)	128.9	128.78	CH	7.25	u	–
C3(2)	128.5	127.96	CH	7.25	u	–
C4	127.6	127.94	CH	7.25	u	–
C5	54.8	54.93	CH	3.65	u	C7, C6, C2, C1
C6	63.9	63.78	CH_2	3.93	u	C1, C7
C6	63.9	63.78	CH_2	3.58	u	C1, C7, C5
C7	174.2	177.37	C	–	–	–

structure **3.12** (and the associated ^{13}C chemical shift assignment is shown) while the average deviations were less than 1 ppm.



Since the molecule is small the structure generation was repeated with the HMBC data switched off with the result: $k = 46 \rightarrow 38 \rightarrow 19$, $t_g = 1.05$ s. The three top structures of the ranked file are shown in Fig. 3.26.

The correct structure was again reliably selected by its average deviations while the difference in generation time (0.001 and 1.05 s) is so small as to not be noticeable.

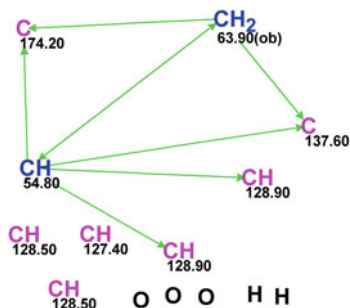


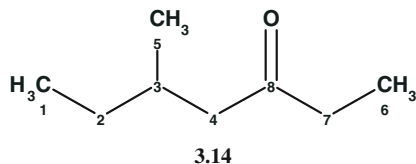
Fig. 3.25 Challenge 6: Molecular connectivity diagram

1	2	3
$d_A(^{13}\text{C}): 0.690$ $d_N(^{13}\text{C}): 0.750$ $d_I(^{13}\text{C}): 0.910$	$d_A(^{13}\text{C}): 3.824$ $d_N(^{13}\text{C}): 3.373$ $d_I(^{13}\text{C}): 3.292$	$d_A(^{13}\text{C}): 4.916$ $d_N(^{13}\text{C}): 4.156$ $d_I(^{13}\text{C}): 5.011$

Fig. 3.26 Challenge 6: The three top structures of the solution found from the 1D and HSQC data

3.7 Challenge 7

The structure of the “unknown” (Table 3.11):



The molecular formula of the compound is so small and simple that one can expect that the structural information carried by the ^{13}C NMR spectrum (Table 3.12) is enough for structure elucidation.

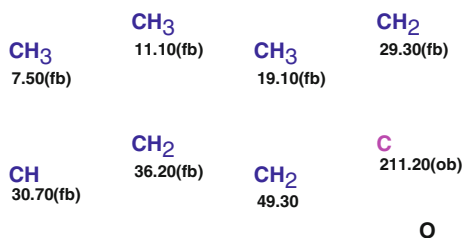
The MCD containing only the skeletal atoms (structural blocks) is presented in Fig. 3.27. The signal multiplicities were determined from the ^{13}C , ^1H , and HSQC data and the carbon atom hybridization states and the possibilities of neighboring

Table 3.11 Challenge 7: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 128, even number	$n_N = 0$ or 2 or 4, no Cl, Br, F, S
^{13}C : 8 signals	Min. $n_C = 8$
^1H : at least 16 H atoms	Min. $n_H = 16$
IR: no bands at 3100–3500, 2200 cm^{-1} , presence of 1,715 cm^{-1}	no OH, no NH, no CN, no NO_2 , presence of C=O, min $n_O = 1$
Ranges of atom numbers	C (8–10), H (16–18), O (1–2), N (0–2), DBE (1–2)
Molecular Formula Generated	$\text{C}_8\text{H}_{16}\text{O}$

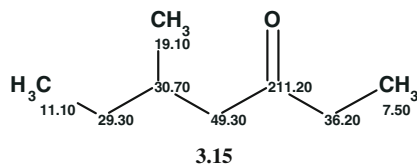
Table 3.12 Challenge 7: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n
C1	11.1	11.47	CH_3
C2	29.3	29.28	CH_2
C3	30.7	30.56	CH
C4	49.3	49.28	CH_2
C5	19.1	20.06	CH_3
C6	7.5	7.68	CH_3
C7	36.2	37.32	CH_2
C8	211.2	212.03	C

Fig. 3.27 Challenge 7: Molecular connectivity diagram

heteroatoms were assigned by the program using the **Atom Property Correlation Table (APCT)**.

Structure generation gave the following results: $k = 360 \rightarrow 360 \rightarrow 13$, $t_g = 0.15$ s. The three top structures of the ranked output file are shown in Fig. 3.28 which confidently identifies the correct structure as **3.15**.



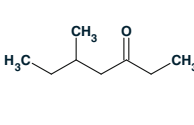
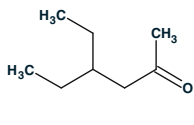
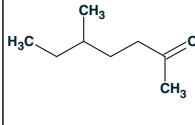
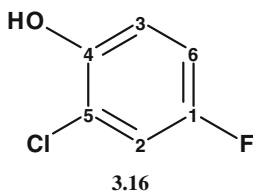
1	2	3
		
$d_A(^{13}\text{C}): 0.455$ $d_N(^{13}\text{C}): 0.555$ $d_I(^{13}\text{C}): 0.466$	$d_A(^{13}\text{C}): 4.529$ $d_N(^{13}\text{C}): 4.364$ $d_I(^{13}\text{C}): 4.826$	$d_A(^{13}\text{C}): 5.092$ $d_N(^{13}\text{C}): 5.236$ $d_I(^{13}\text{C}): 5.305$

Fig. 3.28 Challenge 7: The three top structures of the ranked output file

As the number of isomers corresponding to the molecular formula $\text{C}_9\text{H}_{16}\text{O}$ (N) is equal to 1,684 (all of them are probably stable) and the number of found candidate structures $k = 13$, we can state that 99.2 % of possible isomers were rejected as a result of the application of StrucEluc, while the term describing the moiety of extracted structural information μ is equal to 0.65 (see Part I, Sect. 1.1.1). Therefore, ^{13}C chemical shift prediction allowed us to extract the remaining fraction μ ($^{13}\text{C}_{\text{calc}}$) of the structural information necessary for selection of a single structure—5-methylheptan-3-one. This moiety equals 0.35.

3.8 Challenge 8

The structure of the “unknown”:



The ranges of the atom numbers were selected to take into account the possibility of the presence of undetected carbon and hydrogen atoms, the possibility of two NH groups and a possible OH group and to exclude any doubts regarding the presence of chlorine (Table 3.13). As a result a single molecular formula $\text{C}_6\text{H}_4\text{OCIF}$ was generated matching that from the original work. As in Challenge 7 the molecule is so small that it makes no sense to use the COSY data when StrucEluc is used to identify the unknown. Therefore, only the data taken from the ^{13}C NMR spectrum are included into Table 3.14.

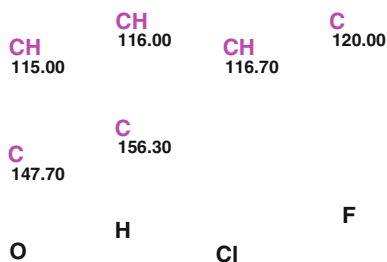
The MCD layout of atoms only is presented in Fig. 3.29.

Table 3.13 Challenge 8: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+ = 146$ ($I = 100\%$), even number	$n_N = 0$ or 2 or 4
MS: $[M+2] = 148$ ($I = 32\%$); $[M]/[M+2] = 3:1$	Presence of Cl atom is possible. No S
^{13}C : 6 signals, all are doublets due to ^{13}C \backslash ^{19}F coupling	Min. $n_C = 6$ Fluorine atom present
^1H : 4 signals, at least 4 hydrogen atoms	Min. $n_H = 4$
IR: 3537, 3460, 1612, 1493 cm^{-1}	OH or NH, no CN, no NO_2 , no C=O. Aromatics, min $n_O = 1$
Ranges of atom numbers	C (6–7), H (4–6), O (1–2), Cl (0–1), F (1), N (0–2), DBE (4–5)
Molecular Formula Generated	$\text{C}_6\text{H}_4\text{OClF}$

Table 3.14 Challenge 8: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n
C1	156.3	156.79	C
C2	115	117.49	CH
C3	116.7	117.2	CH
C4	147.7	149.58	C
C5	120	122.4	C
C6	116	115.29	CH

**Fig. 3.29** Challenge 5.8: Molecular connectivity diagram. The COSY data analyzed in the textbook [1] are not used

Structure generation using the MCD shown in Fig. 3.29 gave the following results: $k = 1,380 \rightarrow 900 \rightarrow 61$, $t_g = 0.42$ s. After the output file is ranked in the usual manner, isomeric structures differing only by the substituent positions on the benzene ring were collected in a table along with their average deviations and positions in the ranked file (Fig. 3.30). Figure 3.30 shows that the best structure #1 coincides with structure 3.16 deduced in the textbook. In the textbook [1, 2], detailed analysis

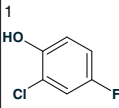
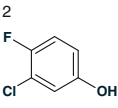
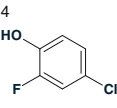
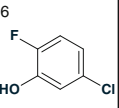
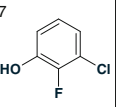
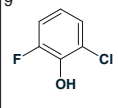
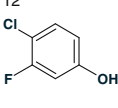
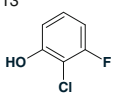
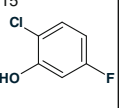
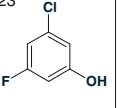
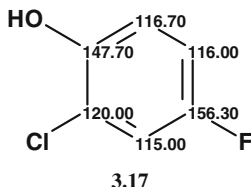
1 	2 	4 	6 	7 
$d_A(^{13}\text{C}): 1.412$ $d_N(^{13}\text{C}): 1.062$ $d_I(^{13}\text{C}): 1.678$	$d_A(^{13}\text{C}): 2.305$ $d_N(^{13}\text{C}): 1.699$ $d_I(^{13}\text{C}): 2.143$	$d_A(^{13}\text{C}): 4.172$ $d_N(^{13}\text{C}): 4.393$ $d_I(^{13}\text{C}): 4.956$	$d_A(^{13}\text{C}): 4.282$ $d_N(^{13}\text{C}): 4.345$ $d_I(^{13}\text{C}): 4.195$	$d_A(^{13}\text{C}): 4.418$ $d_N(^{13}\text{C}): 4.876$ $d_I(^{13}\text{C}): 4.438$
9 	12 	13 	15 	23 
$d_A(^{13}\text{C}): 5.062$ $d_N(^{13}\text{C}): 4.643$ $d_I(^{13}\text{C}): 4.074$	$d_A(^{13}\text{C}): 6.977$ $d_N(^{13}\text{C}): 7.996$ $d_I(^{13}\text{C}): 7.959$	$d_A(^{13}\text{C}): 7.232$ $d_N(^{13}\text{C}): 7.514$ $d_I(^{13}\text{C}): 7.059$	$d_A(^{13}\text{C}): 7.740$ $d_N(^{13}\text{C}): 8.560$ $d_I(^{13}\text{C}): 8.359$	$d_A(^{13}\text{C}): 9.768$ $d_N(^{13}\text{C}): 9.967$ $d_I(^{13}\text{C}): 10.620$

Fig. 3.30 Challenge 8: Isomeric structures differing only by the positions of the substituents

of ^{13}C and ^1H chemical shifts and coupling constants J_{HH} , J_{CH} , J_{FC} , and J_{FH} allowed the authors to reduce the number of alternative structures down to two isomers that are shown in Fig. 3.30 as #1 and #15. The correct structure #1 was distinguished on the basis of comparing experimental and calculated ^{13}C chemical shifts with the measured J_{CH} taken into account.

The expert system has generated all possible isomers without any detailed analysis of the NMR spectrum parameters. It convincingly distinguished the correct structure (#1) from all others (#15) by considering the large difference of their deviations (~ 1.5 and 8 ppm correspondingly).

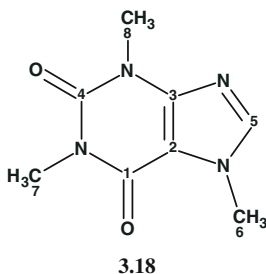
The correct structure **3.17** with assigned chemical shifts:



The number N_s of isomers corresponding to the molecular formula $\text{C}_6\text{H}_4\text{OCIF}$ was calculated under the following constraints: O–F and O–Cl bonds were placed into the User BadList and the Structural Filter was used during classic structure generation. As a result it was found that $N_s = 12,000$. Hence at $k = 61$ the calculated moieties of extracted structural information are equal to $\mu = 0.44$ (the contribution from CASE methodology) and $\mu(^{13}\text{C}_{\text{calc}}) = 0.56$ (the contribution from ^{13}C chemical shift prediction). We see that ^{13}C chemical shift prediction has played a decisive role in the unambiguous structure elucidation of the 2-chloro-4-fluorophenol molecule because neither HMBC nor COSY data were used by the program for structure elucidation.

3.9 Challenge 9

The structure of the “unknown”:



The IR spectrum shows a strong H_2O band at $3,440\text{ cm}^{-1}$ which can mask the absorption of the OH/NH groups. The range of the number of hydrogen atoms was set to 10–11 while the range for the number of carbon atoms was set as 8–9 to take into account the possible presence of a very weak (not observed) signal from a quaternary carbon atom (Table 3.15). The numbers of oxygen and nitrogen atoms were allowed to vary over a wide range. Under these conditions two conceivable molecular formulae— $\text{C}_9\text{H}_{10}\text{O}_3\text{N}_2$ and $\text{C}_8\text{H}_{10}\text{O}_2\text{N}_4$ —were produced. StrucEluc allows the true molecular formula of the unknown to be determined during the structure elucidation process. For this goal both alternative molecular formulae were selected in the **Molecular Formula Generator** dialog window used for structure generation (see Fig. 3.31).

In order to perform the computer-assisted structure elucidation the data presented in Table 3.16 were used. To try the two alternative molecular formulae it is possible to create two **MCDs** and run structure generation in one of the following three modes: (i) perform structure generation sequentially from both MCDs “in one run” and get a *unified* structural file (**Structure Elucidation\Run CSB Generator...**),

Table 3.15 Challenge 9: Molecular Formula Inference

Factual data	Conclusions
MS: $[\text{M}]^+ = 194$ (I = 100 %), even number	$n_{\text{N}} = 0$ or 2 or 4...
MS: $[\text{M}+2]^+$ absent	No S, no Cl
MS: $109 - 82 = 27$, $82 - 55 = 27$	Loss of 2 HCN, $n_{\text{N}} = 2$ or 4 or 6
^{13}C : 8 signals	Min. $n_{\text{C}} = 8$
^1H : 4 signals, sum of integrals is 10	Min. $n_{\text{H}} = 10$
IR: 1704 , 1693 cm^{-1} (very strong) 3114 cm^{-1}	No CN, no NO_2 , 2 C=O are possible, =C–H, min $n_{\text{O}} = 1$
Ranges of atom numbers	C (8–9), H (10–11), O (1–5), N (0–6) DBE (4–8)
Molecular Formulae Generated	1. $\text{C}_8\text{H}_{10}\text{N}_4\text{O}_2$ 2. $\text{C}_9\text{H}_{10}\text{N}_2\text{O}_3$

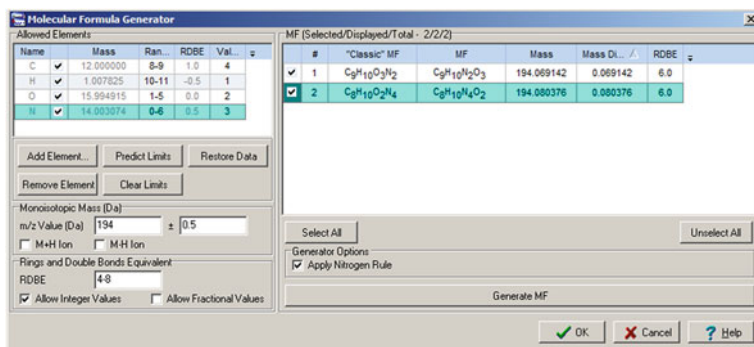


Fig. 3.31 Challenge 9: The **Molecular Formula Generator** dialog window with two molecular formulae

Table 3.16 Challenge 9: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	155.9	155.63	C	–	–	–
C2	107.6	106.65	C	–	–	–
C3	148.3	149.52	C	–	–	–
C4	152.4	151.87	C	–	–	–
C5	143.7	141.96	CH	8.4	s	C3
C6	33.6	33.62	CH_3	4.07	s	C2, C5
C7	28.1	28.32	CH_3	3.42	s	C4
C8	30	29.28	CH_3	3.6	s	C4

(ii), run the structure generation from the first MCD using the command **Structure Elucidation\Run CSB Generator from Current MCD...** and then from the second one (an option **Clear Generated Molecules List before Generation** must be deselected), (iii) divide the problem into two independent tasks and then compare the results of the output file ranking for both tasks. For the sake of clarity we will perform the work with the third approach.

To demonstrate the performance of the program even under those conditions when minimum spectrum-structure information is available we will use only the data from the 1D NMR spectra initially.

MCDs created from the ^{13}C NMR data and the two possible molecular formulae are shown in Fig. 3.32.

Note that the MCD shown on the right-hand side (b) contains a hypothetical carbon atom for which sp^2 hybridization was assigned by the user (the ^{13}C signals from the quaternary carbon atoms can be very weak), but its chemical shift is unknown. Two evident N– CH_3 bonds were drawn by hand.

Structure generation from the MCD for (a) led to the following results: $k = 36,984 \rightarrow 21,054 \rightarrow 1,066$, $t_g = 38$ s. The top ten structures of the ranked file are shown in Fig. 3.33.

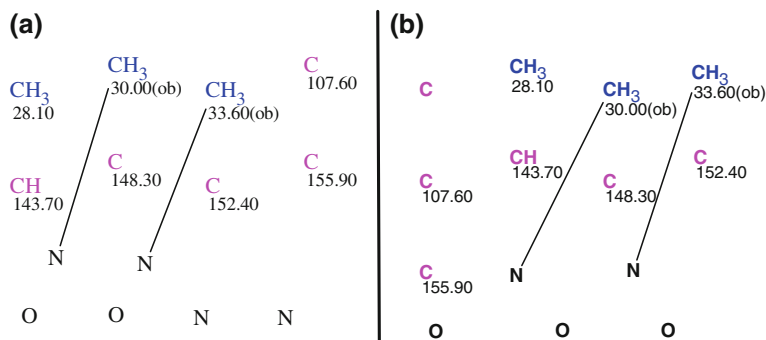


Fig. 3.32 Challenge 9: The MCDs created from the molecular formulae $C_8H_{10}N_4O_2$ (a) and $C_9H_{10}N_2O_3$ (b)

 $d_A(^{13}C): 0.486$ $d_N(^{13}C): 1.281$ $d_I(^{13}C): 0.804$	 $d_A(^{13}C): 2.079$ $d_N(^{13}C): 2.781$ $d_I(^{13}C): 4.474$	 $d_A(^{13}C): 2.560$ $d_N(^{13}C): 2.560$ $d_I(^{13}C): 2.694$	 $d_A(^{13}C): 2.858$ $d_N(^{13}C): 2.681$ $d_I(^{13}C): 3.631$	 $d_A(^{13}C): 2.862$ $d_N(^{13}C): 3.600$ $d_I(^{13}C): 3.083$
 $d_A(^{13}C): 2.879$ $d_N(^{13}C): 2.875$ $d_I(^{13}C): 2.669$	 $d_A(^{13}C): 2.909$ $d_N(^{13}C): 2.182$ $d_I(^{13}C): 4.438$	 $d_A(^{13}C): 3.360$ $d_N(^{13}C): 3.508$ $d_I(^{13}C): 3.405$	 $d_A(^{13}C): 3.360$ $d_N(^{13}C): 1.922$ $d_I(^{13}C): 2.201$	 $d_A(^{13}C): 3.429$ $d_N(^{13}C): 2.353$ $d_I(^{13}C): 2.827$

Fig. 3.33 Challenge 9: The top ten structures of the ranked output file obtained with the molecular formula $C_8H_{10}N_4O_2$

We see that ^{13}C chemical shift prediction correctly selected the right structure from about 1.5 thousand candidates. It is noteworthy that structure #9 was considered in the textbook [1] as an alternative one to structure #1 and it was eliminated on the basis of comprehensive analysis of the mass spectrum. Hypothetical structures #3 and #8 were generated and involved in assessment only on the basis of the structure generation. Figure 3.34 showing the deviation tolerances of the calculated chemical shifts are labeled using colors (see Sect. 2.2.2 and Fig. 3.8) and visually depicts the consistent priority of the correct structure ranked in the first position.

Structure generation was then performed from the MCD (b). The following results were obtained: $k = 207,348 \rightarrow 21,544 \rightarrow 394$, $t_g = 43$ s. The top structures of the ranked output file are shown in Fig. 3.35. Comparison of Fig. 3.9c–e allows us to readily conclude that the best structure corresponding to the molecular formula $C_9H_{10}N_2O_3$ is wrong (deviations are much larger for the best structure) and

1	2	3	4	5
$d_A(^{13}\text{C}): 0.486$ $d_N(^{13}\text{C}): 1.281$ $d_I(^{13}\text{C}): 0.804$	$d_A(^{13}\text{C}): 2.079$ $d_N(^{13}\text{C}): 2.781$ $d_I(^{13}\text{C}): 4.474$	$d_A(^{13}\text{C}): 2.560$ $d_N(^{13}\text{C}): 2.560$ $d_I(^{13}\text{C}): 2.694$	$d_A(^{13}\text{C}): 2.858$ $d_N(^{13}\text{C}): 2.681$ $d_I(^{13}\text{C}): 3.631$	$d_A(^{13}\text{C}): 2.862$ $d_N(^{13}\text{C}): 3.600$ $d_I(^{13}\text{C}): 3.083$

Fig. 3.34 Challenge 9: Visual demonstration of the priority of structure #1. The *green circles* denote atoms for which deviations between the experimental and calculated ^{13}C chemical shifts are less than 3 ppm

1	2	3	4
$d_A(^{13}\text{C}): 3.422$ $d_N(^{13}\text{C}): 3.365$ $d_I(^{13}\text{C}): 3.361$	$d_A(^{13}\text{C}): 3.431$ $d_N(^{13}\text{C}): 4.320$ $d_I(^{13}\text{C}): 4.244$	$d_A(^{13}\text{C}): 4.678$ $d_N(^{13}\text{C}): 6.344$ $d_I(^{13}\text{C}): 5.296$	$d_A(^{13}\text{C}): 5.291$ $d_N(^{13}\text{C}): 8.549$ $d_I(^{13}\text{C}): 6.764$

Fig. 3.35 Challenge 9: Top structures of the ranked output file which was generated from the molecular formula $\text{C}_9\text{H}_{10}\text{N}_2\text{O}_3$

consequently the molecular formula is wrong. The approach described here is usually successfully utilized to allow selection of the *true molecular formula* when there are two or more possible molecular formulae and it is necessary to make a choice between them.

To demonstrate the role of constraints as an aid to eliminate isomers which do not match the experimental data the structure generation from MCD (**a**) was repeated but the HMBC correlations were used (Table 3.16) and the signal multiplicities in the ^1H NMR spectrum (column M(J)) were set by hand on the MCD (using the tool **Edit Atom Properties**). Results: $k = 239 \rightarrow 49 \rightarrow 28$, $t_g = 0.067$ s and structure **3.18** was again selected as the best one. We see that utilization of all information presented in Table 3.16 reduces the number of initially generated structures (207,348) by a factor of ~ 900 and the generation time by 300 times. When HMBC data were used for creation of the MCD followed by then structure generation, the correct structure was quickly obtained. No alternative structure considered in the textbook was generated.

For completeness, structure generation from the correct molecular formula was repeated with the following combination of constraints:

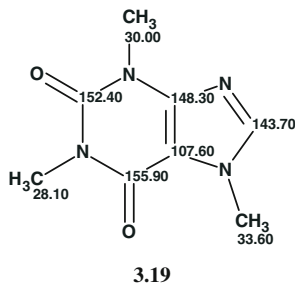
^1H signal multiplicities were set on the MCD (**a**) and HMBC data were used but the chemical bonds between heteroatoms were allowed (check boxes were selected as shown below)

- Allow Bonds between Heteroatoms
- Allow Bonds between Heteroatoms of the Same Atom Type

Results of structure generation: $k = 2,497 \rightarrow 477 \rightarrow 357$, $t_g = 0.627$ s

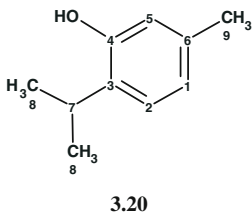
Comparison of the last solution with the previous one allows assessment of the influence of chemical bonds between heteroatoms on the size of the output structure file and the processor time consumed for structure generation—both parameters increased approximately 10 times.

The target structure **3.19** along with ^{13}C chemical shift assignment is shown below



3.10 Challenge 10

The structure of the “unknown” (Table 3.17):



Spectroscopic NMR data for Challenge 10 are collected in Table 3.18.

The MCD where the tabulated data are displayed graphically is shown in Fig. 3.36.

To demonstrate the effects of successively imposing different constraints the first run of the structure generator was performed using only the ^{13}C NMR spectrum and HSQC data. ^{13}C signal multiplicities were determined from DEPT 135 and DEPT 90 experiments. Result: $k = 43,776 \rightarrow 29,484 \rightarrow 391$, $t_g = 21.6$ s. The top 12 structures of the ranked structure file are presented in Fig. 3.37. We see that 9 of the 12 isomers are produced by different substituent arrangements on the benzene ring, the best structure coinciding with structure **3.18**. The correct structure was therefore selected from a set of similar structures by utilizing only the ^{13}C NMR spectrum and its chemical shift prediction.

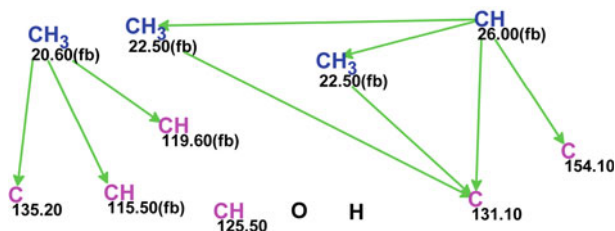
Table 3.17 Challenge 10: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+ = 150$ (I = 36.5 %), even number	$n_N = 0$ or 2 or 4...
MS: $[M+2]^+$ no	No S, no Cl, no Br
MS: all peaks are odd	Presence of N is questionable
^{13}C : no splitting	No F
^{13}C : 9 signals	Min. $n_C = 9$
^1H : $\Sigma\text{H} = 14$	Min. $n_H = 14$, 2 equivalent CH_3 groups
IR: 3240 cm^{-1} $1622, 1586, 1517\text{ cm}^{-1}$	OH, min. $n_O = 1$ Presence of aromatics, no C=O
Ranges of atom numbers	C (9–10), H (14) ^a , O (1–5), N (0–4), DBE (4–6)
Molecular Formulae Generated	$\text{C}_{10}\text{H}_{14}\text{O}$

^a ^1H NMR spectrum contains distinct multiplets and there are no hints of any hidden signals from exchangeable protons

Table 3.18 Challenge 10: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	119.6	121.83	CH	6.55	d(7.7)	–
C2	125.5	126.28	CH	6.97	d(7.7)	–
C3	131.1	131.63	C	–	–	–
C4	154.1	152.21	C	–	–	–
C5	115.5	116.21	CH	6.63	s	–
C6	135.2	136.59	C	–	–	–
C7	26	26.76	CH	3.19	sept (6.9)	C4, C8, C3
C8 (2)	22.5	22.59	CH_3	1.15	d(6.9)	C3
C9	20.6	20.91	CH_3	2.19	s	C6, C1, C5

**Fig. 3.36** Challenge 10: Molecular connectivity diagram

The next run of the structure generator was performed using both the HSQC and HMBC data, but the multiplicities determined in the ^1H NMR spectrum were ignored. Results: $k = 54 \rightarrow 23 \rightarrow 12$, $t_g = 0.023$ s, and the t_g value was reduced approximately 1,000 times while the size of the output file was reduced by a factor













1 	2 	3 	4 	5 	6 
$d_A(^{13}\text{C}): 0.894$ $d_N(^{13}\text{C}): 0.810$ $d_I(^{13}\text{C}): 0.921$	$d_A(^{13}\text{C}): 1.930$ $d_N(^{13}\text{C}): 1.091$ $d_I(^{13}\text{C}): 1.703$	$d_A(^{13}\text{C}): 2.141$ $d_N(^{13}\text{C}): 2.321$ $d_I(^{13}\text{C}): 1.956$	$d_A(^{13}\text{C}): 2.193$ $d_N(^{13}\text{C}): 2.281$ $d_I(^{13}\text{C}): 2.617$	$d_A(^{13}\text{C}): 3.168$ $d_N(^{13}\text{C}): 3.057$ $d_I(^{13}\text{C}): 3.052$	$d_A(^{13}\text{C}): 3.727$ $d_N(^{13}\text{C}): 3.061$ $d_I(^{13}\text{C}): 1.933$
7 	8 	9 	10 	11 	12 
$d_A(^{13}\text{C}): 3.867$ $d_N(^{13}\text{C}): 4.323$ $d_I(^{13}\text{C}): 3.776$	$d_A(^{13}\text{C}): 3.925$ $d_N(^{13}\text{C}): 4.280$ $d_I(^{13}\text{C}): 4.165$	$d_A(^{13}\text{C}): 4.097$ $d_N(^{13}\text{C}): 5.009$ $d_I(^{13}\text{C}): 4.178$	$d_A(^{13}\text{C}): 4.400$ $d_N(^{13}\text{C}): 4.488$ $d_I(^{13}\text{C}): 3.314$	$d_A(^{13}\text{C}): 4.436$ $d_N(^{13}\text{C}): 6.451$ $d_I(^{13}\text{C}): 4.185$	$d_A(^{13}\text{C}): 4.467$ $d_N(^{13}\text{C}): 4.022$ $d_I(^{13}\text{C}): 3.974$

Fig. 3.37 Challenge 10: The top 12 structures of the ranked output structure file. Only the ^{13}C NMR spectrum and HSQC data were used for the structure elucidation






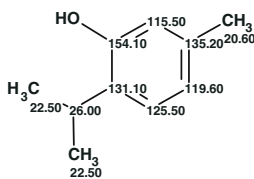
1 	2 	3 	4 	5 
$d_A(^{13}\text{C}): 0.878$ $d_N(^{13}\text{C}): 0.810$ $d_I(^{13}\text{C}): 0.921$	$d_A(^{13}\text{C}): 4.061$ $d_N(^{13}\text{C}): 4.536$ $d_I(^{13}\text{C}): 4.303$	$d_A(^{13}\text{C}): 5.578$ $d_N(^{13}\text{C}): 7.457$ $d_I(^{13}\text{C}): 4.528$	$d_A(^{13}\text{C}): 5.919$ $d_N(^{13}\text{C}): 6.636$ $d_I(^{13}\text{C}): 5.563$	$d_A(^{13}\text{C}): 6.206$ $d_N(^{13}\text{C}): 7.898$ $d_I(^{13}\text{C}): 6.226$

Fig. 3.38 Challenge 10: The top five structures of the output file. 1D, HSQC, and HMBC data were used for the structure elucidation and ^1H signal multiplicities were ignored

of 30. The top five structures of the output file are shown in Fig. 3.38. It should be noted that isomer #3 (Fig. 3.37) is identical to isomer #2 (Fig. 3.38) while the corresponding average and maximum deviations differ significantly.

This is accounted for by the fact that when the HMBC data were used the chemical shift assignment was carried out by the program using constraints imposed by the “axioms” of the HMBC spectrum, which resulted in a more convincing priority of the best structure: $\Delta = d(2) - d(1) \cong 3$ ppm. When information about the ^1H signal multiplicities was involved in the process of structure elucidation the following result was obtained: $k = 9 \rightarrow 7 \rightarrow 5$, $t_g = 003$ s. The effectiveness of employing as many constraints as possible is obvious but at the same time we should always remember that if at least one constraint proves to be erroneous (erroneous axiom) then the solution to the problem will be invalid.

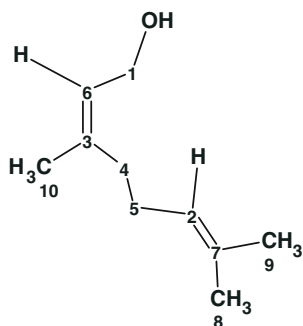
Structure 3.21 shows the carbon chemical shift assignments



3.21

3.11 Challenge 11

The structure of the “unknown” (Table 3.19):

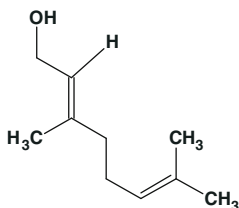


3.22

NMR spectroscopic data used for the computer-assisted structure elucidation are collected in Table 3.20.

The MCD is presented in Fig. 3.39.

All NMR data presented in Table 3.20 (including the signal multiplicities in the ^1H NMR spectrum) were used for structure generation. Results: $k = 6 \rightarrow 2 \rightarrow 1$, $t_g = 0,001$ s. The following average deviations were calculated for the single chemical structure **3.23** which proved to be topologically identical to the target structure **3.22**: $d_A = 1.981$, $d_N = 1.839$, $d_I = 2.175$ ppm.



3.23

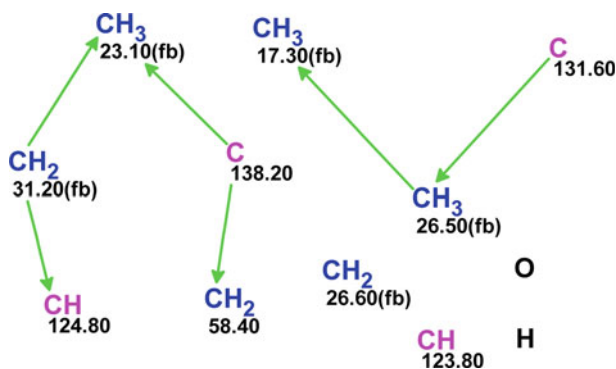
The values of the deviations allow us to accept the structure suggested by the program but the stereoisomeric configuration as *Z* or *E* remains in question. As we see the program automatically depicted structure **3.23** in the *E* form, while in the textbook [1], a NOESY spectrum was used as evidence that the spectral data collected in Table 3.20 corresponds to the *Z* stereo configuration. It was shown [1] that application of ^{13}C NMR chemical shift prediction can frequently help to determine a true stereoisomer without using NOE data. Therefore, the hydrogen atom attached to C6 was represented in an explicit form (right mouse click on atom C6 of the generated structure and select the command **Add Explicit Hydrogens**

Table 3.19 Challenge 11: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 154, very weak, even number	$n_N = 0$ or 2 or 4,
MS: $[M+2] = \text{no}$	No S, no Cl, no Br
MS: all peaks except 1 are odd	Presence of N is questionable
^{13}C : no splitting	No F
^{13}C : 10 signals	Min. $n_C = 10$
^1H : $\Sigma\text{H} = 18$	Min. $n_H = 18$
IR: 3331 cm^{-1} 1668 cm^{-1} , medium	OH, min $n_O = 1$ C=C, no C=O
Ranges of atom numbers	C (10), H (18), O (1–2), N (0–4), RDBE (1–3)
Molecular Formula Generated	C₁₀H₁₈O

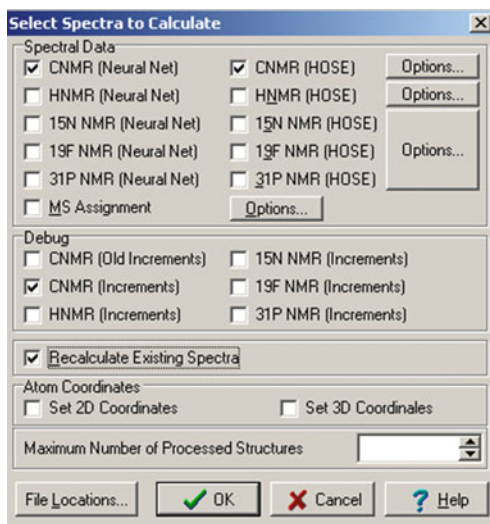
Table 3.20 Challenge 11: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	58.4	58.71	CH ₂	3.9	d(6.0)	C3
C2	124.8	124.04	CH	5.23	t(6.9)	C4
C3	138.2	137.98	C	–	–	–
C4	31.2	32.94	CH ₂	1.84	u	–
C5	26.6	26.7	CH ₂	1.84	u	–
C6	123.8	125.76	CH	4.95	u	–
C7	131.6	131.81	C	–	–	–
C8	26.5	25.52	CH ₃	1.52	u	C7
C9	17.3	17.48	CH ₃	1.44	u	C8
C10	23.1	23.18	CH ₃	1.58	u	C3, C4

**Fig. 3.39** Challenge 11: Molecular connectivity diagram

in the context menu) and then the structure was copied in the **Mol** window (right mouse click on the **Mol** window and select the command **Copy to\Generated Molecules**). The copied structure was transformed into the Z configuration by hand

Fig. 3.40 Dialog window
Select Spectra to Calculate



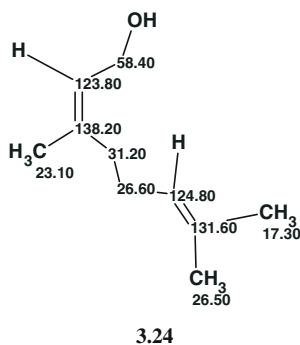
and ^{13}C chemical shift prediction was performed for both structures. To provide for spectrum recalculation, the checkbox **Recalculate Existing Spectra** has to be selected in the dialog window as shown below (Fig. 3.40):

Both stereoisomers with their calculated average deviations are shown in Fig. 3.41. The figure shows clearly that the molecule under investigation exists in the Z configuration in agreement with the conclusions made on the basis of NOE data analysis. Utilizing the StrucEluc expert system allowed us not only to unambiguously determine the skeletal structure of an unknown but also establish its stereoisomeric form.

1	2
$d_A(^{13}\text{C}): 0.655$ $d_N(^{13}\text{C}): 0.671$ $d_I(^{13}\text{C}): 1.049$	$d_A(^{13}\text{C}): 1.981$ $d_N(^{13}\text{C}): 1.839$ $d_I(^{13}\text{C}): 2.175$

Fig. 3.41 Challenge 11: The two possible stereoisomers, Z and E, ranked by average ^{13}C chemical shift deviations

The elucidated structure **3.24** with its ^{13}C chemical shift assignment is shown below



It is noteworthy that a qualitatively correct solution to this simple problem can also be found by structure generation simply from the ^{13}C and HSQC data only ($k = 20,520 \rightarrow 1,416 \rightarrow 27$, $t_g = 1.4$ s), but NMR chemical shift assignment of the correct structure differs from that shown in structure **1a** due to the absence of constraints imposed by the HMBC data on the atoms arrangement in a structure.

3.12 Challenge 12

The structure of the “unknown” (Table 3.21):

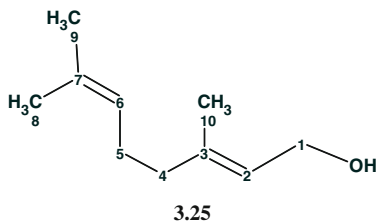


Table 3.21 Challenge 12: Molecular Formula Inference

Factual data	Conclusions
MS: $[\text{M}]^+$ = 154, very weak, even number	$n_{\text{N}} = 0$ or 2 or 4
MS: $[\text{M}+2] = \text{no}$	No S, no Cl, no Br
MS: all peaks except 1 are odd	Presence of N is questionable
^{13}C : no splitting	No F
^{13}C : 10 signals	Min. $n_{\text{C}} = 10$
^1H : $\Sigma\text{H} = 18$	Min. $n_{\text{H}} = 18$
IR: $3,353\text{ cm}^{-1}$ $1,669\text{ cm}^{-1}$, medium	OH, min $n_{\text{O}} = 1$ C=C, no C=O
Ranges of atom numbers	C (10), H (18), O (1–2), N (0–4), RDBE (1–3)
Molecular Formula Generated	$\text{C}_{10}\text{H}_{18}\text{O}$

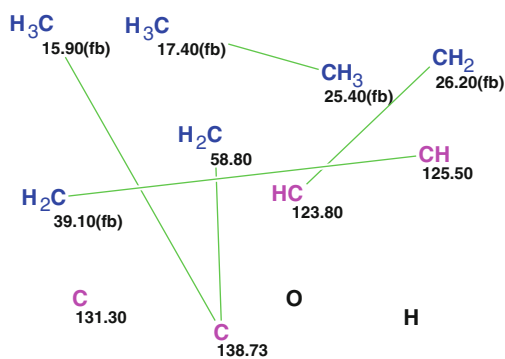
Table 3.22 Challenge 12: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	HMBC
C1	58.8	59.31	CH_2	4.02	D(6.7)	C3
C2	125.5	121.5	CH	5.29	T(6.7)	C4
C3	138.73	138.88	C	–	–	–
C4	39.1	39.46	CH_2	1.91	u	–
C5	26.2	26.27	CH_2	2.02	u	C6
C6	123.8	124.05	CH	5	u	–
C7	131.3	131.31	C	–	–	–
C8	25.4	25.52	CH_3	1.58	S	C9
C9	17.4	17.48	CH_3	1.5	S	–
C10	15.9	16.18	CH_3	1.56	S	C3

NMR spectroscopic data used for CASE are collected in Table 3.22.

The spectroscopic data presented in Table 3.22 were used to create the MCD (Fig. 3.42).

Structure generation was initiated from the MCD and the signal multiplicities in the ^1H NMR spectrum were taken into account (column M(J), Table 3.22). Results: $k = 10 \rightarrow 8 \rightarrow 3$, $t_g = 0.020$ s. The ranked output file is shown in Fig. 3.43. As in the previous Challenge 11 the correct chemical structure was unambiguously determined while its stereo configuration was left to be elucidated. Two possible stereoisomers were copied in the window Proposed Molecule (PM) and the ^{13}C chemical shifts were recalculated. The possible stereoisomers ranked by the values

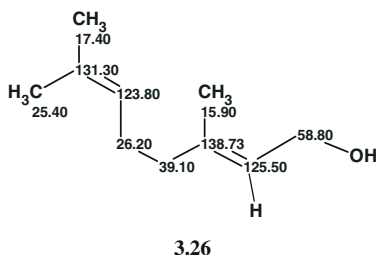
Fig. 3.42 Challenge 12: Molecular connectivity diagram**Fig. 3.43** Challenge 12: The ranked output file

$d_A(^{13}\text{C}): 0.612$ (v. $d_N(^{13}\text{C}): 0.655$ $d_I(^{13}\text{C}): 0.660$	$d_A(^{13}\text{C}): 3.009$ (v. $d_N(^{13}\text{C}): 3.063$ $d_I(^{13}\text{C}): 2.954$	$d_A(^{13}\text{C}): 8.981$ (v. $d_N(^{13}\text{C}): 8.819$ $d_I(^{13}\text{C}): 9.215$

Fig. 3.44 Challenge 12: E (#1) and Z (#2) stereoisomers of structure **3.25**. The values of the average deviations confirm the E configuration (**3.26**) to be correct

1	2
$d_A(^{13}\text{C}): 0.612$ $d_I(^{13}\text{C}): 0.660$ $d_N(^{13}\text{C}): 0.655$ $\text{max_}d_A(^{13}\text{C}): 4.300$ $\text{max_}d_I(^{13}\text{C}): 1.446$ $\text{max_}d_N(^{13}\text{C}): 1.951$	$d_A(^{13}\text{C}): 1.679$ $d_I(^{13}\text{C}): 1.752$ $d_N(^{13}\text{C}): 1.992$ $\text{max_}d_A(^{13}\text{C}): 7.500$ $\text{max_}d_I(^{13}\text{C}): 6.843$ $\text{max_}d_N(^{13}\text{C}): 8.136$

of the average deviations are presented in Fig. 3.44 which clearly identifies the E stereoisomer (see structure **3.26**).



Application of ^{13}C NMR spectrum prediction allowed us once again to unambiguously select the correct stereoisomer configuration without using NOE data or detailed analysis of spectral parameters such as the coupling.

The last two examples show that application of the StrucEluc expert system allows one to quickly and reliably determine the structure and stereo configuration of an unknown compound without any comprehensive spectral analysis.

3.13 Challenge 13

The structure of the “unknown” (Table 3.23):

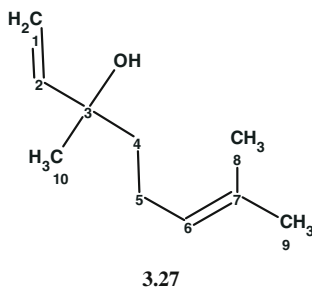


Table 3.23 Challenge 13: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 154, very weak, even number	$n_N = 0$ or 2 or 4,
MS: $[M+2] = \text{no}$	No S, no Cl, no Br
MS: all peaks except 1 are odd	Presence of N is questionable
^{13}C : no splitting	No F
^{13}C : 10 signals	Min. $n_C = 10$
^1H : $\Sigma\text{H} = 18$	Min. $n_H = 18$
IR: 3404 cm^{-1} 1642 cm^{-1} medium,	OH, min $n_O = 1$ C=C, C=O (?)
Ranges of atom numbers	C (10), H (18), O (1–2), N (0–4), RDBE (1–3)
Molecular Formula Generated	C₁₀H₁₈O

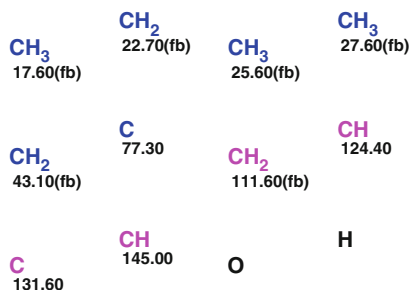
Table 3.24 Challenge 13:
Spectroscopic NMR data
(^{13}C , ^1H , and HSQC)

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)
C1	111.6	112.07	CH_2	5.13	u
C1	111.6	112.07	CH_2	4.96	u
C2	145	145.18	CH	5.82	u
C3	77.3	75.15	C	–	–
C4	43.1	41.65	CH_2	1.48	u
C5	22.7	23.42	CH_2	1.95	u
C6	124.4	123.97	CH	5.04	u
C7	131.6	131	C	–	–
C8	17.6	17.54	CH_3	1.54	u
C9	25.6	25.69	CH_3	1.6	u
C10	27.6	27.91	CH_3	1.19	u

The data used for the computer-assisted structure elucidation of the unknown are presented in Table 3.24.

To elucidate the structure of the molecule only the ^{13}C , ^1H , and HSQC data were input into the program and the MCD was created (Fig. 3.45).

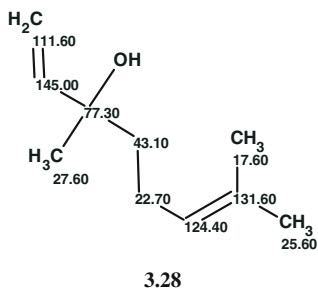
Structure generation was completed with the following results: $k = 2,760 \rightarrow 1,584 \rightarrow 88$, $t_g = 1.7$ s. The first four top structures of the ranked

Fig. 3.45 Challenge 13:
Molecular connectivity
diagram

1 	2 	3 	4
$d_A(^{13}\text{C}): 0.646$ $d_N(^{13}\text{C}): 0.858$ $d_I(^{13}\text{C}): 1.021$	$d_A(^{13}\text{C}): 2.355$ $d_N(^{13}\text{C}): 2.080$ $d_I(^{13}\text{C}): 2.059$	$d_A(^{13}\text{C}): 3.052$ $d_N(^{13}\text{C}): 3.247$ $d_I(^{13}\text{C}): 3.005$	$d_A(^{13}\text{C}): 3.535$ $d_N(^{13}\text{C}): 2.809$ $d_I(^{13}\text{C}): 3.130$

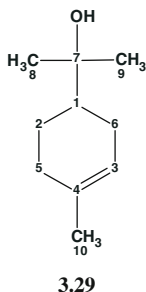
Fig. 3.46 Challenge 13: The first four top structures of the ranked output file

output file are presented in Fig. 3.46 which shows that the correct structure #1 was unambiguously established very quickly without utilizing either HMBC or NOE data (see structure 3.28).



3.14 Challenge 14

The structure of the “unknown” (Table 3.25):



The NMR data input into the program are collected in Table 3.26.

Figure 3.47 shows the MCD created by the program.

Note that the MCD contains two carbon atoms, CH_2 23.29 and CH_2 26.80, which have no HMBC connectivities to other skeletal atoms. This means that there

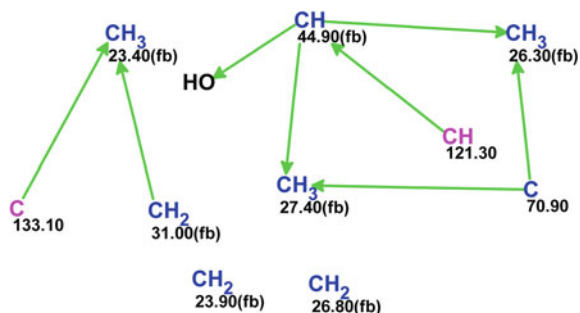
Table 3.25 Challenge 14: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$: Not present, $m/z = 154$ was deduced using information from NMR and IR spectra.	$n_N = 0$ or 2 or 4,
MS: $[M+2]^+$: no	No S, no Cl, no Br
MS: all peaks except 1 are odd	N is questionable
^{13}C : no splitting	No F
^{13}C : 10 signals	Min. $n_C = 10$
1H : $\Sigma H = 18$	Min. $n_H = 18$
IR: 3384 cm^{-1} 1678 cm^{-1} (very weak), $3049, 3010\text{ cm}^{-1}$	OH, min $n_O = 1$ C=C, no C=O, no C \equiv C
Ranges of atom numbers	C (10), H (18), O (1–4), N (0–4), DBE (1–3)
Molecular Formula Generated	C₁₀H₁₈O

Table 3.26 Challenge 14: Spectroscopic NMR data

Label	δX	δC_{calc}	XH _n	δH	M(J)	C HMBC
C1	44.9	46.11	CH	1.33	u	C3
C2	23.9	24.71	CH ₂	1.85	u	–
C2	23.9	24.71	CH ₂	1.1	u	–
C3	121.3	120.48	CH	5.29	u	–
C4	133.1	134.03	C	–	–	–
C5	31	30.87	CH ₂	1.85	u	–
C5	31	30.87	CH ₂	1.85	u	–
C6	26.8	27.93	CH ₂	1.85	u	–
C6	26.8	27.93	CH ₂	1.67	u	–
C7	70.9	72.46	C	–	–	–
C8	26.3	26.31	CH ₃	0.99	u	C1, C7
C9	27.4	27.11	CH ₃	1.01	u	C7, C1
C10	23.4	23.32	CH ₃	1.55	u	C4, C5
O1	100 ^a	–	OH	4.12	u	C1

^a Fictitious ^{17}O chemical shift input to introduce a 1H chemical shift of 4.12 ppm assigned to an OH group

Fig. 3.47 Challenge 14: Molecular connectivity diagram


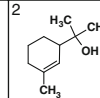
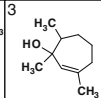

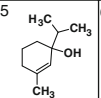
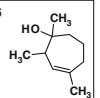
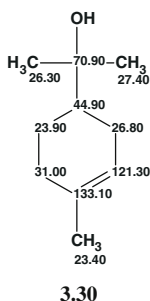
1	2	3	4	5	6
					
$d_A(^{13}\text{C}): 0.697$ $d_N(^{13}\text{C}): 0.865$ $d_I(^{13}\text{C}): 0.781$	$d_A(^{13}\text{C}): 1.281$ $d_N(^{13}\text{C}): 1.299$ $d_I(^{13}\text{C}): 1.550$	$d_A(^{13}\text{C}): 4.533$ $d_N(^{13}\text{C}): 5.000$ $d_I(^{13}\text{C}): 4.077$	$d_A(^{13}\text{C}): 5.163$ $d_N(^{13}\text{C}): 5.132$ $d_I(^{13}\text{C}): 4.835$	$d_A(^{13}\text{C}): 5.336$ $d_N(^{13}\text{C}): 5.484$ $d_I(^{13}\text{C}): 4.955$	$d_A(^{13}\text{C}): 5.429$ $d_N(^{13}\text{C}): 4.955$ $d_I(^{13}\text{C}): 5.175$

Fig. 3.48 Problem 14: Ranked output file

are no constraints imposed by the HMBC correlations on their position in generated molecules and, consequently, the true positions will be determined as a result of chemical shift prediction for the output file. Structure generation was performed and gave the following results: $k = 42 \rightarrow 18 \rightarrow 6$, $t_g = 0.09$ s. All structures of the ranked output file are presented in Fig. 3.48.

The correct structure is ranked first by ^{13}C chemical shift prediction and automatic chemical shift assignment is shown on structure 3.30 below



Comparison of the assignment with that deduced in the original work [1] shows that the chemical shifts 23.90 and 26.80 related to “free” CH_2 groups are interchanged in structure 3.30. The structures with two possible chemical shift assignments were placed in the Proposed Molecule (PM) window. For this goal structure #1 was copied into the PM window twice and the chemical shifts 23.90 and 26.80 were exchanged by hand in the second structure. Then the ^{13}C chemical shifts were recalculated and the two structures with their corresponding average deviations are presented in Fig. 3.49.

The priority of structure #1 is evident from comparing the average and maximum deviations. To additionally confirm the ^{13}C chemical shift assignment for structure #1, the protocol of chemical shift prediction for C 26.80 was displayed (Fig. 3.50).

The protocol shows that 37 hits (structures) were found in the ACD/DB for predicting chemical shifts of the selected CH_2 group (in square). The lowest chemical shift value observed in the reference structures is 25.8, while the calculated value is equal to 26.11 ppm. Therefore, the automatically performed chemical

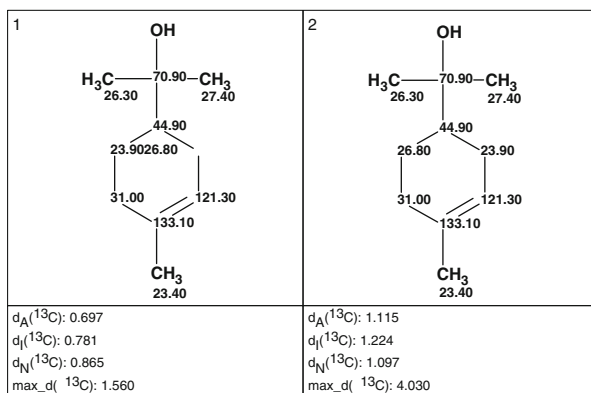


Fig. 3.49 Challenge 14: The correct structure #1 with inverted positions of atoms C 23.90 and C 26.80. Assignment of structure #1 was performed automatically. Exchange of chemical shifts for structure #2 was performed manually

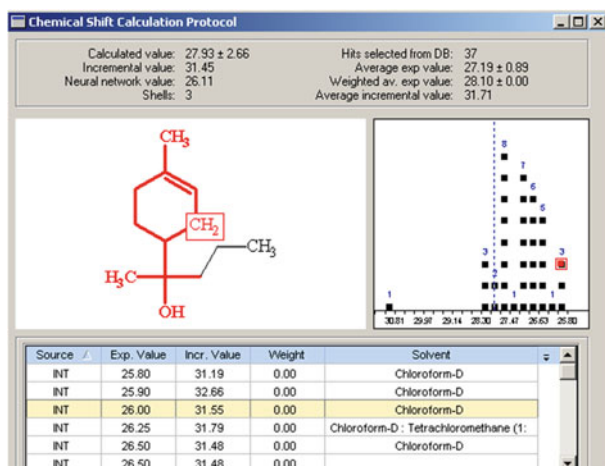
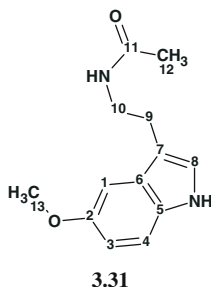


Fig. 3.50 Challenge 5.14: Protocol of chemical shift calculation for atom C 26.80 in structure #1 (Fig. 3.49)

shift assignment looks more probable. The suggested reassignment was additionally confirmed by a structural search against the ACD/NMR DB: it was revealed that structure **1** existed in the database with this assignment.

3.15 Challenge 15

The structure of the “unknown” (Table 3.27):



NMR spectroscopic data used for the CASE study are presented in Table 3.28.

Even though the molecule is not large the presence of two NH groups and DBE = 7 justify utilizing the HMBC data and the reliable correlations observed in the COSY spectrum. One of the two NH groups identified in the ^1H NMR spectrum (δH 10.61 ppm) shows correlations to carbon atoms C5, C6, C7 in the ^1H - ^{13}C HMBC spectrum. As the ^1H - ^{15}N HSQC spectrum was not acquired the chemical shifts of N_1 and N_2 nitrogen atoms are unknown. To involve these atoms in the “net” of HMBC connectivities the following formal approach can be used: different fictitious chemical shifts are assigned to the nitrogen atoms and fictitious “ ^{15}N NMR user” and fictitious ^1H - ^{15}N HSQC spectra are introduced into the program. As shown in Table 3.28, chemical shifts of $\delta\text{N}_1 = 100$ and $\delta\text{N}_2 = 150$ ppm were arbitrarily assigned to the nitrogen atoms. Under this condition the last line of Table 3.28 becomes meaningful in spite of the fictitious chemical shift value of nitrogen atom N_2 . This artificial approach is commonly employed in similar cases

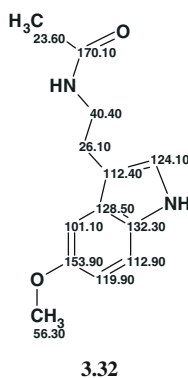
Table 3.27 Challenge 15: Molecular Formula Inference

Factual data	Conclusions
MS: $[\text{M}]^+ = 232$	$n_{\text{N}} = 0$ or 2 or 4...
MS: $[\text{M}+2]^+$: no	No S, no Cl, no Br
MS: 5 peaks are even	N is probable
^{13}C : no splitting	No F
^{13}C : 13 signals	Min. $n_{\text{C}} = 13$
^1H : $\Sigma\text{H} = 16$	Min. $n_{\text{H}} = 16$
IR: 3244 cm^{-1} very strong	OH or NH or H_2O
$1587, 1568\text{ cm}^{-1}$ mean, 1630 very strong	Ar, C=O
Ranges of atom numbers	C (13), H (16–18), O (0–5), N (0–6), DBE (5–10)
Molecular Formulas Generated	$\text{C}_{13}\text{H}_{16}\text{N}_2\text{O}_2$ RDBE = 7

1	2	3	5
$d_A(^{13}\text{C}): 1.363$ $d_N(^{13}\text{C}): 1.556$ $d_I(^{13}\text{C}): 1.607$	$d_A(^{13}\text{C}): 3.545$ $d_N(^{13}\text{C}): 4.053$ $d_I(^{13}\text{C}): 3.719$	$d_A(^{13}\text{C}): 4.087$ $d_N(^{13}\text{C}): 4.452$ $d_I(^{13}\text{C}): 5.876$	$d_A(^{13}\text{C}): 4.184$ $d_N(^{13}\text{C}): 4.480$ $d_I(^{13}\text{C}): 3.551$
7	10	33	37
$d_A(^{13}\text{C}): 4.352$ $d_N(^{13}\text{C}): 4.685$ $d_I(^{13}\text{C}): 5.044$	$d_A(^{13}\text{C}): 5.177$ $d_N(^{13}\text{C}): 4.805$ $d_I(^{13}\text{C}): 5.280$	$d_A(^{13}\text{C}): 7.142$ $d_N(^{13}\text{C}): 6.868$ $d_I(^{13}\text{C}): 7.653$	$d_A(^{13}\text{C}): 7.492$ $d_N(^{13}\text{C}): 7.048$ $d_I(^{13}\text{C}): 7.979$

Fig. 3.52 Challenge 15: A set of structures from the ranked output file

The molecular structure generation from the MCD was completed with the following results: $k = 2,417 \rightarrow 1,246 \rightarrow 379$, $t_g = 3.8$ s. A set of selected structures which are included into the ranked output file are presented in Fig. 3.52. Note that they are displayed not in the “one after another” order (see structure numbers in the ranked file) to demonstrate, on the one hand, the diversity of generated structures and, on the other hand, the degree of their similarity. The correct structure **3.32** (see below) was unambiguously distinguished from 379 candidate structures. Note that molecules containing a C=N group were generated but the ranking procedure placed the closest ranked structure containing C=N in the 5th position.



For the sake of completeness in the investigation, structure generation was repeated with the C=O bond is drawn by hand in the MCD. Results: $k = 744 \rightarrow 468 \rightarrow 116$, $t_g = 1.2$ s. We see that the single and almost evident constraint reduces both the number of generated structures and the generation time approximately by a factor of 3.

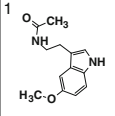
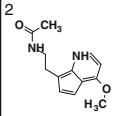
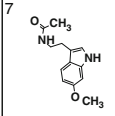
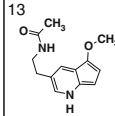
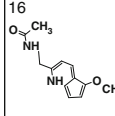
				
$d_A(^{13}\text{C}): 1.363$ $d_N(^{13}\text{C}): 1.556$ $d_I(^{13}\text{C}): 1.607$	$d_A(^{13}\text{C}): 3.545$ $d_N(^{13}\text{C}): 4.053$ $d_I(^{13}\text{C}): 3.719$	$d_A(^{13}\text{C}): 5.177$ $d_N(^{13}\text{C}): 4.805$ $d_I(^{13}\text{C}): 5.280$	$d_A(^{13}\text{C}): 6.359$ $d_N(^{13}\text{C}): 5.695$ $d_I(^{13}\text{C}): 5.916$	$d_A(^{13}\text{C}): 6.525$ $d_N(^{13}\text{C}): 6.602$ $d_I(^{13}\text{C}): 6.357$

Fig. 3.53 Challenge 15: The correct structure along with similar ones extracted from the ranked output file

Figure 3.53 demonstrates the value of ^{13}C chemical shift prediction to discriminate similar structures containing the carbonyl group.

3.16 Challenge 16

The structure of the “unknown” (Table 3.29):

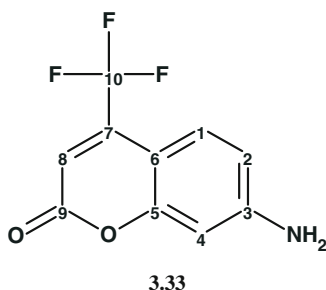


Table 3.29 Challenge 16: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+ = 229$, odd number	$n_N = 1$ or 3 or 5
MS: $[M+2]^+$: no	No S, no Cl, no Br
MS: 5 peaks are even	N is probable
^{13}C : 3 quartets, coupling with F	Hint to CF_3
^{13}C : 10 signals	Min. $n_C = 10$
^1H : $\Sigma\text{H} = 6$, strong peak from H_2O ,	Min. $n_H = 6$
IR: 3455, 3380 cm^{-1} very strong 1616, 1605 cm^{-1} strong, 1712 cm^{-1} very strong	OH, NH, NH_2 , H_2O possible Aromatics, $\text{C}=\text{O}$, min. $n_O = 1$
Ranges of atom numbers	C (10–12), H (6–7), O (1–5), N (1–5), F (3), DBE (4–10)
Molecular Formula Generated	$\text{C}_{10}\text{H}_6\text{NO}_2\text{F}_3$

Table 3.30 Challenge 16:
Spectroscopic NMR data

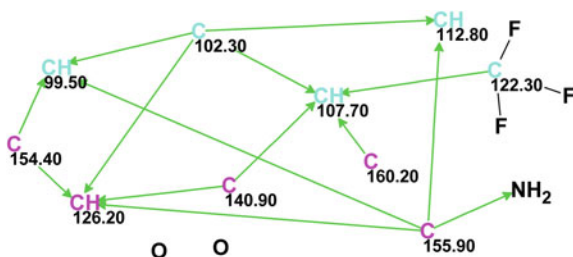
Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	126.2	125.67	CH	7.29	u	C3, C6, C7, C5
C2	112.8	114.72	CH	6.61	u	C6, C3
C3	155.9	152.76	C	–	–	–
C4	99.5	101.38	CH	6.48	u	C5, C3, C6
C5	154.4	155.53	C	–	–	–
C6	102.3	103.82	C	–	–	–
C7	140.9	139.91	C	–	–	–
C8	107.7	113.81	CH	6.34	u	C6, C10, C7, C9
C9	160.2	158.53	C	–	–	–
C10	122.3	121.81	C	–	–	–

NMR spectroscopic data that were input into the program are presented in Table 3.30.

The MCD generated is shown in Fig. 3.54.

As the ^{13}C NMR spectrum contains a quartet at 122.3 ppm for which the measured coupling constant is 275 Hz ($^1J_{\text{CF}} = 272$ Hz is given in the literature) this signal was assigned as a CF_3 group and three chemical bonds from the F atoms to C 122.3 were drawn by hand. As the signal at δH 6.42 (2H) showed no correlations in the HSQC spectrum it was unambiguously related to NH_2 , which is in agreement with ^1H spectrum-structure correlations for Ar-NH_2 . For the sake of caution no other MCD edits were made. Structure generation initiated from the MCD gave the following results: $k = 697 \rightarrow 562 \rightarrow 290$, $t_g = 7$ s. Selected structures from the ranked output file are presented in Fig. 3.56.

Figure 3.55 shows that the true structure was unambiguously determined ($\Delta = d(2) - d(1) \cong 3$ ppm), while two similar structures which are characterized by huge deviations (15–20 ppm) were ranked as #86 and #166. In the textbook [1], four alternative structures (Fig. 3.56) were considered and structure 3.33 was chosen on the basis of detailed analysis of the 1D and 2D NMR spectra.

Fig. 3.54 Problem 16:
Molecular connectivity
diagram

1 	2 	64 	86 	166
$d_A(^{13}\text{C}): 1.938$ $d_N(^{13}\text{C}): 1.935$ $d_I(^{13}\text{C}): 2.889$	$d_A(^{13}\text{C}): 5.255$ $d_N(^{13}\text{C}): 4.706$ $d_I(^{13}\text{C}): 5.657$	$d_A(^{13}\text{C}): 14.832$ $d_N(^{13}\text{C}): 13.991$ $d_I(^{13}\text{C}): 14.485$	$d_A(^{13}\text{C}): 17.058$ $d_N(^{13}\text{C}): 15.102$ $d_I(^{13}\text{C}): 15.378$	$d_A(^{13}\text{C}): 19.490$ $d_N(^{13}\text{C}): 19.061$ $d_I(^{13}\text{C}): 20.882$

Fig. 3.55 Challenge 16: Selected structures in the ranked output file

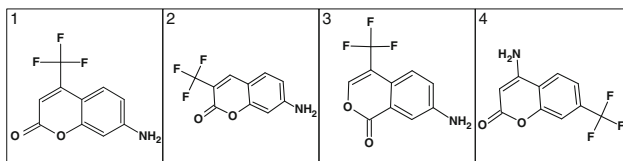
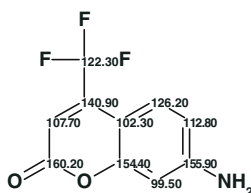


Fig. 3.56 Problem 16: The four alternative structures suggested in the textbook [1]

To select the most probable structure from all alternative structural hypotheses the following simple and fast approach can be utilized. The structures shown in Fig. 3.56 were placed in the window Proposed Molecules (PM) and ^{13}C chemical shift prediction was carried out using the Neural Nets approach. The command **File \Create Report\Stick Spectra...** was initiated and in the options window **Neural Nets** were selected as the method of chemical shift calculation. As a result the program generated the picture presented in Fig. 3.57. Visual comparison of the experimental ^{13}C NMR spectrum (top) with the spectra of all of the alternative structures allows us to easily conclude that the spectrum of structure #1 (correct structure) is closest to the experimental spectrum. The spectra of the wrong alternative structures contain significant outliers. Note that the approach described is very useful and convenient for fast elimination of the wrong structural suggestions. What is especially attractive is that when this approach is used there is *no need for chemical shift assignment* for the set of examined structures.

It is noteworthy that suggested structure #3 (Fig. 3.56) is identical to the generated structure ranked as #86 (Fig. 3.55). ^{13}C chemical shift assignment of the true structure is shown below (3.34):



3.34

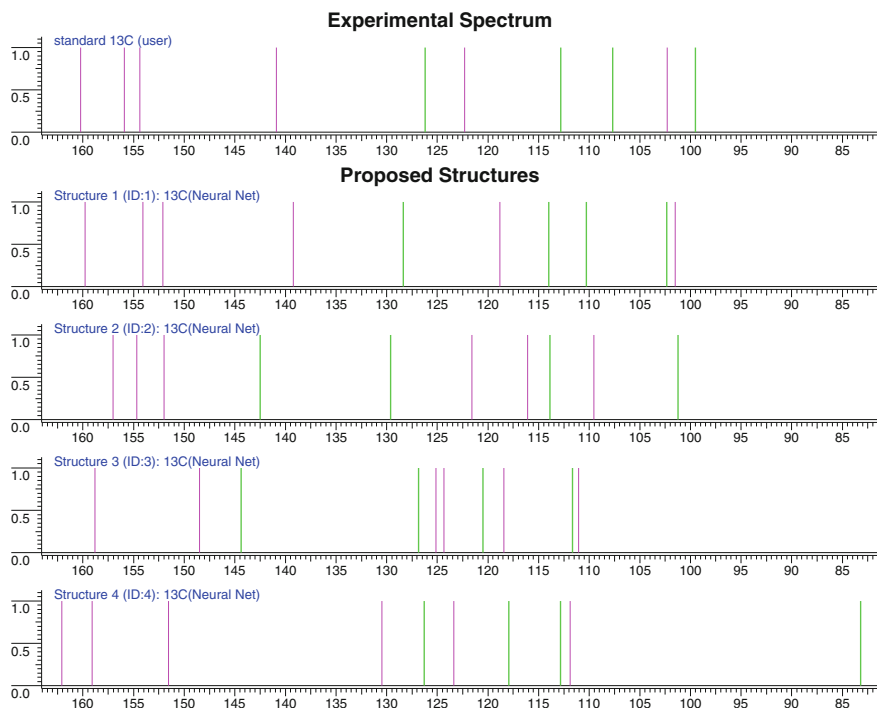
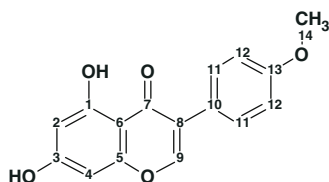


Fig. 3.57 Challenge 16: The experimental ^{13}C NMR spectrum (*top*) compared with spectra predicted for the four alternative structures presented in Fig. 3.56. The *violet lines* denote singlets while the *green lines* show doublets

3.17 Challenge 17

The structure of the “unknown” (Table 3.31):



3.35

^{13}C , ^1H , HSQC, and HMBC NMR spectra were used for the structure elucidation with the aids of StrucEluc expert system (Table 3.32).

The MCD created from the data presented in Table 3.32 is shown in Fig. 3.58.

What immediately attracts attention in the MCD is the presence of seven carbon atoms (colored in light blue) for which an ambiguous state of hybridization “*not sp*”

Table 3.31 Challenge 17: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 284, even (chemical ionization gives m/z 285 = $(M+1)^+$)	$n_N = 0$ or 2 or 4...
MS: $[M+2]^+$ no	No S, no Cl, no Br
^{13}C : no splitting	No F
^{13}C : 14 signals	Min. $n_C = 14$, hint to para-benzene
1H : $\Sigma H \geq 9$, strong signal from H_2O	Min. $n_H = 9$, hint to para-benzene
IR: 3490, 1653 cm^{-1} very strong, 1582, 1515 cm^{-1}	OH, min $n_O = 1$ C=O, Aromatics
Ranges of atom numbers	C (16–18), H (9–15), O (1–10), N (0–6), DBE (6–15)
Molecular Formula Generated	C₁₆H₁₂O₅ DBE 11

Table 3.32 Challenge 17: Spectroscopic NMR data

Label	δX	δC_{calc}	XHn	δH	M(J)	C HMBC
C1	162.3	163.03	C	–	–	–
C2	99.5	99.84	CH	6.17	d(2)	–
C3	164.7	164.64	C	–	–	–
C4	94.2	94.05	CH	6.32	d(2)	C6, C5, C3
C5	158	158.29	C	–	–	–
C6	104.9	105.32	C	–	–	–
C7	180.5	180.05	C	–	–	–
C8	122.4	122.47	C	–	–	–
C9	154.5	154.4	CH	8.2	s	C8, C7, C5
C10	123.3	123.43	C	–	–	–
C11(2)	130.6	131.2	CH	7.47	u	C8, C13
C12(2)	114.2	117.3	CH	6.93	u	C10, C13
C13	159.6	160.12	C	–	–	–
C14	55.6	55.36	CH ₃	3.73	s	C13
O1	100 ^a	–	OH	3.81	u	–
O2	150 ^a	–	OH	12.8	u	C2, C1, C6

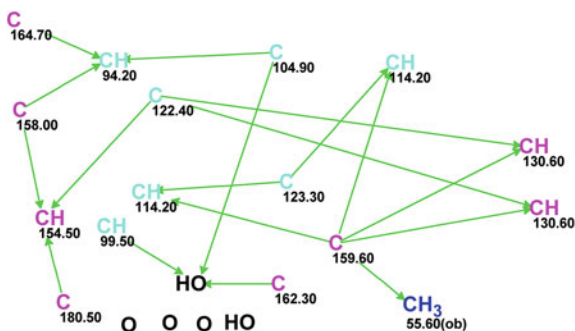
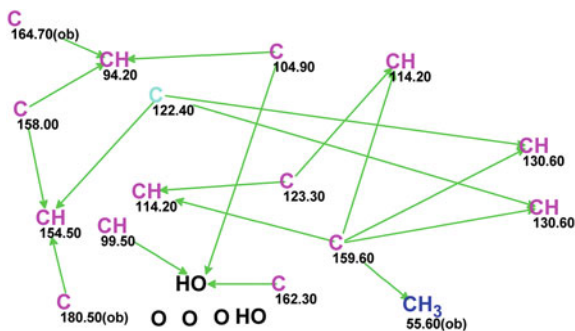
^a Fictitious ^{17}O chemical shifts**Fig. 3.58** Challenge 17: The initial MCD

Fig. 3.59 Challenge 17:
Edited MCD



(sp^2 or sp^3) was assigned by the program. Taking into account that the degree of unsaturation is equal to 11 and that an aromatic ring is definitely present, we can suggest that all these atoms may be supplied with the property sp^2 . In the presence of five oxygen atoms at least two carbons C 180.5 and 164.7 can be allocated the label “*ob*”. According to the signal multiplicities in the ^1H spectrum (Table 3.32) the numbers of hydrogen atoms attached to the carbon atoms belonging to the first sphere were set for C 99.5 $n_e(\text{H}) = 1$, 164.7 $n_e(\text{H}) = 1$, 154.5 $n_e(\text{H}) = 0$, and 55.6 $n_e(\text{H}) = 0$. The edited MCD is shown in Fig. 3.59.

The structure generation from the edited MCD was completed with the following results:

1	3	4	6
$d_A(^{13}\text{C}): 0.681$ $d_N(^{13}\text{C}): 0.533$ $d_I(^{13}\text{C}): 0.665$	$d_A(^{13}\text{C}): 3.849$ $d_N(^{13}\text{C}): 4.233$ $d_I(^{13}\text{C}): 4.357$	$d_A(^{13}\text{C}): 4.264$ $d_N(^{13}\text{C}): 3.821$ $d_I(^{13}\text{C}): 4.481$	$d_A(^{13}\text{C}): 5.064$ $d_N(^{13}\text{C}): 4.812$ $d_I(^{13}\text{C}): 4.777$
7	9	19	25
$d_A(^{13}\text{C}): 5.388$ $d_N(^{13}\text{C}): 5.486$ $d_I(^{13}\text{C}): 6.062$	$d_A(^{13}\text{C}): 6.177$ $d_N(^{13}\text{C}): 6.375$ $d_I(^{13}\text{C}): 7.357$	$d_A(^{13}\text{C}): 7.909$ $d_N(^{13}\text{C}): 8.160$ $d_I(^{13}\text{C}): 7.679$	$d_A(^{13}\text{C}): 8.373$ $d_N(^{13}\text{C}): 7.926$ $d_I(^{13}\text{C}): 8.755$

Fig. 3.60 Challenge 17: The selected top structures of the ranked output file. Similar structures are shown along with their average deviations

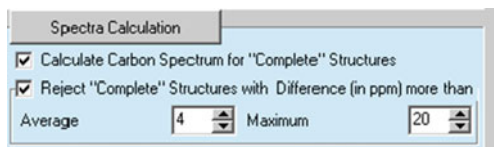


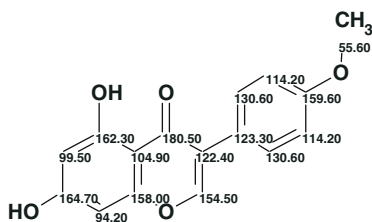
Fig. 3.61 Challenge 17: Option settings for ^{13}C chemical shift calculation during structure generation

$k = 334 \rightarrow 174 \rightarrow 96$, $t_g = 0.33$ s. The selected top structures of the ranked output file are shown in Fig. 3.60. We see that the correct structure is reliably distinguished, while other similar structures are rejected by the corresponding average deviations.

This problem can be used to demonstrate the influence of the number of atoms marked with a light blue color on the results of structure generation. For this goal structure generation was initiated from the initial MCD (Fig. 3.58). The program was stopped (**Stop** button was pressed) after 1 h of the program running, when $\sim 9,000$ structures were generated. This is a hint to add some constraints to accelerate the structure generation process. For the second run, the label “ob” was set for carbon atoms C 162.3, C 164.7, and C 180.5, as well as ^1H signal multiplicities (Table 3.32, column M(J)) were used. To reduce the number of structures that have to be saved on disk the option providing ^{13}C chemical shift calculation during structure generation was utilized (see Fig. 3.61).

The calculation is performed using the incremental approach (the fastest method of chemical shift prediction, $\sim 30,000$ shifts per second), and structures for which $d_I > 4$ ppm and $d_{I(\text{max})} > 20$ ppm are rejected and not saved on disk. Results: $k = 4,534 \rightarrow 1$, $t_g = 7$ min 38 s. The single structure turned out to be the correct one. Thus, we automatically obtained the “safest” solution (i.e., with minimum assumptions), but at the cost of an increase in generation time by 1,500 times in comparison with the solution found from the MCD shown in Fig. 3.59.

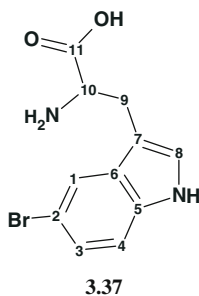
The ^{13}C chemical shift assignment is shown below (3.36):



3.36

3.18 Challenge 18

The structure of the “unknown” (Table 3.33):



Computer-assisted structure elucidation was carried out with the data presented in Table 3.34.

The spectroscopic data contained in Table 3.34 were used to create the MCD (Fig. 3.62).

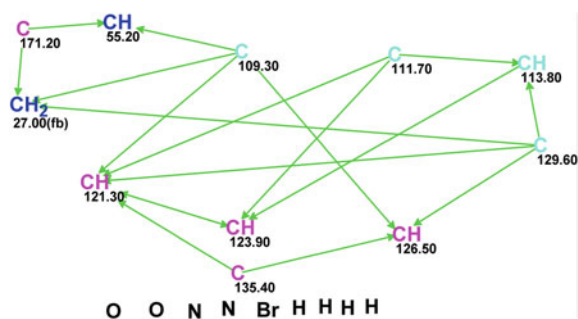
The MCD contains four light blue carbon atoms which are allowed to be sp^3 - or sp^2 -hybridized. When the structure elucidation was carried out manually [1] a suggestion about the presence of NH, NH_2 , and COOH groups was made from the spectral data, particularly from the IR spectrum. To solve the problem in a “safe mode” no edits of the initial MCD were made. The program was allowed to generate all atomic combinations which are possible according to the experimental 1D and 2D NMR spectra and the system knowledge. Structure generation was

Table 3.33 Challenge 18: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 282, even number	$n_N = 0$ or 2 or 4 or 6...
MS: $[M+2]^+$ present, $I[M]^+ : I[M+2] = 1:1$	$n_{Br} = 1$, no hints to S
MS: 10 even peaks	N is very probable
^{13}C : no splitting	No F
^{13}C : 11 signals	Min. $n_C = 11$
1H : $\Sigma H = 8$, strong peak of H_2O at 3.8 ppm	Min. $n_H = 8$
IR: 3439 cm^{-1} $3400\text{--}2600\text{ cm}^{-1}$ 1629 cm^{-1} very strong, 1494 cm^{-1}	OH or NH or NH_2 or H_2O Probable OH of COOH C=O, Aromatics, min $n_O = 1$
Ranges of atom numbers	C (11), H (8–12), O (1–5), N (0–4), Br (1), DBE (4–10)
Molecular Formula Generated	$C_{11}H_{11}N_2O_2Br$, DBE 7

Table 3.34 Challenge 18:
Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	C HMBC
C1	121.3	121.4	CH	7.79	u	C7, C5, C3, C6, C2
C2	111.7	111.7	C	–	–	–
C3	123.9	124	CH	7.14	u	C4, C1, C2
C4	113.8	113.9	CH	7.3	u	C6, C2
C5	135.4	134.6	C	–	–	–
C6	129.6	127.9	C	–	–	–
C7	109.3	111.1	C	–	–	–
C8	126.5	124.2	CH	7.23	u	C7, C6, C5
C9	27	26.99	CH ₂	2.96	u	–
C9	27	26.99	CH ₂	3.24	u	C7, C6, C11
C10	55.2	54.21	CH	3.5	u	C7, C11
C11	171.2	173.8	C	–	–	–

Fig. 3.62 Challenge 18:
Molecular connectivity
diagram

initiated from the initial MCD which was completed with the results: $k = 50 \rightarrow 29 \rightarrow 27$, $t_g = 7$ s.

To indicate the visual diversity of similar structures Fig. 3.63 presents 15 top structures of the ranked output file.

Figure 3.63 shows that structure #1 coincides with structure 3.37 and therefore a correct solution to the problem was found automatically and instantaneously without utilizing any user assumptions. Comparison of the structures reveals many functionalities which are produced by different atomic combinations. Changes observed in the end of the chain in combination with transfers from indole to benzofuran moieties markedly influence the predicted ¹³C NMR chemical shifts and gives the possibility to follow how structural changes are reflected in the values of

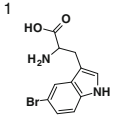
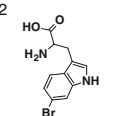
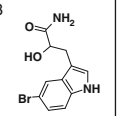
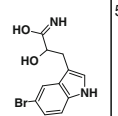
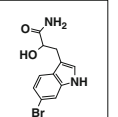
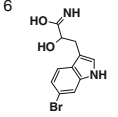
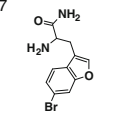
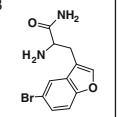
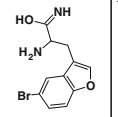
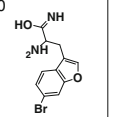
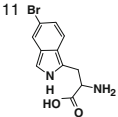
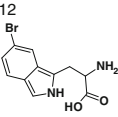
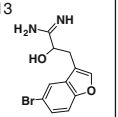
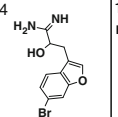
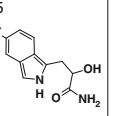
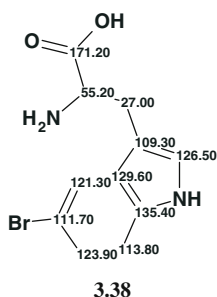
1 	2 	3 	4 	5 
$d_A(^{13}\text{C})$: 0.950 $d_N(^{13}\text{C})$: 1.041 $d_I(^{13}\text{C})$: 1.126	$d_A(^{13}\text{C})$: 1.598 $d_N(^{13}\text{C})$: 1.985 $d_I(^{13}\text{C})$: 1.434	$d_A(^{13}\text{C})$: 2.885 $d_N(^{13}\text{C})$: 3.184 $d_I(^{13}\text{C})$: 3.074	$d_A(^{13}\text{C})$: 3.119 $d_N(^{13}\text{C})$: 2.670 $d_I(^{13}\text{C})$: 3.053	$d_A(^{13}\text{C})$: 3.228 $d_N(^{13}\text{C})$: 3.900 $d_I(^{13}\text{C})$: 3.340
6 	7 	8 	9 	10 
$d_A(^{13}\text{C})$: 3.376 $d_N(^{13}\text{C})$: 3.517 $d_I(^{13}\text{C})$: 3.501	$d_A(^{13}\text{C})$: 4.385 $d_N(^{13}\text{C})$: 5.043 $d_I(^{13}\text{C})$: 5.535	$d_A(^{13}\text{C})$: 4.409 $d_N(^{13}\text{C})$: 5.116 $d_I(^{13}\text{C})$: 5.440	$d_A(^{13}\text{C})$: 4.816 $d_N(^{13}\text{C})$: 5.259 $d_I(^{13}\text{C})$: 4.893	$d_A(^{13}\text{C})$: 4.929 $d_N(^{13}\text{C})$: 5.312 $d_I(^{13}\text{C})$: 5.009
11 	12 	13 	14 	15 
$d_A(^{13}\text{C})$: 5.318 $d_N(^{13}\text{C})$: 6.278 $d_I(^{13}\text{C})$: 8.289	$d_A(^{13}\text{C})$: 5.834 $d_N(^{13}\text{C})$: 5.686 $d_I(^{13}\text{C})$: 8.072	$d_A(^{13}\text{C})$: 6.627 $d_N(^{13}\text{C})$: 6.992 $d_I(^{13}\text{C})$: 6.953	$d_A(^{13}\text{C})$: 6.790 $d_N(^{13}\text{C})$: 6.926 $d_I(^{13}\text{C})$: 7.060	$d_A(^{13}\text{C})$: 7.053 $d_N(^{13}\text{C})$: 8.047 $d_I(^{13}\text{C})$: 10.303

Fig. 3.63 Challenge 18: Top structures of the ranked output file

the average deviations and structure ranking. Structure **3.38** shows the automatically executed ^{13}C chemical shift assignment:



3.19 Challenge 19

The structure of the “unknown”:

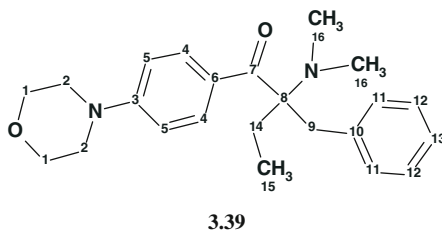


Table 3.35 Challenge 19: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 366	$n_N = 0$ or 2 or 4...
MS: $[M+2]^+$ no	No S, no Cl, no Br
MS: 8 peaks are even	N is very probable
^{13}C : no splitting	No F
^{13}C : 202.2 (C _q), 65.9 (CH ₂), 38.5 2 CH ₃ groups, 46.7(CH ₂), 8 signals in aromatic area	C=O, CH ₂ -O, CH ₃ -N, CH ₂ -N, aromatics, min. $n_O = 2$, min. $n_N = 2$
^{13}C : 16 signals, $\Sigma H = 19$	Min. $n_C = 16$
1H : $\Sigma H = 26$	Min. $n_H = 26$
IR: not available	
Ranges of atom numbers	C (16–26); H (26–36), O (2–5), N (2–4), DBE (8–10)
Molecular Formula Generated	C₂₃H₃₀N₂O₂

In the textbook [1] the molecular formula was determined on the basis of the molecular ion $M^+ = 366$ amu and the elemental chemical analysis data: 75.4 % C, 8.20 % H, 7.65 % N. The ranges of the atom numbers in the **Molecular Formula Generator** were set on the basis of signal intensities observed in the 1H and ^{13}C NMR spectra (Table 3.35). The spectroscopic NMR data necessary to perform the CASE analysis were extracted from the Extra Materials [2] (Table 3.36).

Table 3.36 Challenge 19: Spectroscopic NMR data

Label	δC	δC_{calc}	CH _n	δH	M(J)	COSY	C HMBC
C1(2)	65.9	66.5	CH ₂	3.74	u	3.3	C2
C2(2)	46.7	47.2	CH ₂	3.3	u	3.74	C3
C3	153.2	154.14	C	–	–	–	–
C4(2)	131.4	128.91	CH	6.93	u	8.23	C5, C6
C5(2)	112.5	112.19	CH	8.23	u	6.93	C7
C6	126.6	128.08	C	–	–	–	–
C7	202.2	198.15	C	–	–	–	–
C8	73	73.94	C	–	–	–	–
C9	35.1	39.46	CH ₂	3.11	u		C10, C11, C7, C14, C8
C10	138.8	138.91	C	–	–	–	–
C11(2)	127.7	129.58	CH	<u>7.19</u>	u	–	–
C12(2)	130.8	128.41	CH	<u>7.19</u>	u	–	C10, C9
C13	125.9	126.53	CH	<u>7.19</u>	u	–	–
C14	26.9	28.47	CH ₂	1.78	u	0.59	C9, C15
C14	26.9	28.47	CH ₂	1.96	u	–	C7, C8
C15	9.35	9.49	CH ₃	0.59	u	1.78	C8, C14
C16	38.5	39.88	CH ₃	2.3	u	–	C8

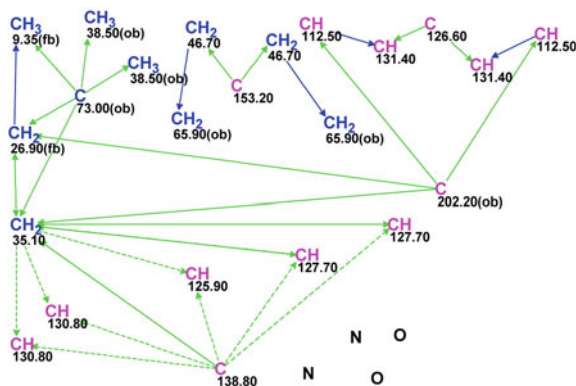
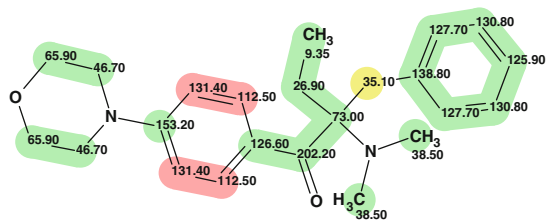


Fig. 3.64 Challenge 19: Molecular connectivity diagram

On the basis of the spectroscopic data presented in Table 3.36 the MCD was created (Fig. 3.64)

Detailed analysis of the available NMR data was carried out in the textbook [1, 2] which suggested there were two aromatic rings (1,-AR and 1,4-AR) in the molecule under analysis. This information was used for manual molecular structure elucidation. Figure 3.64 shows the presence of four *light blue* carbon atoms and many ambiguous HMBC connectivities caused by overlapping signals at 7.19 ppm in the ^1H NMR spectrum (see Table 3.36, underlined ^1H chemical shifts). Strict Structure Generation was initiated from the initial MCD. Results: $k = 6 \rightarrow 6 \rightarrow 6$, $t_g = 0.8$ s. The ranked output structure file is presented in Fig. 3.65.

Figure 3.65 shows that structure #1 is identical to structure 3.39 but its average and maximum deviations are surprisingly high. This fact looks especially odd if we take into account that the structure contains no unusual chemical moieties. To immediately detect the “weak points” of structure #1 a function was used which produces a picture of the structure (see structure 3.40 below) where chemical shift prediction accuracy is marked in different colors for the carbon atoms:



3.40

The picture shows that the problem is in the bad prediction of the chemical shifts of atoms C 131.40 and C 112.5. As mentioned above (see Chap. 1, Sect. 1.4.1.4) a **Chemical Shift Calculation Protocol** can help to understand the cause of the great

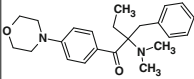
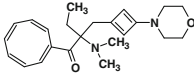
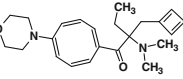
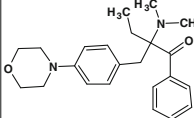
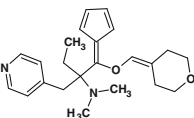
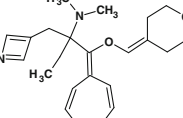
		
$d_A(^{13}\text{C}): 4.303$ $d_N(^{13}\text{C}): 4.412$ $d_I(^{13}\text{C}): 4.231$ $\text{max_}d_A(^{13}\text{C}): 19.210$	$d_A(^{13}\text{C}): 4.716$ $d_N(^{13}\text{C}): 5.495$ $d_I(^{13}\text{C}): 4.860$ $\text{max_}d_A(^{13}\text{C}): 19.930$	$d_A(^{13}\text{C}): 5.577$ $d_N(^{13}\text{C}): 4.535$ $d_I(^{13}\text{C}): 6.565$ $\text{max_}d_A(^{13}\text{C}): 18.320$
		
$d_A(^{13}\text{C}): 5.834$ $d_N(^{13}\text{C}): 5.504$ $d_I(^{13}\text{C}): 5.270$ $\text{max_}d_A(^{13}\text{C}): 19.790$	$d_A(^{13}\text{C}): 9.308$ $d_N(^{13}\text{C}): 8.996$ $d_I(^{13}\text{C}): 9.129$ $\text{max_}d_A(^{13}\text{C}): 42.200$	$d_A(^{13}\text{C}): 12.665$ $d_N(^{13}\text{C}): 11.298$ $d_I(^{13}\text{C}): 11.143$ $\text{max_}d_A(^{13}\text{C}): 45.670$

Fig. 3.65 Challenge 19: Ranked output structure file (Strict Structure Generation)

discrepancy between the calculated and experimental chemical shift values. Figures 3.66 and 3.67 display protocols created by the program for chemical shifts 131.4 and 112.5 correspondingly.

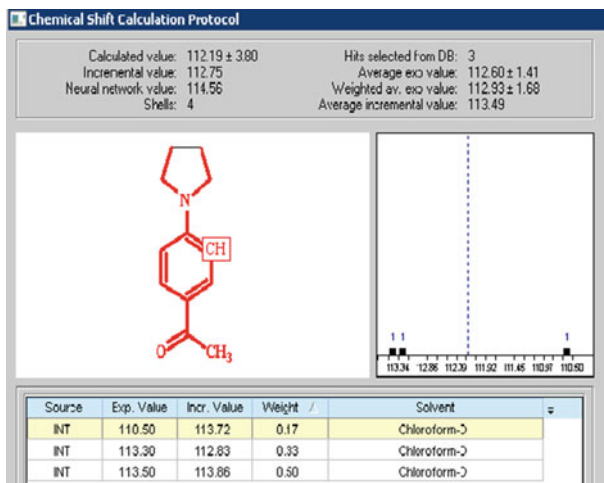


Fig. 3.66 Challenge 19: Protocol of ^{13}C chemical shift prediction for carbon atom C 131.4

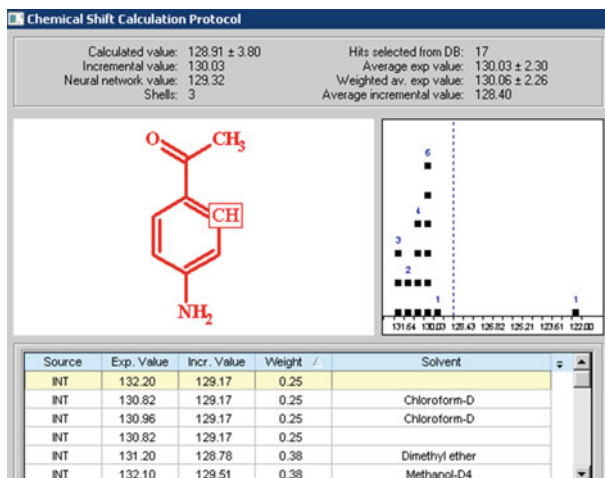
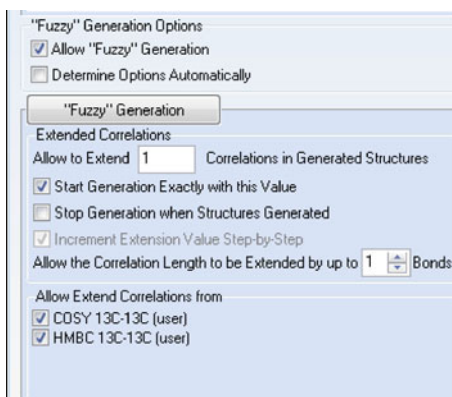


Fig. 3.67 Challenge 19: The ^{13}C chemical shift prediction protocol for carbon atom C 112.5

Figures 3.66 and 3.67 show clearly that the considered chemical shifts should be exchanged. However, the automatic assignment was carried out by the program in exact correspondence with the constraints that are imposed on the generated structures by HMBC and COSY correlations. Consequently, it is very probable that there are connectivities of nonstandard length (see Sects. 1.2.2 and 2.1.4). To circumvent this problem Fuzzy Structure Generation (FSG) was run from the MCD. The following Options of FSG were set: $m = 1$, $a = 1$, which means that for the first step the existence of at least one nonstandard connectivity (NSC) was suggested, which should be augmented by one chemical bond. A part of the dialog window **CSB Generator Options** is shown in Fig. 3.68.

The results of FSG gave: $k = 37 \rightarrow 35 \rightarrow 22$, $t_g = 8$ s, and the program reported that 1 from 17 connectivities was extended during structure generation, and 17

Fig. 3.68 Challenge 19: The dialog window for **Fuzzy Generation Options**



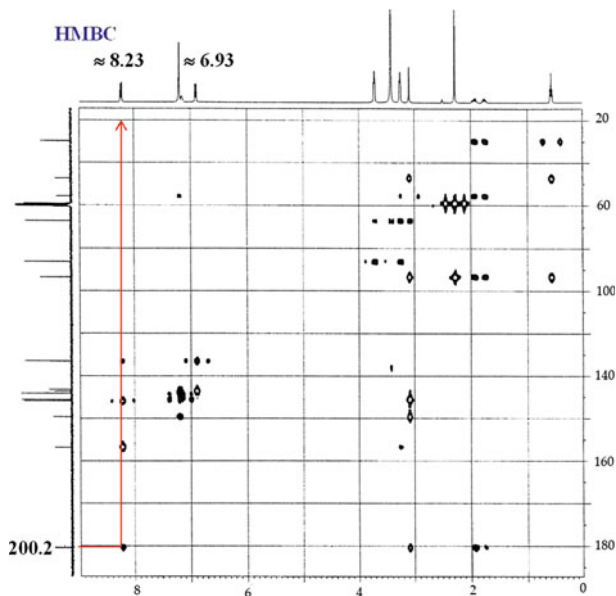


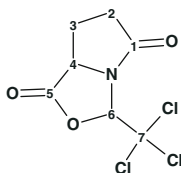
Fig. 3.70 Challenge 19: HMBC spectrum

is frequently violated, and that is a case in the current problem. Figure 3.70 shows the HMBC pattern where the correlation of H 8.23 to C 200.2 is drawn with red lines.

We see that the intensity of this HMBC signal corresponding to a $^4J_{\text{CH}}$ coupling constant has similar intensity to the $^{2-3}J_{\text{CH}}$ signals. This example visually demonstrates the important role of nonstandard correlations in molecular structure elucidation (manual or computer-aided). This warns the researcher to be careful when operating with 2D NMR data. On the other hand the example shows how the CASE approach allows for detection of a misassignment and how it can be corrected. Furthermore, we will see many examples where FSG helps to correctly elucidate the structure of an unknown whose 2D NMR spectra contain many nonstandard correlations of different and unknown length.

3.20 Challenge 20

The structure of the “unknown”:



3.42

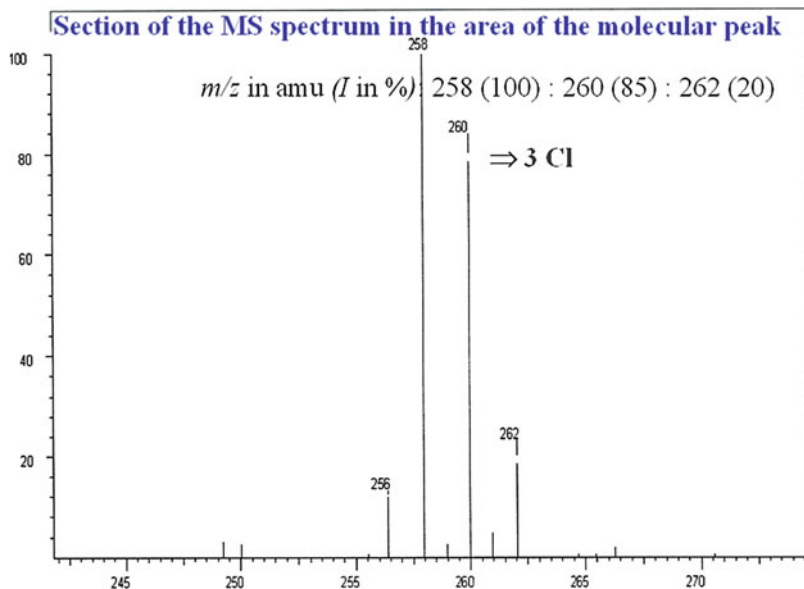


Fig. 3.71 Challenge 20: The molecular ion cluster in the mass spectrum (peak intensities are scaled)

Some difficulties were encountered during the molecular formula determination because the MS peaks of the molecular ion cluster have very low intensities. The peak m/z 258 was tentatively assigned as the molecular ion $[\text{M}]^+ = 258$ amu and the peak intensity ratio $[\text{M}]:[\text{M}+2]:[\text{M}+4]$ shown in Fig. 3.71 suggests the presence of three atoms of chlorine while bromine was not absolutely excluded. The IR spectrum contains a very strong band at $3,440 \text{ cm}^{-1}$ and a band of medium intensity at $1,632 \text{ cm}^{-1}$, both of them are probably accounted for by water absorption. Strictly speaking, in such a case we can say nothing about the presence or absence of OH, NH, and NH_2 groups (Table 3.37).

As explained in the textbook [1, 2] when all suggested elements of the chemical composition were summed it was realized that the true molecular peak was $m/z = 257$ which was not observed in the mass spectrum. Therefore, for the sake of safety, the following ranges of atom numbers and DBE values were used: C (7–8), H (6–7), O (2–5), N (0–5), Cl (0–3), Br (0–2), DBE (2–5). To remove any doubts regarding the value of the molecular ion the **Molecular Formula Generator** was run three times with $m/z = 257$, 258, and 259. As a result the single molecular formula $\text{C}_7\text{H}_6\text{NO}_3\text{Cl}_3$ was generated for $m/z = 257$, while in other cases the program informed that no molecular formula was generated for the specified data.

To elucidate the structure, ^{13}C , ^1H , HSQC, and HMBC data were used (Table 3.38).

Table 3.37 Challenge 20: Molecular Formula Inference

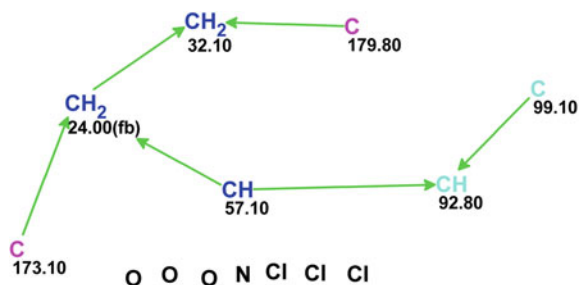
Factual data	Conclusions
MS: $[M]^+ = 258?$	$n_N = 0$ or 2 or 4
MS: $[M]: [M+2]: [M+4] = 100:85:20$	$3 \times \text{Cl}$ are expected
MS: 10 peaks are even	N is probable
^{13}C : no splitting	No F
^{13}C : 7 signals	Min. $n_C = 7$
^{13}C DEPT and ^1H : $\Sigma\text{H} = 6$	Min. $n_H = 6$
IR: $3,440\text{ cm}^{-1}$ very strong (H_2O) $1,632\text{ cm}^{-1}$ medium	OH, NH, NH_2 possible no CN, no NO_2
IR: $1811, 1737\text{ cm}^{-1}$	CO, min $n_O = 2$
Ranges of atom numbers Real $[M]^+ = 257$ [2]	C (7–8), H (6–7), O (2–5), N (0–5), Cl (0–3), Br (0–2), DBE (2–5)
Molecular Formulas Generated	$\text{C}_7\text{H}_6\text{NO}_3\text{Cl}_3$, RDBE = 4

Table 3.38 Challenge 20: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	179.8	174.71	C	–	–	–
C2	32.1	31.3	CH_2	2.4	u	C3, C1
C2	32.1	31.3	CH_2	2.8	u	–
C3	24	25.68	CH_2	2.54	u	–
C3	24	25.68	CH_2	2.3	u	C4, C5
C4	57.1	61.2	CH	4.79	u	–
C5	173.1	171.71	C	–	–	–
C6	92.8	95.26	CH	6.05	u	C4, C7
C7	99.1	96.05	C	–	–	–

The MCD is presented in Fig. 3.72.

No MCD edits were made. Structure generation from the MCD gave the following results: $k = 156 \rightarrow 124 \rightarrow 92$, $t_g = 0.08$ s. The top six structures of the ranked output file are presented in Fig. 3.73.

Fig. 3.72 Challenge 20: Molecular connectivity diagram

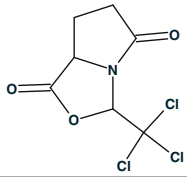
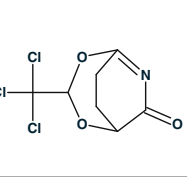
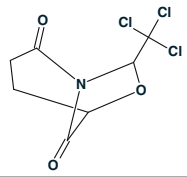
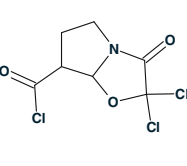
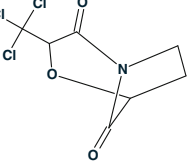
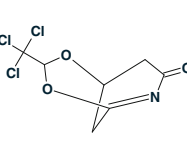
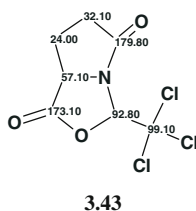
		
$d_A(^{13}\text{C}): 2.653$ $d_N(^{13}\text{C}): 1.654$ $d_I(^{13}\text{C}): 2.406$ $\text{max_}d_A(^{13}\text{C}): 5.090$	$d_A(^{13}\text{C}): 4.706$ $d_N(^{13}\text{C}): 5.549$ $d_I(^{13}\text{C}): 8.244$ $\text{max_}d_A(^{13}\text{C}): 13.850$	$d_A(^{13}\text{C}): 5.640$ $d_N(^{13}\text{C}): 5.673$ $d_I(^{13}\text{C}): 6.457$ $\text{max_}d_A(^{13}\text{C}): 18.080$
		
$d_A(^{13}\text{C}): 6.280$ $d_N(^{13}\text{C}): 6.023$ $d_I(^{13}\text{C}): 5.799$ $\text{max_}d_A(^{13}\text{C}): 20.210$	$d_A(^{13}\text{C}): 6.633$ $d_N(^{13}\text{C}): 6.705$ $d_I(^{13}\text{C}): 7.606$ $\text{max_}d_A(^{13}\text{C}): 16.770$	$d_A(^{13}\text{C}): 6.743$ $d_N(^{13}\text{C}): 8.123$ $d_I(^{13}\text{C}): 6.587$ $\text{max_}d_A(^{13}\text{C}): 11.880$

Fig. 3.73 Challenge 20: Top six structures of the ranked output file

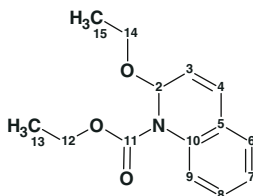
The figure shows that ^{13}C chemical shift prediction confidently selected the correct structure. The chemical shift assignment is shown below (**3.43**)



The total number N of isomers corresponding to the molecular formula $\text{C}_7\text{H}_6\text{NO}_3\text{Cl}_3$ is equal to 678,461,822. Hence at $k = 92$, it turns out that $\mu = 0.78$ (the moiety of structural information extracted by StrucEluc) and $\mu (^{13}\text{C}_{\text{calc}}) = 0.22$ (the contribution from ^{13}C chemical shift prediction).

3.21 Challenge 21

The structure of the “unknown” (Table 3.39):



3.44

1D and 2D NMR data used for the molecular structure elucidation are presented in Table 3.40.

Figure 3.74 presents the MCD.

In Fig. 3.74 the HMBC and COSY connectivities connected to two carbon atoms accidentally coincide with the ^{13}C chemical shift at 127.40 ppm and are ambiguous (marked by dotted lines). Two carbon atoms—C 77.90 and 154.6—have no connectivities at all and one carbon C 125.5 is colored in light blue (sp^2 or sp^3). It is expected that the structure generation from this MCD will produce many structures, the majority of which will be wrong. To prevent saving deliberately incorrect structures, structure generation was accompanied by ^{13}C chemical shift prediction. In so doing the structures that have large average deviations (see Sect. 2.2.1 and Fig. 2.10) are ignored. Result: $k = 10,882 \rightarrow 48 \rightarrow 12$ while the total time for structure generation and ^{13}C chemical shift prediction for ~ 11 thousand structures is about 12 s. Eight of the twelve structures from the ranked output file are presented in Fig. 3.75.

We see that the ^{13}C chemical shift prediction correctly selected the true structure. The most similar structure is placed in 11th position and its maximum deviation is

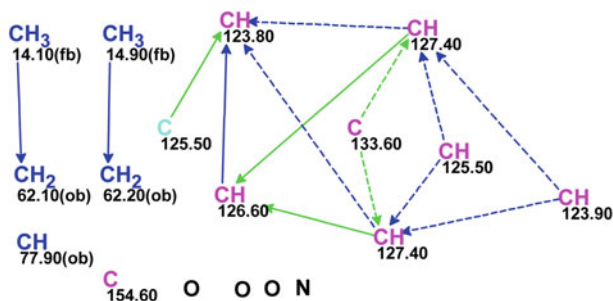
Table 3.39 Challenge 21: Molecular Formula Inference

Factual data	Conclusions
MS: $[\text{M}]^+ = 247$ odd number	$n_{\text{N}} = 1$ or 3 or 5
MS: no $[\text{M}+2]^+$, no isotope pattern	no Cl, no Br, no S
MS: 5 peaks are even	N is probable
^{13}C : no splitting	No F
^{13}C and HSQC: 14 signals	Min $n_{\text{C}} = 14$
^{13}C DEPT and ^1H : $\Sigma\text{H} = 17$	Min. $n_{\text{H}} = 17$
IR: $3,440\text{ cm}^{-1}$ very strong	H_2O , OH, NH possible, no CN, no NO_2
IR: $1704, 1490\text{ cm}^{-1}$ (strong)	CO, min $n_{\text{O}} = 1$, min DBE = 5, Benzene ring (?)
Ranges of atom numbers	C (14–16); H (17–20), O (1–5), N (1–5)
Molecular Formula Generated	$\text{C}_{14}\text{H}_{17}\text{NO}_3$, DBE (5–10)

Table 3.40 Challenge 21: Spectroscopic NMR data

Label ^a	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C2	77.9	79.47	CH	6.19	u	–	–
C3	<u>125.5</u>	119.69	CH	6.18	u	6.75	–
C4	<u>127.4</u>	127.45	CH	6.75	u	6.18	–
C5	<u>125.5</u>	123.33	C	–	–	–	–
C6	126.6	126.87	CH	7.23	u	7.09	C4, C8
C7	123.8	123.88	CH	7.09	u	7.30, 7.23	C5
C8	<u>127.4</u>	128.8	CH	7.3	u	7.09, 7.65	C10
C9	123.9	120.82	CH	7.65	u	7.3	–
C10	133.6	137.02	C	–	–	–	–
C11	154.6	152.77	C	–	–	–	–
C12	62.1	61.9	CH ₂	4.34	u	1.37	–
C13	14.1	14.16	CH ₃	1.37	u	4.34	–
C14	62.2	63.18	CH ₂	3.68	u	1.17	–
C15	14.9	15.04	CH ₃	1.17	u	3.68	–

^a Note that the atom numbering is started from C2 as in the textbook [2] (see structure 3.44)

**Fig. 3.74** Challenge 21: Molecular connectivity diagram

about 20 ppm. The degree of candidate structure correspondence to the ¹³C experimental spectrum can be visually evaluated not only by the command which marks the atoms by different colors depending on the difference of a predicted chemical shift from the experimental one (see Sect. 2.2.1.1 and Fig. 3.8 in Sect. 3.1). For this goal a **Graph Window** is also provided in the Structure Elucidator program. The system possesses a function that displays a linear regression line showing the dependence of calculated chemical shifts with the experimental ones, the regression statistical data is collected and displayed in the main program window. To initiate

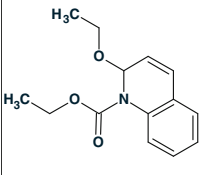
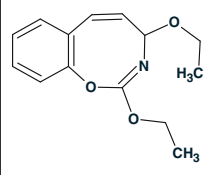
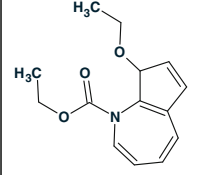
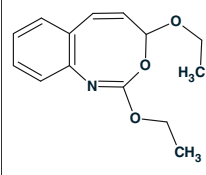
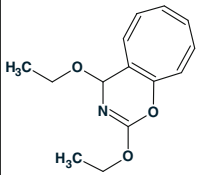
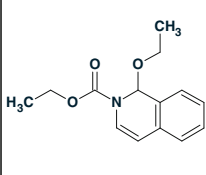
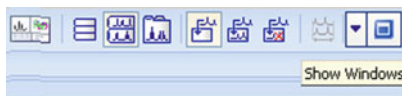
<p>1</p> 	<p>2</p> 	<p>3</p> 
<p>$d_A(^{13}\text{C})$: 1.504 $d_N(^{13}\text{C})$: 1.296 $d_I(^{13}\text{C})$: 1.510 $\text{max_}d_A(^{13}\text{C})$: 5.810</p>	<p>$d_A(^{13}\text{C})$: 3.144 $d_N(^{13}\text{C})$: 2.308 $d_I(^{13}\text{C})$: 3.563 $\text{max_}d_A(^{13}\text{C})$: 13.160</p>	<p>$d_A(^{13}\text{C})$: 3.362 $d_N(^{13}\text{C})$: 2.974 $d_I(^{13}\text{C})$: 3.820 $\text{max_}d_A(^{13}\text{C})$: 15.210</p>
<p>4</p> 	<p>6</p> 	<p>11</p> 
<p>$d_A(^{13}\text{C})$: 3.447 $d_N(^{13}\text{C})$: 3.563 $d_I(^{13}\text{C})$: 3.637 $\text{max_}d_A(^{13}\text{C})$: 14.970</p>	<p>$d_A(^{13}\text{C})$: 3.561 $d_N(^{13}\text{C})$: 4.484 $d_I(^{13}\text{C})$: 2.612 $\text{max_}d_A(^{13}\text{C})$: 10.440</p>	<p>$d_A(^{13}\text{C})$: 4.311 $d_N(^{13}\text{C})$: 3.374 $d_I(^{13}\text{C})$: 3.097 $\text{max_}d_A(^{13}\text{C})$: 19.330</p>

Fig. 3.75 Challenge 21: Eight of twelve structures from the ranked output file

this function it is necessary to press the button **Show Windows** on the right end of the toolbar (see below) and then choose **Graph Window** in the pulldown menu.



The functions of the buttons arranged on the toolbar of the **Graph Window** are intuitively clear (see below) and the purposes of these buttons are displayed by tool tips.



The regression lines built for structures #1 and #11 on the basis of predictions computed by all three methods (HOSE, Increments and Neural Nets) are presented in Figs. 3.76 and 3.77 correspondingly.

Comparison of the graphs shows clearly that for the wrong structure #11 the maximum deviation corresponds to C 8 (structure 3.44) and there is a poorly predicted shift for C 10 (C 133.6). Other predictions fit the line $Y = X$ fairly well. At the same time the R^2 values calculated for the correct structure #1 is 0.998 while rather lower values were found for the wrong structure—from 0.976 to 0.984.

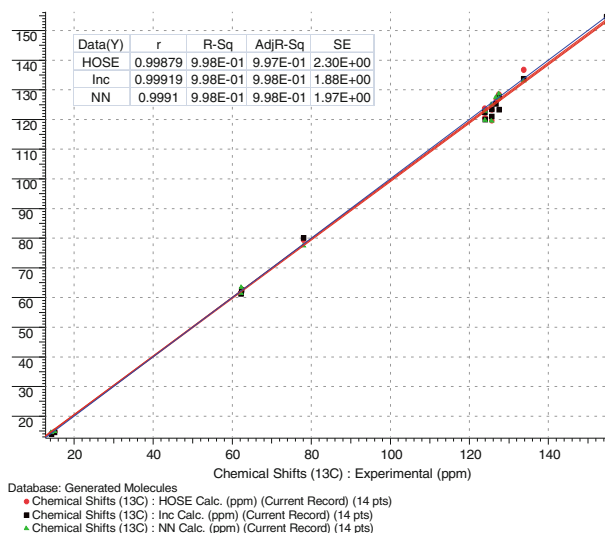


Fig. 3.76 Challenge 21: The comparison of calculated and experimental ^{13}C chemical shifts for the correct structure #1

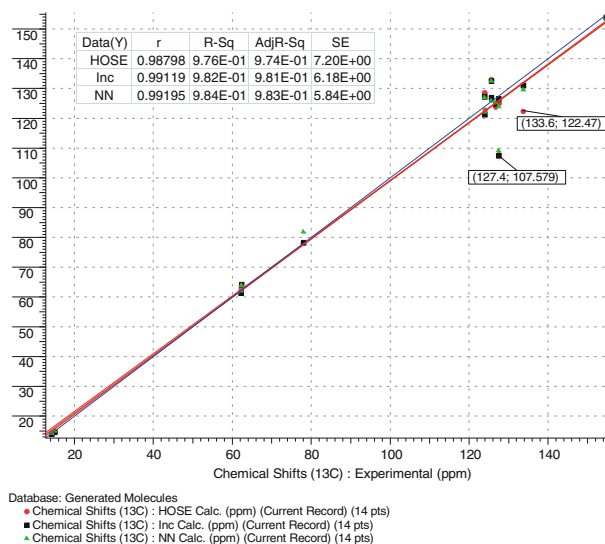
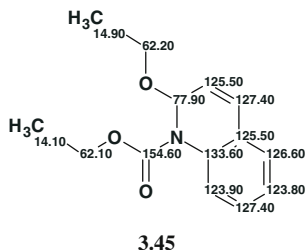


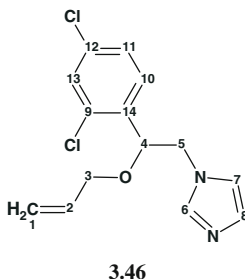
Fig. 3.77 Challenge 21: The comparison of calculated and experimental ^{13}C chemical shifts for the similar wrong structure #11. The numbers in the *boxes* show the experimental and calculated values of the corresponding chemical shifts

R^2 values are frequently used as quantitative indicators of the correspondence between the experimental and calculated chemical shifts. The ^{13}C chemical shift assignment of the elucidated structure is shown below (3.45)



3.22 Challenge 22

The structure of the “unknown”:



The molecular ion cluster shown in Fig. 3.78 is typical for a molecule containing two chlorine atoms. There is no reliable evidence of the presence or absence of oxygen and nitrogen atoms, therefore the ranges of atom numbers were set as 0–5 and 0–4 for these atoms correspondingly. Bearing in mind that the ^{13}C signals of quaternary carbons can be missed due to their very low intensity the range of the number of carbons was set as 14–16. As a result the Molecular Formula Generator produced two molecular formulae of close accurate molecular masses: $\text{C}_{14}\text{H}_{14}\text{N}_2\text{OCl}_2$ (296.048318) and $\text{C}_{15}\text{H}_{14}\text{O}_2\text{Cl}_2$ (296.037087). Structure Elucidator provides the aids to try both of them and find the best solution (Table 3.41).

The experimental data extracted from the Extra Materials and used for CASE analysis are collected in Table 3.42.

We should remember that there are several possible ways to solve this problem (see Sect. 3.9).

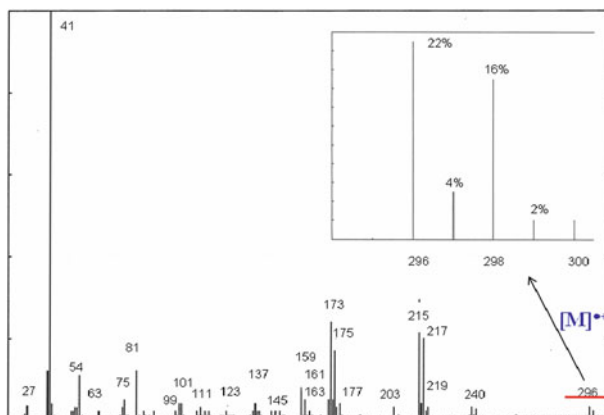


Fig. 3.78 Challenge 20: Mass spectrum of the unknown. The area of the molecular ion cluster is expanded in the *upper part* of the figure

Table 3.41 Challenge 22: Molecular Formula Inference

Factual data	Conclusions
MS: $[M]^+$ = 296, even number	$n_N = 0$ or 2 or 4
MS: $[M]:[M+2]:[M+4]$ 100:78:10	$n_{Cl} = 2$
MS: 3 peaks are even	N is probable
^{13}C : no splitting, MS: no mass differences of 19 or 20	No F
^{13}C : 14 signals	Min. $n_C = 14$
^{13}C DEPT and 1H : $\Sigma H = 14$	Min. $n_H = 14$
IR: $3,437\text{ cm}^{-1}$ very strong, $1,646\text{ cm}^{-1}$ medium	H_2O , OH, NH possible, C=C (?), no CN, NO_2
IR: 1589 cm^{-1} , 1505 cm^{-1}	AR, no CO, no $C\equiv C$, min DBE 4
Ranges of atom numbers	C (14–16), H (14–16), O (0–5), Cl (2), N (0–4), DBE (4–10)
Molecular Formulas Generated	1. $C_{14}H_{14}N_2OCl_2$ 2. $C_{15}H_{14}O_2Cl_2$ DBE = 8

Again, as in the case of Challenge 9, to provide a clear explanation we will use the third way of solving the problem—both molecular formulas will be examined separately.

Solution 1 $C_{14}H_{14}N_2OCl_2$.

The MCD is presented in Fig. 3.79.

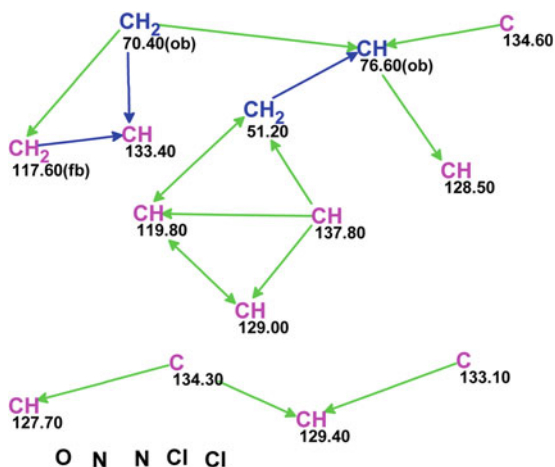
Structure generation initiated from the MCD, Fig. 3.79, gave the following results: $k = 12,417 \rightarrow 9,641 \rightarrow 2,735$, $t_g = 12$ s, (prediction of ^{13}C chemical shifts using simultaneously incremental and neural nets approaches took 37 s for 9,641 structures).

The eight top structures of the ranked output file are presented in Fig. 3.80. The correct structure was selected from among ca. 2,700 candidates but all eight structures are

Table 3.42 Challenge 22: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	117.6	117.07	CH ₂	5.11	u	5.7	C3
C1	117.6	117.07	CH ₂	5.12	u	–	–
C2	133.4	132.89	CH	5.7	u	5.11, 3.80	C3
C3	70.4	70.19	CH ₂	3.88	u	–	–
C3	70.4	70.19	CH ₂	3.8	u	5.7	–
C4	76.6	78.01	CH	4.88	u	3.98	C5, C9, C3
C5	51.2	52.62	CH ₂	4.14	u	–	–
C5	51.2	52.62	CH ₂	3.98	u	4.88	C6, C7
C6	137.8	137.41	CH	7.39	u	–	–
C7	119.8	119.2	CH	6.87	u	–	C6, C8, C5
C8	129	128.68	CH	6.96	u	–	C6, C7
C9	134.6	133.64	C	–	–	–	–
C10	128.5	129.33	CH	7.22	u	–	C4
C11	127.7	128.09	CH	7.23	u	–	C12
C12	134.3	133.13	C	–	–	–	–
C13	129.4	128.54	CH	7.36	u	–	C12, C14
C14	133.1	133	C	–	–	–	–

Fig. 3.79 Challenge 22:
Molecular connectivity
diagram created from the
molecular formula
 $C_{14}H_{14}N_2OCl_2$



characterized by average deviations whose values are quite small. Note that structures #1–#5 differ only by the positions of chlorine atoms on the benzene ring while structures #7 and #8 are very exotic and hardly can compete with the best structure.

Solution 2 $C_{15}H_{14}O_2Cl_2$.

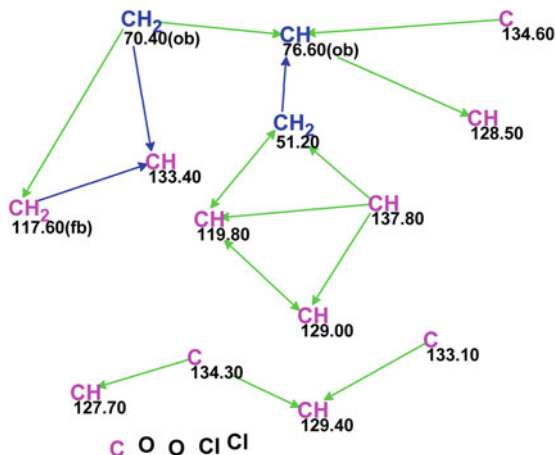
The MCD created for the second solution is presented in Fig. 3.81.

The MCD contains an additional quaternary carbon atom which is marked in light blue because the check box **Allow sp Carbons** was deselected in the dialog

1	2	3	4
$d_A(^{13}\text{C}): 0.693$ $d_N(^{13}\text{C}): 0.722$ $d_I(^{13}\text{C}): 0.945$ $\text{max}_dA(^{13}\text{C}): 1.420$	$d_A(^{13}\text{C}): 1.115$ $d_N(^{13}\text{C}): 1.277$ $d_I(^{13}\text{C}): 1.029$ $\text{max}_dA(^{13}\text{C}): 3.410$	$d_A(^{13}\text{C}): 1.241$ $d_N(^{13}\text{C}): 1.152$ $d_I(^{13}\text{C}): 1.219$ $\text{max}_dA(^{13}\text{C}): 3.880$	$d_A(^{13}\text{C}): 1.300$ $d_N(^{13}\text{C}): 1.293$ $d_I(^{13}\text{C}): 1.371$ $\text{max}_dA(^{13}\text{C}): 4.320$
5	6	7	8
$d_A(^{13}\text{C}): 1.643$ $d_N(^{13}\text{C}): 1.520$ $d_I(^{13}\text{C}): 1.818$ $\text{max}_dA(^{13}\text{C}): 8.510$	$d_A(^{13}\text{C}): 2.316$ $d_N(^{13}\text{C}): 2.619$ $d_I(^{13}\text{C}): 2.738$ $\text{max}_dA(^{13}\text{C}): 9.880$	$d_A(^{13}\text{C}): 2.400$ $d_N(^{13}\text{C}): 2.245$ $d_I(^{13}\text{C}): 2.909$ $\text{max}_dA(^{13}\text{C}): 12.950$	$d_A(^{13}\text{C}): 2.474$ $d_N(^{13}\text{C}): 2.078$ $d_I(^{13}\text{C}): 2.717$ $\text{max}_dA(^{13}\text{C}): 10.430$

Fig. 3.80 Challenge 22: Eight top structures of the ranked output file generated from the molecular formula $\text{C}_{14}\text{H}_{14}\text{N}_2\text{OCl}_2$

Fig. 3.81 Challenge 22:
Molecular connectivity
diagram created from the
molecular formula
 $\text{C}_{15}\text{H}_{14}\text{O}_2\text{Cl}_2$



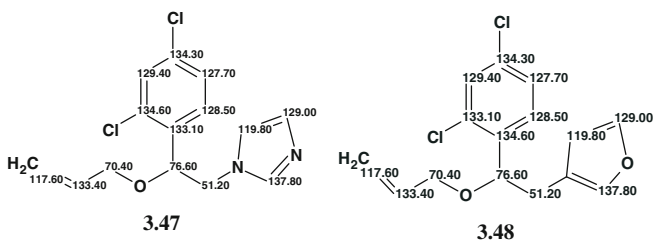
window **Create MCDs Options**, i.e., it is assumed that this atom can have sp^2 or sp^3 hybridizations. The structure generation from the MCD, Fig. 3.81, was completed with the results: $k = 17,772 \rightarrow 14,050 \rightarrow 1,617$, $t_g = 13.5$ s. The eight top structures of the ranked output file produced with the molecular formula $\text{C}_{15}\text{H}_{14}\text{O}_2\text{Cl}_2$ are presented in Fig. 3.82.

Average deviations found for structure #1 shown in Fig. 3.82 are much greater than those calculated for the true structure (Fig. 3.80), which allows us to establish

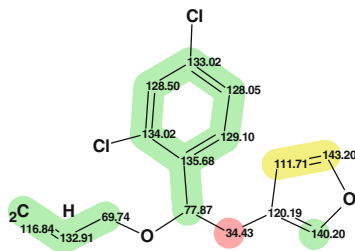
1	2	3	4
$d_A(^{13}\text{C}): 3.555$ $d_N(^{13}\text{C}): 3.720$ $d_I(^{13}\text{C}): 4.019$ $\text{max_}d_A(^{13}\text{C}): 16.770$	$d_A(^{13}\text{C}): 3.799$ $d_N(^{13}\text{C}): 4.092$ $d_I(^{13}\text{C}): 4.353$ $\text{max_}d_A(^{13}\text{C}): 16.080$	$d_A(^{13}\text{C}): 3.920$ $d_N(^{13}\text{C}): 6.156$ $d_I(^{13}\text{C}): 4.494$ $\text{max_}d_A(^{13}\text{C}): 15.340$	$d_A(^{13}\text{C}): 4.253$ $d_N(^{13}\text{C}): 4.271$ $d_I(^{13}\text{C}): 4.403$ $\text{max_}d_A(^{13}\text{C}): 16.540$
5	6	7	8
$d_A(^{13}\text{C}): 4.289$ (v.14.0) $d_N(^{13}\text{C}): 4.011$ $d_I(^{13}\text{C}): 4.558$ $\text{max_}d_A(^{13}\text{C}): 16.080$	$d_A(^{13}\text{C}): 4.551$ (v.14.0) $d_N(^{13}\text{C}): 5.486$ $d_I(^{13}\text{C}): 5.260$ $\text{max_}d_A(^{13}\text{C}): 21.070$	$d_A(^{13}\text{C}): 4.601$ (v.14.0) $d_N(^{13}\text{C}): 4.586$ $d_I(^{13}\text{C}): 4.896$ $\text{max_}d_A(^{13}\text{C}): 15.550$	$d_A(^{13}\text{C}): 4.894$ (v.14.0) $d_N(^{13}\text{C}): 6.754$ $d_I(^{13}\text{C}): 4.864$ $\text{max_}d_A(^{13}\text{C}): 19.830$

Fig. 3.82 Challenge 22: The eight top structures of the ranked output file generated from the molecular formula $\text{C}_{15}\text{H}_{14}\text{O}_2\text{Cl}_2$

both the true molecular and structural formulae of the unknown. Below the best structures found in both problem solutions are shown for comparison: **3.47**—#1 (Fig. 3.80), $\text{C}_{14}\text{H}_{14}\text{N}_2\text{OCl}_2$ and **3.48**—#1 (Fig. 3.82), $\text{C}_{15}\text{H}_{14}\text{O}_2\text{Cl}_2$.



We see that the structures are very similar and the ^{13}C chemical shift assignment is the same for carbons which are “topological twins” (exclusions are C 133.1 and C 134.6 which are permuted). The additional carbon atom (unassigned) was incorporated into the furan ring. Structure **3.49** (#1, $\text{C}_{15}\text{H}_{14}\text{O}_2\text{Cl}_2$) supplied with ^{13}C chemical shifts calculated by the HOSE approach establishes the expected chemical shift of the additional carbon atom (120 ppm) and detects carbons whose predicted chemical shifts differ significantly from the experimental values (see colored spheres).



3.49

The considered example demonstrates a methodology which is efficient for obtaining the correct structural formula of an unknown when several molecular formulae can be expected. The solution was found automatically in a very short time.

References

1. Reichenbacher M, Popp J (2012) Challenges in molecular structure determination. Springer, Heidelberg
2. "Extra Materials", <http://extras.springer.com/2012/978-3-642-24389-9>. Accessed 7 Jan 2015
3. Elyashberg ME, Williams AJ, Blinov KA (2012) Contemporary computer-assisted approaches to molecular structure elucidation, vol 1. New Developments in NMR. RSC Publishing, Cambridge
4. www.pubchem.ncbi.nlm.nih.gov
5. www.chemspider.com
6. emolecules.com

Part III
Solution of Real World Problems

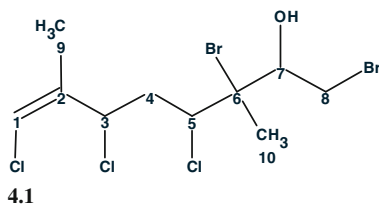
Chapter 4

Structure Elucidation Using Strict Structure Generation

Abstract Part III is divided into two chapters—Chaps. 4 and 5. These chapters are the most important for those who want to become proficient in the routine application of CASE analysis for solving structural problems which appear in analytical laboratories. Here CASE-based solutions to 66 *real-world* structural problems are fully explained. Spectroscopic data for the problems 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, the problems we selected were primarily related to molecules possessing *unique* or *unprecedented* skeletons. Spectroscopic data for these problems are also (as described in Chap. 3) 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. This chapter describes 33 real-world 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 $^2J_{\text{CH}}$ and $^{2-3}J_{\text{HH}}$ correspondingly.

4.1 Costatol D

A new natural product was isolated by Motti and co-workers [1] from the red algae *Plocamium costatum*, collected from South Australian waters. An extensive spectroscopic analysis (1D and 2D NMR and MS) allowed the researchers to determine the molecular formula ($\text{C}_{10}\text{H}_{15}\text{Br}_2\text{Cl}_3\text{O}$), and structure of this highly halogenated small molecule, Costatol D **4.1**. The absolute stereochemistry was determined by single-crystal X-ray crystallographic analysis [1].



The spectroscopic data used for challenging the Structure Elucidator software are presented in Table 4.1 and Fig. 4.1 where the Molecular Connectivity Diagram (MCD) is displayed.

Checking the MCD showed that the HMBC data were consistent and strict structure generation was initiated. Results: $k = 34 \rightarrow 34$, $t_g = 0.05$ s. Visual examination of the output structural file showed that all structures contained a $X-C=C-CH_3$ group in α -position. Because two substituent orientations relative to

Table 4.1 Costatol D: 1H and ^{13}C NMR spectroscopic data (CD₃OD)

Label	δC	δC_{calc}	CH	δH	M(J)	C HMBC
C1	116.2	115.51	CH	6.11	u	C9, C3, C2
C2	138.7	136.6	C	–	–	–
C3	57.9	57.51	CH	5.55	u	C2, C9, C1, C5
C4	41.1	39.26	CH ₂	2.94	u	C2, C6, C3
C4	41.1	39.26	CH ₂	2.02	u	C5
C5	64.5	63.36	CH	4.93	u	C10, C4, C6, C3
C6	75.5	72.74	C	–	–	–
C7	76.8	75.31	CH	4	u	C8
C8	37.4	33.98	CH ₂	3.85	u	C7
C8	37.4	33.98	CH ₂	3.55	u	C7
C9	16.1	16.39	CH ₃	1.93	u	C1, C3, C2
C10	22.5	22.97	CH ₃	1.75	u	C5, C6, C7

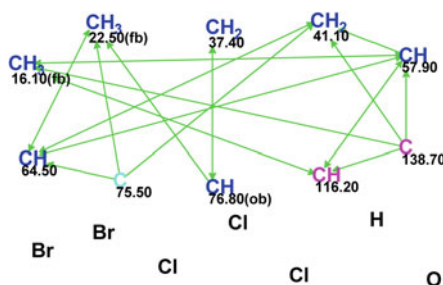


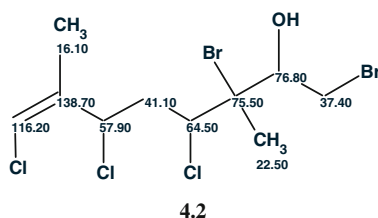
Fig. 4.1 Costatol D: Molecular connectivity diagram

1	2	3	4
$d_A(^{13}\text{C}): 1.917$ $d_N(^{13}\text{C}): 1.865$ $d_I(^{13}\text{C}): 2.501$ $\text{max_}d_A(^{13}\text{C}): 3.420$	$d_A(^{13}\text{C}): 2.049$ $d_N(^{13}\text{C}): 2.264$ $d_I(^{13}\text{C}): 1.830$ $\text{max_}d_A(^{13}\text{C}): 6.970$	$d_A(^{13}\text{C}): 2.527$ $d_N(^{13}\text{C}): 1.987$ $d_I(^{13}\text{C}): 2.826$ $\text{max_}d_A(^{13}\text{C}): 9.770$	$d_A(^{13}\text{C}): 2.710$ $d_N(^{13}\text{C}): 3.156$ $d_I(^{13}\text{C}): 3.211$ $\text{max_}d_A(^{13}\text{C}): 8.650$
5	6	7	8
$d_A(^{13}\text{C}): 2.927$ $d_N(^{13}\text{C}): 3.937$ $d_I(^{13}\text{C}): 2.552$ $\text{max_}d_A(^{13}\text{C}): 8.450$	$d_A(^{13}\text{C}): 2.934$ $d_N(^{13}\text{C}): 2.910$ $d_I(^{13}\text{C}): 2.450$ $\text{max_}d_A(^{13}\text{C}): 9.540$	$d_A(^{13}\text{C}): 3.004$ $d_N(^{13}\text{C}): 2.756$ $d_I(^{13}\text{C}): 4.049$ $\text{max_}d_A(^{13}\text{C}): 8.590$	$d_A(^{13}\text{C}): 3.190$ $d_N(^{13}\text{C}): 2.952$ $d_I(^{13}\text{C}): 2.693$ $\text{max_}d_A(^{13}\text{C}): 7.980$

Fig. 4.2 Costatol D: Top eight *cis*-structures of the ranked output file

the double bond are possible (*cis*- and *trans*-), ^{13}C NMR chemical shift prediction was performed for both configurations (*X* substituents were moved by hand to corresponding positions in the top ten structures of the preliminary ranked file). Figures 4.2 and 4.3 show the top eight structures of both ranked files.

Comparison of the figures allows us to conclude that the correct structure for **4.1** was distinguished in both cases, but the deviation values $d_A(^{13}\text{C}) = 1.917$ ppm (*cis*) and $d_A(^{13}\text{C}) = 1.459$ ppm (*trans*) confirm the priority of the *trans*-configuration. The correct structure **4.2** with the ^{13}C chemical shift assignments is presented below:



As the molecule under investigation contains only 16 skeletal atoms, the theoretically conceivable number of isomers associated with the given molecular formula can be calculated using the structure generator. It turned out that this number was equal to 66,983,298 (remember that the total number of all commercially

1	2	3	4
$d_A(^{13}\text{C}): 1.459$ $d_N(^{13}\text{C}): 1.914$ $d_I(^{13}\text{C}): 2.364$ $\text{max_}d_A(^{13}\text{C}): 3.420$	$d_A(^{13}\text{C}): 2.354$ $d_N(^{13}\text{C}): 2.893$ $d_I(^{13}\text{C}): 2.436$ $\text{max_}d_A(^{13}\text{C}): 10.060$	$d_A(^{13}\text{C}): 2.526$ $d_N(^{13}\text{C}): 3.965$ $d_I(^{13}\text{C}): 2.431$ $\text{max_}d_A(^{13}\text{C}): 8.450$	$d_A(^{13}\text{C}): 2.597$ $d_N(^{13}\text{C}): 2.017$ $d_I(^{13}\text{C}): 2.988$ $\text{max_}d_A(^{13}\text{C}): 10.160$
5	6	7	8
$d_A(^{13}\text{C}): 2.886$ $d_N(^{13}\text{C}): 2.932$ $d_I(^{13}\text{C}): 2.324$ $\text{max_}d_A(^{13}\text{C}): 9.540$	$d_A(^{13}\text{C}): 2.922$ $d_N(^{13}\text{C}): 3.784$ $d_I(^{13}\text{C}): 3.824$ $\text{max_}d_A(^{13}\text{C}): 10.020$	$d_A(^{13}\text{C}): 3.136$ $d_N(^{13}\text{C}): 3.657$ $d_I(^{13}\text{C}): 3.854$ $\text{max_}d_A(^{13}\text{C}): 9.900$	$d_A(^{13}\text{C}): 3.171$ $d_N(^{13}\text{C}): 2.791$ $d_I(^{13}\text{C}): 4.217$ $\text{max_}d_A(^{13}\text{C}): 9.900$

Fig. 4.3 Costatol D: First eight *trans*-structures of the ranked output file

available chemical compounds is approximately the same). It was interesting to learn which solution to the problem would be obtained if the HMBC data were omitted. For this goal all HMBC connectivities were deleted in the MCD (a new MCD can also be created with the HMBC spectrum switched off) and structure generation was repeated with the fragments O–Cl and O–Br placed into the User BadList. At the same time ^{13}C chemical shift calculation and structural filtering with a threshold of $d = 4$ ppm were allowed during structure generation. Results: $k = 1,541,400 \rightarrow 7,950 \rightarrow 3,670$, $t_g = 30$ min.

Again, as in the first program run, the top structures of the output structural file were ranked for *trans*- and *cis*-double bond configurations separately. Figure 4.4 shows that the correct structure was placed in first position by the HOSE code-based ^{13}C chemical shift prediction when all structures were *trans*-configuration.

1	2	3	4
$d_A(^{13}\text{C}): 1.459$ $d_N(^{13}\text{C}): 1.914$ $d_I(^{13}\text{C}): 2.364$ $\text{max_}d_A(^{13}\text{C}): 3.420$	$d_A(^{13}\text{C}): 1.564$ $d_N(^{13}\text{C}): 1.960$ $d_I(^{13}\text{C}): 1.562$ $\text{max_}d_A(^{13}\text{C}): 3.510$	$d_A(^{13}\text{C}): 1.676$ $d_N(^{13}\text{C}): 2.080$ $d_I(^{13}\text{C}): 2.110$ $\text{max_}d_A(^{13}\text{C}): 3.950$	$d_A(^{13}\text{C}): 1.676$ $d_N(^{13}\text{C}): 2.271$ $d_I(^{13}\text{C}): 3.009$ $\text{max_}d_A(^{13}\text{C}): 3.500$

Fig. 4.4 Costatol D: The top four structures of the ranked output file (*trans*-configurations) obtained when the HMBC data were ignored

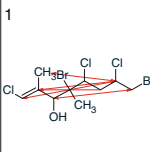
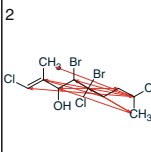
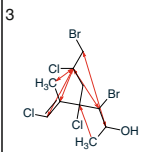
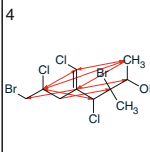
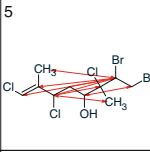
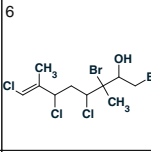
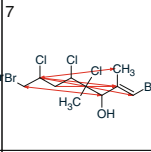
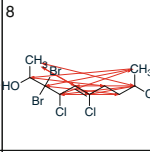
			
$d_A(^{13}\text{C}): 1.182$ $d_N(^{13}\text{C}): 1.712$ $d_l(^{13}\text{C}): 1.461$ $\text{max_}d_A(^{13}\text{C}): 2.500$	$d_A(^{13}\text{C}): 1.467$ $d_N(^{13}\text{C}): 1.714$ $d_l(^{13}\text{C}): 2.891$ $\text{max_}d_A(^{13}\text{C}): 3.330$	$d_A(^{13}\text{C}): 1.633$ $d_N(^{13}\text{C}): 2.297$ $d_l(^{13}\text{C}): 2.911$ $\text{max_}d_A(^{13}\text{C}): 4.090$	$d_A(^{13}\text{C}): 1.732$ $d_N(^{13}\text{C}): 3.073$ $d_l(^{13}\text{C}): 1.753$ $\text{max_}d_A(^{13}\text{C}): 4.020$
			
$d_A(^{13}\text{C}): 1.828$ $d_N(^{13}\text{C}): 1.905$ $d_l(^{13}\text{C}): 1.687$ $\text{max_}d_A(^{13}\text{C}): 4.240$	$d_A(^{13}\text{C}): 1.917$ $d_N(^{13}\text{C}): 1.865$ $d_l(^{13}\text{C}): 2.501$ $\text{max_}d_A(^{13}\text{C}): 3.420$	$d_A(^{13}\text{C}): 1.978$ $d_N(^{13}\text{C}): 1.989$ $d_l(^{13}\text{C}): 1.916$ $\text{max_}d_A(^{13}\text{C}): 11.020$	$d_A(^{13}\text{C}): 1.992$ $d_N(^{13}\text{C}): 2.117$ $d_l(^{13}\text{C}): 2.464$ $\text{max_}d_A(^{13}\text{C}): 3.450$

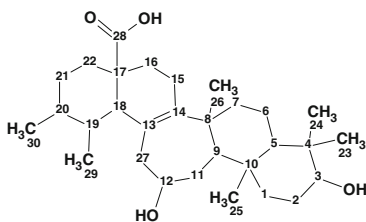
Fig. 4.5 Costatol D: The top eight structures of the ranked *cis*-configured file obtained when the HMBC data were ignored. HMBC connectivities are depicted by *arrows*

However, in the *cis*-configured file the correct structure occupied the sixth position which accounted for the fact that no HMBC constraints were imposed on the assigned carbon atom locations. Figure 4.5 shows the top eight structures of the ranked *cis*-configured file where the HMBC connectivities are depicted by arrows.

Figure 4.5 clearly demonstrates the crucial role of HMBC correlations in distinguishing the correct structure (4.1): all structures ranked before and after the correct one (#6) are characterized by a large number of unrealistic long nonstandard connectivities. In spite of the small average deviations these structures cannot be generated in accordance with structural constraints from the HMBC spectrum.

4.2 Ilesane

Wang et al. [2] isolated from the leaves of *Ilex latifolia* an *unusual* triterpene, ilelic acid (4.3), which represented a new type of triterpenoid with a seven-membered ring, named as “Ilesane”. Its structure with absolute configurations was elucidated by spectroscopic analysis and the modified Mosher’s method.



4.3

The molecular formula of **4.3** was determined to be $C_{30}H_{48}O_4$ on the basis of a quasimolecular ion at m/z 495.3444 $[M+Na]^+$ (calculated 495.3444 for $C_{30}H_{48}O_4Na$) in the HRESIMS. The degree of unsaturation was equal to seven. The IR spectrum of the unknown was acquired. The authors concluded that it “showed the characteristic absorptions attributable to carboxyl ($1,708\text{ cm}^{-1}$) and hydroxyl ($3,410\text{ cm}^{-1}$) groups”. However, the absorption band observed at $1,710\text{ cm}^{-1}$ in a complex molecule like ilesane can be related to either ketone, or ester or carboxyl group. Assignment of this frequency to a carboxyl group will be likely only by taking into account a broadband observed at $\sim 2,600\text{ cm}^{-1}$ characteristic for an acid hydroxyl group. A band at 203 nm observed in UV spectrum confirms the suggestion of the presence of carboxyl group.

The analysis of the 1H and ^{13}C NMR spectra revealed that **4.3** possessed 30 carbons including two olefinic carbons (δC 131.7 and 142.5) as well as one carbonyl carbon (δC 180.6), which suggested that the unknown could be a pentacyclic compound (RDBE = 7). The 1D and 2D NMR spectral data (the latter were presented in the article graphically as *selected* key COSY and HMBC correlations) are shown in Table 4.2.

The MCD is shown in Fig. 4.6.

MCD overview The MCD contains a significant number of ambiguous connectivities, which is accounted for by the presence of two pairs of overlapping chemical shifts of hydrogens attached to carbon atoms C2, C22 and C9, C16. Because only selected key HMBC and COSY correlations were available from the article [2], four quaternary carbons (38.6, 39.3, 43.8, and 49.9) have no connectivities and consequently all admissible positions of these atoms in the generated molecules should be tried. It is expected that the presence of ambiguous connectivities and “free” atoms on the MCD can increase the time for structure generation. On the other hand, all mentioned four atoms are automatically marked by the label “*fb*” from the Atom Property Characteristic Table (APCT). This can lead to a drastic reduction in the number of possible atomic combinations which simultaneously satisfy both the constraints imposed on the atom environments (chemical bonds with oxygens are forbidden for these atoms) and the COSY /HMBC structural constraints.

Table 4.2 Ilesane: NMR spectroscopic data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	HMBC
C1	40.1	37.78	CH ₂	1.14	u	1.82, 1.82	–
C1	40.1	37.78	CH ₂	1.92	u	–	–
C2	28.4	28.07	CH ₂	1.82	u	3.40, 1.14	–
C3	77.8	78.87	CH	3.4	u	1.82, 1.82	–
C4	39.3	40.14	C	–	–	–	–
C5	55.4	55.66	CH	0.9	u	1.34	–
C6	19.4	20.73	CH ₂	1.34	u	2.01, 0.90	–
C6	19.4	20.73	CH ₂	1.55	u	–	–
C7	42.3	38.41	CH ₂	2.01	u	1.34	–
C7	42.3	38.41	CH ₂	1.33	u	–	–
C8	43.8	40.35	C	–	–	–	–
C9	52.2	51.18	CH	1.94	u	1.93	C12
C10	38.6	39.31	C	–	–	–	–
C11	35.1	37.47	CH ₂	2.24	u	–	–
C11	35.1	37.47	CH ₂	1.93	u	4.39, 1.94, 1.94	C27
C12	69.4	72.41	CH	4.39	u	1.93, 2.66	–
C13	131.7	127.99	C	–	–	–	–
C14	142.5	138.56	C	–	–	–	–
C15	22.8	21.6	CH ₂	2.15	u	2.19	C13
C15	22.8	21.6	CH ₂	2.45	u	–	–
C16	24.2	25.68	CH ₂	2.19	u	2.15	C28
C16	24.2	25.68	CH ₂	1.94	u	1.93	–
C17	49.9	48.09	C	–	–	–	–
C18	58.4	51.52	CH	2.48	d(11.2)	1.22	C27, C14, C28
C19	39	43.25	CH	1.22	u	2.48, 1.06, 1.01	–
C20	38.7	38.92	CH	1.06	u	1.22, 1.48, 0.91	–
C21	31.1	31.31	CH ₂	1.48	u	1.06, 2.04	–
C22	34.8	35.08	CH ₂	1.82	u	3.40, 1.14	–
C22	34.8	35.08	CH ₂	2.04	u	1.48	C20, C28
C23	28.8	23.15	CH ₃	1.21	s	–	C5
C24	17	10.63	CH ₃	1.03	s	–	C23, C3
C25	16.6	24.31	CH ₃	0.94	s	–	C1, C5, C9
C26	19.7	24.51	CH ₃	1.12	s	–	C9, C14, C7
C27	49.1	44.82	CH ₂	2.66	u	4.39	C18, C11, C14
C27	49.1	44.82	CH ₂	3.24	u	–	–
C28	180.6	182.8	C	–	–	–	–
C29	18	16.34	CH ₃	1.01	d(6.5)	1.22	C18, C20
C30	20.2	18.09	CH ₃	0.91	d(6.5)	1.06	C19, C21

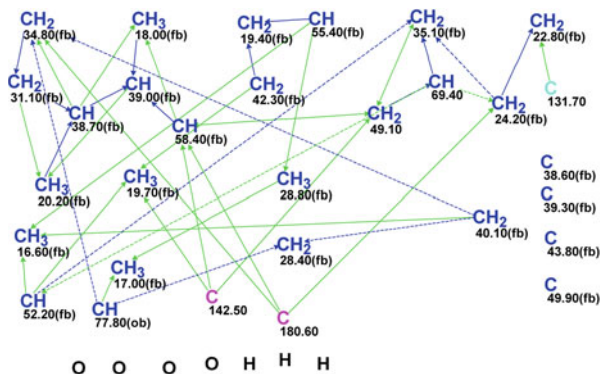
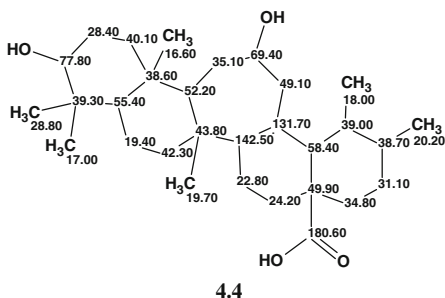


Fig. 4.6 Ilesane: Molecular connectivity diagram

No user editing of atom properties was performed and structure generation was carried out from the MCD without further adjustment. Result: $k = 24 \rightarrow 24 \rightarrow 1$, $t_g = 0.4$ s, $d_A = 2.65$, $d_N = d_I = 3$ ppm, and the single structure **4.4** coincided with structure **4.3**:

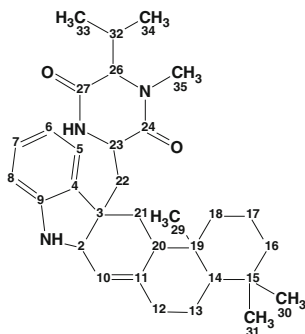


Thus, the structure of an unusual triterpene representing a new type of triterpenoid with a seven-membered ring was instantly ($t_g = 0.4$ s) and unambiguously elucidated in fully automatic mode in spite of the limited number of available 2D NMR correlations and the presence of significant number of ambiguous connectivities in the MCD.

4.3 Indotertine A

Hybrid isoprenoids are biosynthesized by attaching terpenoid moieties of different levels of complexity to molecules produced via nonterpenoid biosynthetic routes. Production of these compounds is quite rare among prokaryotes, however the reported literature for hybrid isoprenoids metabolites of bacterial origin have interesting structural diversity and biological activities. Che et al. [3] isolated a new

hybrid isoprenoid *indotertine A* from a reeds rhizosphere soil and elucidated its structure. Indotertine A (**4.5**) has a *novel scaffold* characterized by the condensed ring system containing a tryptophane's indole moiety and a sesquiterpene unit and represents a new class of natural products.



4.5

The molecular ion and 1D NMR spectra were input into the ACD/Structure Elucidator software to determine the molecular formula of the unknown. The molecular ion m/z 504.3583 $[M+H]^+$ distinguished in the HRMS spectrum suggests that the molecule contains an odd number of nitrogen atoms. According to the number of signals observed in the ^{13}C NMR spectrum (32), the possible number of carbon atoms was set to 32–34 in the options of the Molecular Formula Generator (it is taken into account that a couple of quaternary carbons may be missed). A range of 43–48 atoms was postulated for the number of hydrogen atoms from an analysis of ^1H NMR spectrum. To assess the possibility of heteroatoms being included into the molecular formula, a fragment search in the ACD Fragment Library by ^{13}C NMR spectrum was performed, and as a result ca. 7,000 fragments were found. Then, a Functional Group Library was “sifted” through the Found Fragments (see Sect. 1.3.3.6 and 3.2) and the following functional groups were displayed by the program as the most probable (Fig. 4.7):

Therefore, it can be expected that the unknown contains both amine and amide groups in agreement with revealing nitrogen as part of the molecular composition. At the same time there were no Found Fragments containing ketone, ester, or acid groups.

Judging from the probable functional groups selected, and the characteristic NMR chemical shift values 63.3, 3.88(CH), 67.3, 3.66(CH), 149.2(C), 165.8(C), 167.3(C),

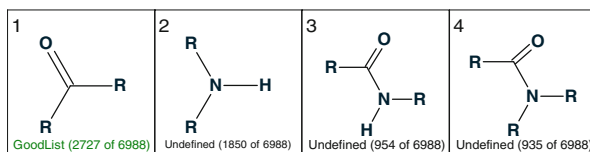


Fig. 4.7 Indotertine A: The most probable functional groups

the molecule can contain nitrogen and oxygen atoms bonded to carbon atoms; therefore the intervals of number of oxygen and nitrogen atoms were set as $n(\text{N}) = 1-10$, $n(\text{O}) = 1-10$ in the options of the Molecular Formula Generator. As a result two molecular formulae were offered by the program at a tolerance value of 0.5 a.m.u.:

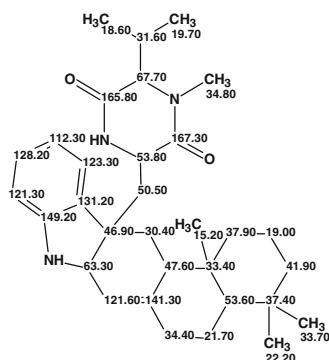
#	MF	m/z	Difference	RDBE
1	$\text{C}_{32}\text{H}_{45}\text{N}_3\text{O}_2$	503.3512	0.000703	12
2	$\text{C}_{33}\text{H}_{45}\text{N}_1\text{O}_3$	503.3399	-0.010531	12

It is evident that the molecular formula of the unknown is $\text{C}_{32}\text{H}_{45}\text{N}_3\text{O}_2$ which requires 12 degree of unsaturation.

2D NMR spectra of the unknown were measured by Che et al. but in the article [3] only the 1D ^{13}C and ^1H chemical shifts were given, while solely selected HMBC (28) and COSY (13) correlations were presented graphically on the elucidated structure. The NMR spectral information input into the program is presented in Table 4.3 and in the MCD (Fig. 4.8).

Overview of MCD The carbon atom 112.3(CH) colored in light blue is characterized with hybridization “ sp^3 or sp^2 ” to take into account the possibility to be included either into a C=C bond or into an acetal group. Ambiguous COSY and HMBC connectivities appeared due to the presence of two pairs of carbon atoms CH_3 19.00, CH_2 21.7 and CH_2 34.4, CH_2 50.5 with the overlapping signals of the attached hydrogens—1.46 and 1.89 ppm for the first and second pairs correspondingly. It is most probable that the chemical shifts 165.8(C) and 167.3(C) are accounted for as carbonyl groups, but labels “ob” were not ascribed to them for the sake of caution. No user MCD edits were performed.

MCD checking confirmed the absence of contradictions in the 2D NMR data and strict structure generation accompanied with ^{13}C chemical shift calculation and structure filtering was initiated. Structures characterized with average deviations $d > 4$ ppm were rejected in the process of structure generation. Results: $k = 290 \rightarrow 4 \rightarrow 1$, $t_g = 9$ s and a single structure **4.6** identical to structure **4.5** was found with the following set of ^{13}C deviations: $d_A = 1.92$, $d_N = 2.42$, $d_I = 2.39$.



4.6

Table 4.3 Indotertine A: NMR spectroscopic data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C2	63.3	62.86	CH	3.88	u	5.06	C9, C4
C3	46.9	48.33	C	–	–	–	–
C4	131.2	131.64	C	–	–	–	–
C5	123.3	123.5	CH	7.02	u	–	C3, C9
C6	121.3	119.93	CH	6.87	u	7.05	–
C7	128.2	128.16	CH	7.05	u	6.87, 6.68	–
C8	112.3	109.11	CH	6.68	u	7.05	C4
C9	149.2	148.79	C	–	–	–	–
C10	121.6	118.99	CH	5.06	u	3.88	C12
C11	141.3	142.31	C	–	–	–	–
C12	34.4	35.6	CH ₂	2.21	u	1.25	C10, C20, C14
C12	34.4	35.6	CH ₂	1.89	u	–	–
C13	21.7	25.25	CH ₂	1.25	u	2.21, 0.93	–
C13	21.7	25.25	CH ₂	1.46	u	1.40, 1.03	C19
C14	53.6	55.3	CH	0.93	u	1.25	–
C15	33.4	33.77	C	–	–	–	–
C16	41.9	42	CH ₂	1.14	u	–	–
C16	41.9	42	CH ₂	1.4	u	1.46	–
C17	19	19	CH ₂	1.46	u	–	–
C18	37.9	38.35	CH ₂	1.76	u	–	C14
C18	37.9	38.35	CH ₂	1.03	u	1.46	–
C19	37.4	40.27	C	–	–	–	–
C20	47.6	45.34	CH	1.7	u	2.13	C10, C11
C21	30.4	34.72	CH ₂	2.13	u	1.7	C2
C21	30.4	34.72	CH ₂	1.53	u	–	C4, C24, C2, C21
C22	50.5	38.87	CH ₂	2.33	u	3.22	–
C22	50.5	38.87	CH ₂	1.89	u	–	–
C23	53.8	51.78	CH	3.22	u	8.59, 2.33	–
C24	167.3	170.22	C	–	–	–	–
C26	67.7	68.13	CH	3.66	u	2.31	–
C27	165.8	167.6	C	–	–	–	–
C29	15.2	22.62	CH ₃	0.78	u	–	C20, C14
C30	33.7	21.18	CH ₃	0.82	u	–	C16, C14
C31	22.2	32.84	CH ₃	0.84	u	3.66, 1.04	C14, C16
C32	31.6	30.43	CH	2.31	u	1.16	–
C33	18.6	20.12	CH ₃	1.04	u	2.31	–
C34	19.7	18.38	CH ₃	1.16	u	2.31	–
C35	34.8	32.79	CH ₃	2.9	u	–	C24, C26
N1	100 ^a		NH	8.59	u	3.22	C26, C27

^a Fictitious ¹⁵N chemical shift

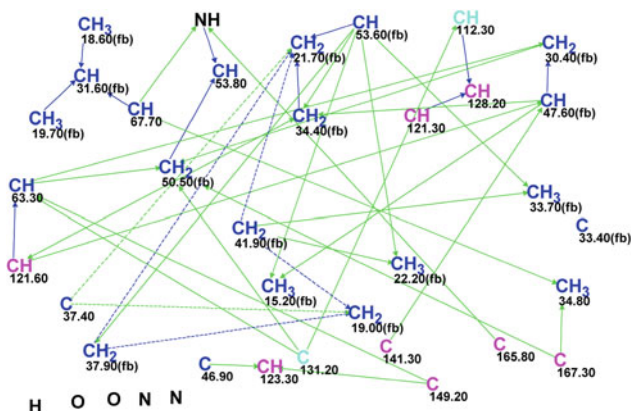
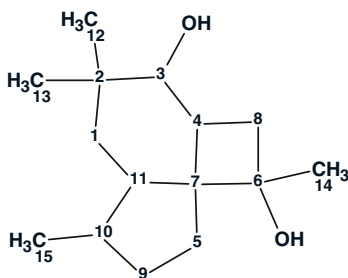


Fig. 4.8 Indotertine A: Molecular connectivity diagram tabl

In conclusion, a complex molecule of a new natural product characterized by a novel scaffold was automatically and unambiguously identified with the assistance of the Structure Elucidator software.

4.4 Trefolane A

Inspired by the potential biological activity of different sesquiterpenes from basidiomycetes, Ding et al. [4] obtained a *novel* sesquiterpenoid, Trefolane A (**4.7**), from cultures of the basidiomycete *Tremella foliacea*, an edible fungus with gelatinous fruiting bodies.



4.7

Trefolane A possesses an *unprecedented* 5/6/4 system which was suggested to be transformed from the humulene skeleton. Its structure was established by

extensive spectroscopic methods, and the absolute configuration was determined by single-crystal X-ray diffraction analysis.

Trefolane A, a colorless crystal, was detected to possess the molecular formula $C_{15}H_{26}O_2$ (HREIMS: $m/z = 238.1932 [M]^+$), indicating three degrees of unsaturation. The ^{13}C and DEPT NMR spectra revealed three sp^3 quaternary carbons (one oxygenated at δC 74.4), four sp^3 methines (one oxygenated at δC 74.5), four sp^3 methylenes, and four methyls. These data suggested that **4.7** was a sesquiterpenoid with a three-ring system.

1D NMR, HSQC data, and key HMBC and COSY correlations, which were presented in the article [4] graphically, are shown in Table 4.4.

The MCD is presented in Fig. 4.9.

MCD overview The overwhelming majority of carbon atoms were recognized by the program as possessing sp^3/fb property, and only three atoms C 55.7, C 74.4, and C 74.5 were not fully characterized by all properties. As there are two non-equivalent CH_2 groups having the same chemical shift 35.6 ppm, all HMBC and COSY correlations directed to these groups are ambiguous. The number of hydrogen atoms attached to the neighboring carbons were set in accordance with the 1H multiplicities displayed in the Table 4.4. No edits of the MCD were carried out. Since no contradictions were detected in the 2D NMR data, structure generation was performed with the following results: $k = 1,762 \rightarrow 19 \rightarrow 14$, $t_g = 1$ s.

The top structures of the ranked output file are shown in Fig. 4.10.

Table 4.4 Trefolane A: The spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	HMBC
C1	40.3	40.97	CH_2	1.6	u	1.83	C2
C1	40.3	40.97	CH_2	1.06	u	–	–
C2	35.7	37.49	C	–	–	–	–
C3	74.5	76.34	CH	3.43	d(6.6)	2.12	C2
C4	37.1	39.65	CH	2.12	u	3.43, 2.02	C7
C5	35.6	30.74	CH_2	2.02	u	2.12	C6
C6	74.4	75.26	C	–	–	–	–
C7	55.7	55.86	C	–	–	–	–
C8	35.6	38.16	CH_2	1.47	u	–	–
C8	35.6	38.16	CH_2	1.93	u	1.73	C7
C9	34.2	32.4	CH_2	0.88	u	–	–
C9	34.2	32.4	CH_2	1.73	u	1.93, 1.58	–
C10	41.8	39.44	CH	1.58	u	1.73, 1.83, 1.01	–
C11	39.9	44.18	CH	1.83	u	1.58, 1.60	C7
C12	23.4	27.13	CH_3	1.04	s	–	C2
C13	29.4	21.08	CH_3	0.91	s	–	C2
C14	23.2	25.51	CH_3	1.25	s	–	C6
C15	19.9	17.19	CH_3	1.01	d(6.8)	1.58	–

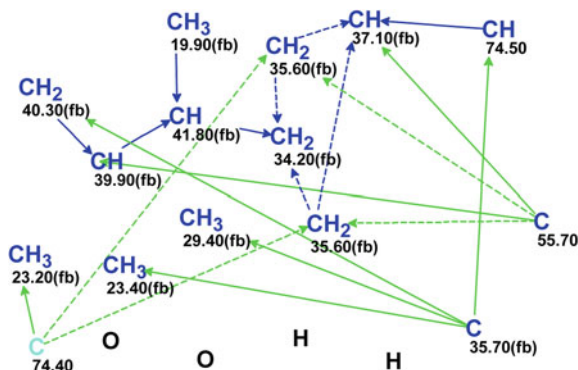
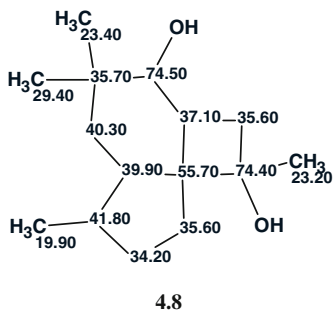


Fig. 4.9 Trefolane A: The MCD

$d_A(^{13}\text{C}): 2.239$ $d_N(^{13}\text{C}): 2.478$ $d_I(^{13}\text{C}): 2.318$ $\max_dA(^{13}\text{C}): 4.680$	$d_A(^{13}\text{C}): 3.460$ $d_N(^{13}\text{C}): 3.488$ $d_I(^{13}\text{C}): 2.963$ $\max_dA(^{13}\text{C}): 11.420$	$d_A(^{13}\text{C}): 3.517$ $d_N(^{13}\text{C}): 3.540$ $d_I(^{13}\text{C}): 3.436$ $\max_dA(^{13}\text{C}): 14.270$	$d_A(^{13}\text{C}): 3.583$ $d_N(^{13}\text{C}): 3.765$ $d_I(^{13}\text{C}): 3.488$ $\max_dA(^{13}\text{C}): 11.450$

Fig. 4.10 Trefolane A: The top structures of the ranked output file

Structure #1 was ranked first by all three ^{13}C chemical shift prediction methods and is identical to the published structure of Trefolane A. The difference between the deviations $\Delta = d(\#2) - d(\#1)$ allows us to conclude that the structure elucidation was performed reliably and unambiguously. The structure of Trefolane A (4.8) with the ^{13}C chemical shift assignments is shown below.



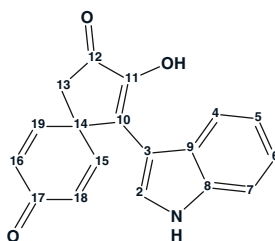
Note that this is the first isolated compound which represents a new type of carbon skeleton in the family of sesquiterpenoids and it was automatically identified by the program from key 2D NMR correlations and without the need for the application of any single-crystal X-ray diffraction experiment.

4.5 Spirobacillene A

One effective approach for new secondary metabolites is to study microorganisms derived from unexplored extreme environments (e.g., acidic mine drainage), where microorganisms can create unique offensive and defensive biochemical metabolism under ecological pressure.

Park et al. [5] isolated a bacterial strain, *Lysinibacillus fusiformis* KMC003 from acidic coal mine drainage that was highly contaminated by iron-rich heavy metal ions and sulfuric acid (pH 3.0). They investigated the production of secondary metabolites from the bacterial strain *L. fusiformis* KMC003 because the *Bacillus* species are known for their ability to produce structurally diverse bioactive molecules, such as polyene, macrolide, and especially peptide antibiotics.

The ethyl acetate extracts were subsequently subjected to reversed-phase HPLC separation to afford two novel compounds, Spirobacillenes A and B. Spirobacillenes A and B featured a *unique* indole and indolenine moiety that contained spirocyclopentenones, respectively. Here we will consider the structure elucidation of Spirobacillene A (**4.9**). The structure of **4.9** contains a single C–C bond between a spiro [4.5] decane moiety and an adjacent indole ring, which is a *novel skeleton* for a natural product.



4.9

Spirobacillene A was isolated as yellow crystals. The molecular formula was determined to be $C_{18}H_{13}NO_3$ ($[M+H]^+$ with m/z 292.0977 (calculated $m/z = 292.0974$) based on a HRFABMS measurement), which indicated that **4.9** contained 13 degrees of unsaturation. Note that the ratio of the total number of skeletal and hydrogen atoms is close to 2:1, which is an attribute of a challenging problem according to Crews rule (see Sect. 1.2.2). The IR spectrum showed

absorption bands that corresponded to hydroxyl ($3,228\text{ cm}^{-1}$) and conjugated carbonyl ($1,693\text{ cm}^{-1}$) functional groups. The absorption band at $1,657\text{ cm}^{-1}$ can be assigned either to another conjugated carbonyl group or to a carbon double bond.

The 1D NMR, HSQC, COSY, and HMBC data tabulated in article [5] are presented in Table 4.5.

The MCD created from the data presented in Table 4.6 is shown in Fig. 4.11.

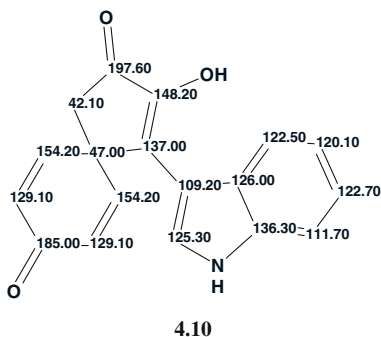
MCD overview The program automatically assigned hybridization states to all carbon atoms, while the label “ob” was assigned only to C 197.6. Ambiguous connectivities allocated to the carbon atoms at C 111.7 and C 125.3 are a consequence of the overlapping ^1H signals of the hydrogen atoms attached to the relevant carbon atoms. The number of hydrogen atoms that should be attached to the skeletal atoms existing in the first sphere around seven carbon atoms of the molecule were set to correspond with the ^1H signal multiplicities specified in Table 4.5. No edits were made to the MCD and strict structure generation accompanied by ^{13}C chemical shift prediction and spectral filtering was initiated.

Table 4.5 Spirobacillene A: The NMR spectroscopic data

Label	δX	$\delta\text{C}_{\text{calc}}$	XH_n	δH	M(J)	COSY	C HMBC
C2	125.3	129.91	CH	7.46	u	7.23, 9.63	C3, C9, C8
C3	109.2	116.53	C	–	–	–	–
C4	122.5	120.55	CH	7.98	d(8.0)	7.16	C3
C5	120.1	118.74	CH	7.16	dd(8.0,7.0)	7.23, 7.98	C9
C6	122.7	122.06	CH	7.23	dd(8.0,7.0)	7.46, 7.16	C8
C7	111.7	110.97	CH	7.46	u	–	–
C8	136.3	134.77	C	–	–	–	–
C9	126	127.51	C	–	–	–	–
C10	137	130.52	C	–	–	–	–
C11	148.2	148.94	C	–	–	–	–
C12	197.6	198.45	C	–	–	–	–
C13	42.1	41.7	CH_2	2.76	s	–	C12, C11, C10, C14
C14	47	48.99	C	–	–	–	–
C15	152.4	153.27	CH	7.06	d(10.0)	6.33	C10, C13, C17, C14
C16	129.1	129.58	CH	6.33	d(10.0)	–	–
C17	185	184.87	C	–	–	–	–
C18	129.1	129.58	CH	6.33	d(10.0)	7.06	C17, C14
C19	152.4	153.27	CH	7.06	d(10.0)	–	–
N1	100 ^a	–	NH	9.63	u	7.46	–

^a Fictitious ^{15}N chemical shift

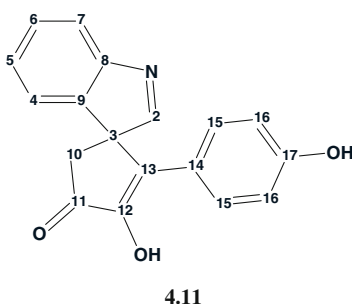
Results: $k = 698 \rightarrow 1$, $t_g = 3$ s, and the single structure **4.10** coincided with structure **4.9** of Spirobacillene A:



The deviation values ($d_A = 1.837$, $d_N = 1.676$, $d_I = 3.763$ ppm) confirm the correctness of the structure inferred by the program. Therefore, a challenging problem for the human expert was solved automatically, instantaneously, and unambiguously by Structure Elucidator.

4.6 Spirobacillene B

As mentioned above (Sect. 4.5) Spirobacillene B was isolated along with Spirobacillene A by Park et al. [5]. Spirobacillene B featured a *unique* indolenine moiety that contained spiro-cyclopentenone. Here we will consider the structure elucidation of Spirobacillene B (**4.11**), an isomer of Spirobacillene A (the molecular formula $C_{18}H_{12}NO_3$).



Although the carbon backbone of **4.11** has been previously reported as an intermediate in the synthesis of carbazole alkaloids, the fact that a highly functionalized spiro-cyclopentenone that contains this carbon backbone has a natural origin is intriguing.

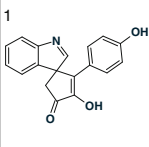
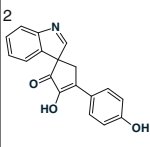
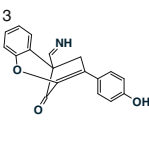
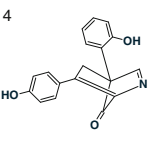
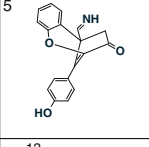
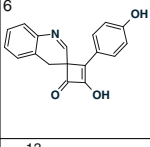
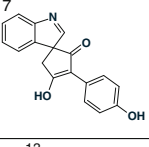
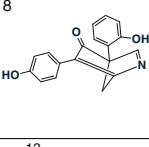
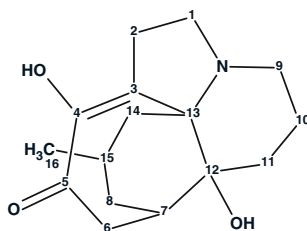
			
$d_A(^{13}\text{C}): 2.896$ $d_N(^{13}\text{C}): 2.749$ $d_I(^{13}\text{C}): 2.608$ $\text{max}_dA(^{13}\text{C}): 10.160$	$d_A(^{13}\text{C}): 3.404$ $d_N(^{13}\text{C}): 2.783$ $d_I(^{13}\text{C}): 3.256$ $\text{max}_dA(^{13}\text{C}): 13.040$	$d_A(^{13}\text{C}): 4.875$ $d_N(^{13}\text{C}): 4.405$ $d_I(^{13}\text{C}): 5.234$ $\text{max}_dA(^{13}\text{C}): 18.850$	$d_A(^{13}\text{C}): 4.929$ $d_N(^{13}\text{C}): 4.441$ $d_I(^{13}\text{C}): 5.053$ $\text{max}_dA(^{13}\text{C}): 18.850$
			
$d_A(^{13}\text{C}): 5.818$ $d_N(^{13}\text{C}): 4.894$ $d_I(^{13}\text{C}): 4.942$ $\text{max}_dA(^{13}\text{C}): 21.300$	$d_A(^{13}\text{C}): 6.684$ $d_N(^{13}\text{C}): 5.264$ $d_I(^{13}\text{C}): 5.735$ $\text{max}_dA(^{13}\text{C}): 27.950$	$d_A(^{13}\text{C}): 6.181$ $d_N(^{13}\text{C}): 5.598$ $d_I(^{13}\text{C}): 5.669$ $\text{max}_dA(^{13}\text{C}): 37.980$	$d_A(^{13}\text{C}): 5.813$ $d_N(^{13}\text{C}): 5.840$ $d_I(^{13}\text{C}): 5.673$ $\text{max}_dA(^{13}\text{C}): 29.830$

Fig. 4.13 Spirobacillene B: The first eight structures of the ranked output file

4.7 Lycojaponicum D

Plants of the *Lycopodium* species (Lycopodiaceae) are known to be a rich source of *Lycopodium* alkaloids possessing unique heterocyclic frameworks, such as huperzine A, fawcettimine, and serratinine, which have attracted great interest from biogenetic, synthetic, and biological perspectives. As a representative plant of the *Lycopodiaceae* family, *Lycopodium japonicum* Thunb. ex Murray has historically been used as a traditional Chinese medicine for the treatment of contusions, strains, and myasthenia. Previously, Wang et al. [6] reported three trace *Lycopodium* alkaloids, Lycojaponicumins A (see Sect. 4.22) and B with a 5/5/5/6 pentacyclic ring system and lycojaponicum C with a 6/5/5/6 tetracyclic ring system, isolated from this plant. In their next work [7], the same group of researchers described the discovery of a *structurally unique* alkaloid, Lycojaponicum D (**4.13**), with an *unprecedented* 5/7/6/6 tetracyclic skeleton formed by an unusual C3–C13 linkage.



4.13

Its structure was determined by extensive spectroscopic methods, CD experiments, ECD calculations, and X-ray crystallography. In this section, we describe the structure elucidation of **4.13** using Structure Elucidator.

Lycojaponicum D (**4.13**) was obtained as a white amorphous powder, and its molecular formula was established as $C_{16}H_{23}NO_3$ by HRESIMS at m/z 278.1747 $[M+H]^+$ (calcd 278.1751), accounting for six degrees of unsaturation. The IR spectrum of **4.13** (see SI to [7]) shows the absorption band of OH/NH groups at $3,390\text{ cm}^{-1}$ and a band at $1,677\text{ cm}^{-1}$ which can be related to either an α,β -unsaturated keto group or an amide group.

The 1D NMR and HSQC data tabulated in the article [7], as well as key HMBC and COSY correlations given by the authors [7], are represented in Table 4.7.

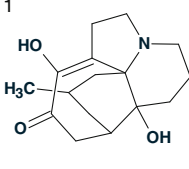
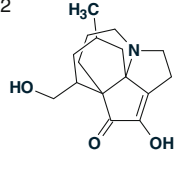
The MCD created on the basis of data collected in Table 4.7 is presented in Fig. 4.14.

In spite of the presence of three light blue carbon atoms (their hybridization is sp^2 or sp^3), no user intervention is necessary because the molecule is small ($C_{16}H_{23}NO_3$) and the number of 2D NMR correlations is large enough. No

Table 4.7 Lycojaponicum D: The NMR spectroscopic data

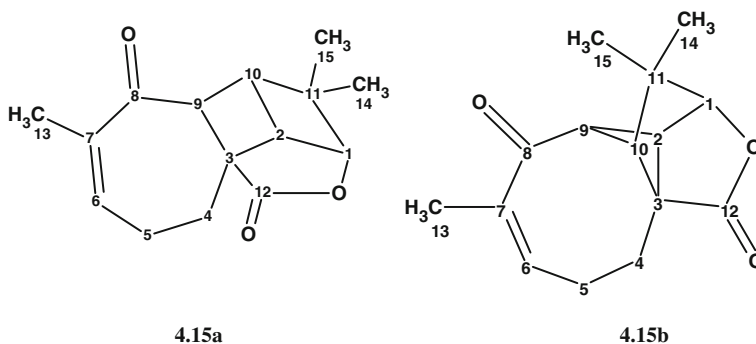
Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	HMBC
C1	49.3	49.84	CH ₂	2.76	u	–	–
C1	49.3	49.84	CH ₂	3.25	u	3.1	C13, C3, C9
C2	29.1	21.67	CH ₂	3.1	u	3.25	C4, C3, C13
C2	29.1	21.67	CH ₂	2.67	u	–	–
C3	133	136.1	C	–	–	–	–
C4	143.1	148.1	C	–	–	–	–
C5	195.8	197.1	C	–	–	–	–
C6	42.6	40.66	CH ₂	2.75	u	–	–
C6	42.6	40.66	CH ₂	2.83	u	2.09	–
C7	41.1	43.41	CH	2.09	u	2.83, 1.98	C12, C13, C5
C8	36	36.15	CH ₂	1.98	u	1.40, 2.09	–
C8	36	36.15	CH ₂	1.35	u	–	–
C9	51.1	47.78	CH ₂	2.86	u	2.16	–
C9	51.1	47.78	CH ₂	2.44	u	–	–
C10	20.4	21.15	CH ₂	2.16	u	2.86, 1.77	C12
C10	20.4	21.15	CH ₂	1.58	u	–	–
C11	32	34.2	CH ₂	1.77	u	2.16	C13
C11	32	34.2	CH ₂	1.5	u	–	–
C12	70.7	74.92	C	–	–	–	–
C13	73.8	72.32	C	–	–	–	–
C14	37.7	40.42	CH ₂	1.37	u	–	–
C14	37.7	40.42	CH ₂	1.74	u	1.4	C3
C15	25.1	27.53	CH	1.4	u	1.98, 1.74, 0.89	C13
C16	21.5	23.65	CH ₃	0.89	u	1.4	–

Fig. 4.15 Lycojaponicum
D: The ranked output
structure file

1	2
	
$d_A(^{13}\text{C}): 2.567$ $d_N(^{13}\text{C}): 3.029$ $d_I(^{13}\text{C}): 2.759$ $\text{max_}d_A(^{13}\text{C}): 7.430$	$d_A(^{13}\text{C}): 6.243$ $d_N(^{13}\text{C}): 6.407$ $d_I(^{13}\text{C}): 5.739$ $\text{max_}d_A(^{13}\text{C}): 20.460$

4.8 Aquatolide

Aquatolide is a humulane-derived sesquiterpenoid lactone isolated from *Asteriscus aquaticus*. The structure of Aquatolide (**4.15a**) originally proposed [8] on the basis of 1D and 2D NMR analysis contains an exceedingly rare ladderane substructure.



Intrigued by this structural unit, Lodewyk and coworkers [9] initiated quantum-mechanical (QM) NMR calculations to verify the reported connectivity. These calculations led the authors [9] to extensive structural revision of this complex natural product. As a result of the quantum-chemical calculations of the ^{13}C and ^1H chemical shifts and associated coupling constants for a series of possible structures (the GIAO approximation of the DFT approach was used), it was proven that the true structure of Aquatolide was **4.15b**. For this goal it was necessary to calculate the chemical shifts for 60 different possible alternative structures, largely based on other related compounds, and found in the same plant.

The article [9] does not contain either tabulated ^1H - ^{13}C HMBC data or a graphical representation of the two-dimensional correlations on the structures. Therefore, these data were directly extracted from the tabulated 1D NMR and HSQC data and the HMBC spectrum pattern presented in the Supporting Information to [9].

The NMR spectroscopic data of Aquatolide are presented in Table 4.8, and the corresponding MCD is displayed in Fig. 4.16.

No edits of the MCD were made. As no contradictions were detected in the 2D NMR data, strict structure generation was performed with the following results: $k = 3 \rightarrow 3$, $t_g = 0.05$ s. The three generated structures along with their ^{13}C and ^1H deviations as well as the maximum ^{13}C deviations are shown in Fig. 4.17.

The first observation that can be made is the absence of the original structure 4.15a in the output file. Therefore, if the CASE approach was utilized it would

Table 4.8 Aquatolide: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	84.2	87.69	CH	4.48	u	C12, C15, C3, C10, C14
C2	54.54	49.01	CH	3.26	u	C11, C8, C3, C10
C3	62.83	54.68	C	–	–	–
C4	22.15	31.27	CH_2	1.96	u	–
C4	22.15	31.27	CH_2	2.52	u	C2, C6, C10, C12, C3, C5
C5	28.63	23.36	CH_2	2.35	u	–
C5	28.63	23.36	CH_2	2.03	u	–
C6	131.1	143.34	CH	5.85	u	C4, C13, C8
C7	135.08	136.93	C	–	–	–
C8	211.94	199.27	C	–	–	–
C9	54.45	51.52	CH	2.92	u	C8, C7, C1, C10, C2, C3, C11
C10	62.59	49.53	CH	2.64	u	C11, C1, C15, C8, C2, C9, C14, C4
C11	41.86	40.09	C	–	–	–
C12	177.5	176.04	C	–	–	–
C13	22.22	19.56	CH_3	1.87	u	C7, C8, C6
C14	22.62	12.31	CH_3	1.05	u	C15, C1, C11, C10
C15	22.84	22.77	CH_3	1.19	u	C11, C14

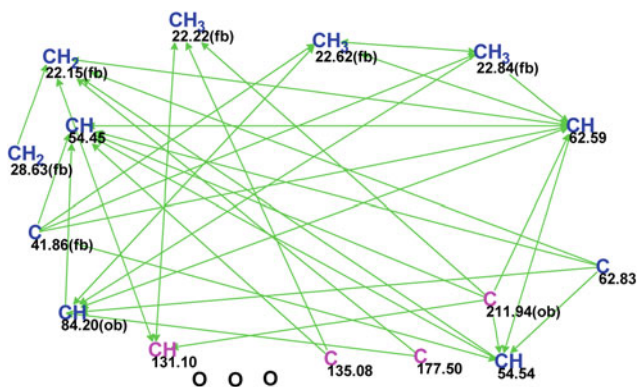


Fig. 4.16 Aquatolide: The MCD

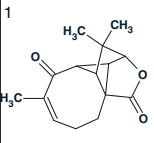
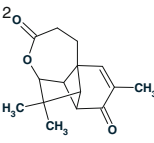
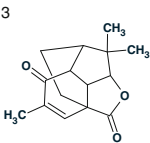
1 	2 	3 
$d_A(^{13}\text{C})$: 6.039 $d_N(^{13}\text{C})$: 5.811 $d_I(^{13}\text{C})$: 4.988 $\text{max_}d_A(^{13}\text{C})$: 13.060 $\text{max_}d_N(^{13}\text{C})$: 14.742 $\text{max_}d_I(^{13}\text{C})$: 11.097 $d_A(^1\text{H})$: 0.443 $d_N(^1\text{H})$: 0.308 $d_I(^1\text{H})$: 0.272	$d_A(^{13}\text{C})$: 6.235 $d_N(^{13}\text{C})$: 6.452 $d_I(^{13}\text{C})$: 6.355 $\text{max_}d_A(^{13}\text{C})$: 20.640 $\text{max_}d_N(^{13}\text{C})$: 19.388 $\text{max_}d_I(^{13}\text{C})$: 21.511 $d_A(^1\text{H})$: 0.379 $d_N(^1\text{H})$: 0.358 $d_I(^1\text{H})$: 0.342	$d_A(^{13}\text{C})$: 7.183 $d_N(^{13}\text{C})$: 6.146 $d_I(^{13}\text{C})$: 6.457 $\text{max_}d_A(^{13}\text{C})$: 15.820 $\text{max_}d_N(^{13}\text{C})$: 15.446 $\text{max_}d_I(^{13}\text{C})$: 14.131 $d_A(^1\text{H})$: 0.485 $d_N(^1\text{H})$: 0.425 $d_I(^1\text{H})$: 0.276

Fig. 4.17 Aquatolide: The ranked structural output file

prevent the generation of an erroneous structural hypothesis. The first ranked structure is identical to the revised structure **4.15b** though its deviation values are significantly higher than those usually obtained. This result is a consequence of the very unusual skeleton of **4.15b** for which there is a lack of associated structures in the prediction databases. Specifically, only single hits were found in the ACD/CNMR database to assist in the prediction of each of the chemical shifts at 131.1 and 211.94 ppm. Figure 4.18 shows the chemical shift distributions displayed in the calculation protocols which were generated to explain the predicted chemical shift values calculated for C 22.15 (5 hits) and C 62.83 (22 hits). The chemical shift distributions clearly demonstrate the reasons for the revealed discrepancies between the experimental and calculated chemical shifts.

The superiority of structure #1 over other structures is not so significant when judged using the average ^{13}C deviation values. Nevertheless, the maximum ^{13}C shift deviations and the average deviations calculated for the ^1H chemical shifts using incremental and neural nets approaches also support the priority ranking of the revised structure **4.15b**. In this case the quantum-chemical calculations could

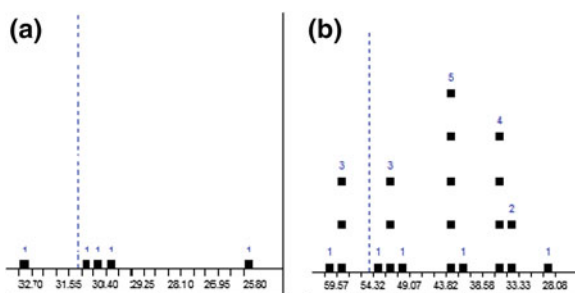


Fig. 4.18 Aquatolide: The calculation protocols which were generated to explain the predicted chemical shift values for C 22.15 (5 hits, **a**) and C 62.83 (22 hits, **b**). The vertical dashed lines indicate the positions of the predicted chemical shifts

play a decisive role in distinguishing the preferable structure among the three candidates illustrated in Fig. 4.17. Note that the quantum-chemical shift calculations would be performed only for the three most probable molecules, but not for the sixty as performed by the authors [9].

Figure 4.19 shows the revised and original structures along with the calculated deviation values.

Figure 4.19 shows that both the average and maximum ^{13}C chemical shift deviations allow the assignment of the revised structure as the most probable. ^1H chemical shift deviations cannot be used for supporting the priority of the revised structure in this case as $d_j\text{H}(\#1) < d_j\text{H}(\#2)$, $d_N\text{H}(\#1) \cong d_N\text{H}(\#2)$, and $d_A\text{H}(\#1) > d_A\text{H}(\#2)$. Such results are not unexpected because the accuracy of empirical ^1H chemical shift prediction is lower than the accuracy of ^{13}C chemical shift calculations and this is the reason why ^1H deviations play a secondary role in computer-assisted structure elucidation.

For completeness, it was interesting to investigate why the original structure **4.15a** was not generated by the program. Two possible reasons seem to be most probable: either structure **4.15a** was rejected by the filter or it was not produced by the generator due to the presence of contradictions between HMBC data and structure **4.15a** (presence of NSCs) that prevented its generation. The structure was checked using HMBC data, and as a result three $^4J_{\text{HC}}$ nonstandard correlations were detected as shown by the red arrows in structure **4.16**

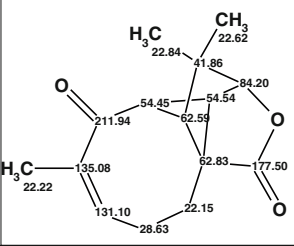
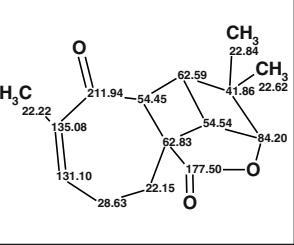
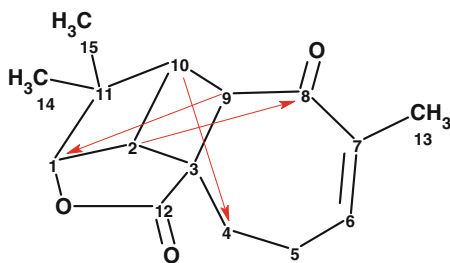
1 Revised	2 Original
 <p> $d_A(^{13}\text{C})$: 6.039 $d_I(^{13}\text{C})$: 4.988 $d_N(^{13}\text{C})$: 5.811 $\text{max}_dA(^{13}\text{C})$: 13.060 $\text{max}_dI(^{13}\text{C})$: 11.097 $\text{max}_dN(^{13}\text{C})$: 14.742 $d_A(^1\text{H})$: 0.443 $d_I(^1\text{H})$: 0.272 $d_N(^1\text{H})$: 0.308 </p>	 <p> $d_A(^{13}\text{C})$: 6.583 $d_I(^{13}\text{C})$: 5.367 $d_N(^{13}\text{C})$: 6.234 $\text{max}_dA(^{13}\text{C})$: 26.500 $\text{max}_dI(^{13}\text{C})$: 22.099 $\text{max}_dN(^{13}\text{C})$: 25.100 $d_A(^1\text{H})$: 0.214 $d_I(^1\text{H})$: 0.307 $d_N(^1\text{H})$: 0.293 </p>

Fig. 4.19 Aquatolide: The original and revised structures with the assigned ^{13}C chemical shifts and calculated ^{13}C and ^1H chemical shift deviations

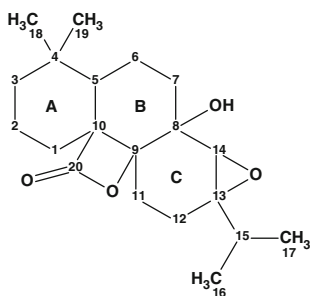


4.16

Therefore, to generate this structure it would be necessary to initiate Fuzzy Structure Generation (FSG) with the parameters $m = 3$, $a = 1$. The example convincingly shows that before performing laborious QM NMR chemical shift predictions for hypothetical structures it is worth employing a CASE system to generate a set of candidate structures and select the most probable ones for the QM calculations.

4.9 Rubesanolide A

Isodon rubescens (Hemsl.) Hara is a well-known folk medicine in China for the treatment of respiratory and gastrointestinal bacterial infections, inflammation, and cancer. Investigation of this plant led Zou et al. [10] to the isolation of two novel diterpenoids, Rubesanolides A (4.17) and B, with an *unprecedented* β -lactone group formed between C9 and C20.



4.17

It has been shown [10] that the compound 4.17 has a very different conformation from that of the abietane diterpenes, in which, all three rings (A, B, and C) are in chair conformations. However, in Rubesanolide A, the three six-membered rings form chair, boat, and twisted-chair conformations, respectively. The rings between A and

B and between B and C are *trans*-fused, which is *the first time* that a diterpene having such a conformation in the skeleton has been discovered.

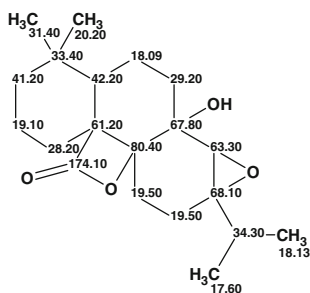
Rubesanolide A was isolated as colorless laminate crystals (MeOH). Its positive EIMS showed an $[M]^+$ at m/z 334, corresponding to a molecular formula of $C_{20}H_{30}O_4$ and requiring six units of unsaturation. The molecular formula was further confirmed by HRESIMS ($[M+Na]^+$, found 357.2035, calculated 357.2041), as well as by the ^{13}C and DEPT NMR spectra. Its IR spectrum showed absorption bands indicating the presence of hydroxyl ($3,451\text{ cm}^{-1}$) and carbonyl ($1,804\text{ cm}^{-1}$) groups. Note that the absorption band at $1,804\text{ cm}^{-1}$ is characteristic for strained four-membered lactones [11].

The 1D NMR and HSQC data tabulated in the article [10] as well as selected key HMBC and COSY correlations (shown in the article graphically) are presented in Table 4.9. The MCD is displayed in Fig. 4.20.

MCD overview It is worthy to note that if a ^{13}C chemical shift is observed in the region 60–70 ppm then the corresponding sp^3 -hybridized carbon atom is usually assigned as that whose neighbor is most probably an oxygen atom [11]. However in the molecule under consideration, such an assumption imposed upon all carbons whose chemical shifts fall into the interval between 61.2 and 68.1 ppm would lead to a pitfall because the atom C 61.2 is connected to carbon atoms only in the Rubesanolide A molecule, and hence the correct solution would never be obtained using this assumption. No user edits of the MCD were made and the signal multiplicities observed in the 1H NMR spectrum (column M(J), Table 4.9) were not used to accelerate the structure generation.

MCD checking did not detect any contradictions in the 2D NMR data. Strict structure generation accompanied by ^{13}C chemical shift prediction and structure filtering gave the following results: $k = 60,991 \rightarrow 4 \rightarrow 3$, $t_g = 1$ min 24 s. The ranked output file is presented in Fig. 4.21.

We see that structure #1 is identical to the originally elucidated structure 4.17 and its automatically performed ^{13}C chemical shift assignment is shown on structure 4.18



4.18

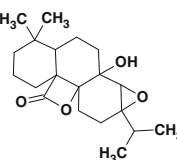
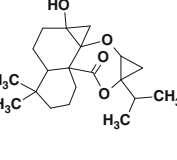
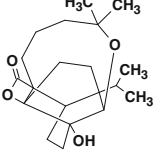
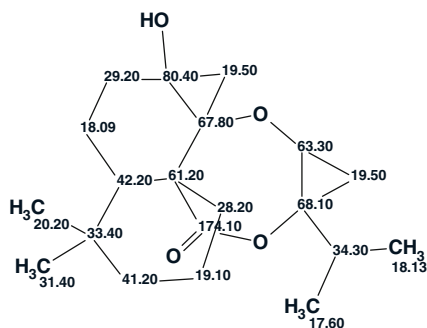
<p>1</p> 	<p>2</p> 	<p>3</p> 
<p>$d_A(^{13}\text{C})$: 2.645 $d_N(^{13}\text{C})$: 3.237 $d_I(^{13}\text{C})$: 3.280 $\text{max}_{d_A}(^{13}\text{C})$: 5.340</p>	<p>$d_A(^{13}\text{C})$: 3.188 $d_N(^{13}\text{C})$: 3.686 $d_I(^{13}\text{C})$: 3.399 $\text{max}_{d_A}(^{13}\text{C})$: 14.730</p>	<p>$d_A(^{13}\text{C})$: 4.398 $d_N(^{13}\text{C})$: 4.532 $d_I(^{13}\text{C})$: 3.935 $\text{max}_{d_A}(^{13}\text{C})$: 12.730</p>

Fig. 4.21 Rubesanolide A: Ranked output file

It was interesting to see how utilization of multiplicities in ^1H NMR spectrum will influence the solution of the problem. When the numbers of neighboring hydrogen atoms were added to the atom properties on the MCD (column M(J), Table 4.9) then the following solution was obtained: $k = 5,366 \rightarrow 3 \rightarrow 2$, $t_g = 15$ s, i.e., the time to perform structure generation became five times shorter. The structures are presented in Fig. 4.22.

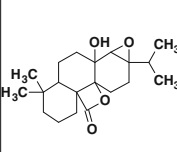
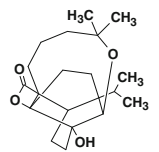
A comparison of Figs. 4.21 and 4.22 indicates that structure #2 displayed in Fig. 4.21 was rejected during the second run of the structure generation. The reason for the elimination of structure #2 can be explained by the analysis of its ^{13}C chemical shift assignment:



Structure #2 (Fig. 4.2)

It is evident that the multiplicity of the ^1H signal associated with the resonance of proton δH 2.73 (singlet) attached to C 63.3 contradicts the predicted multiplicity (triplet) of this signal in structure #2. The example shows that involving the structural information carried by the multiplicities of five (from twenty two) signals in the ^1H spectrum reduces the time for structure generation drastically.

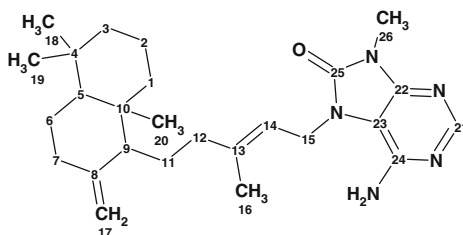
Fig. 4.22 The Rubesanolide
A: Ranked output file
obtained as a result of the
second run

1	2
	
$d_A(^{13}\text{C}): 2.645$	$d_A(^{13}\text{C}): 4.398$
$d_N(^{13}\text{C}): 3.237$	$d_N(^{13}\text{C}): 4.532$
$d_I(^{13}\text{C}): 3.280$	$d_I(^{13}\text{C}): 3.935$
$\text{max_}d_A(^{13}\text{C}): 5.340$	$\text{max_}d_A(^{13}\text{C}): 12.730$

4.10 Oxo-agelasine D

Marine sponges of the genus *Agelas* have proven to be an excellent source of *structurally novel* natural products, ranging from diterpene alkaloids to bromopyrrole alkaloids and glycosphingolipids. Diterpene alkaloids and their analogs have attracted a great deal of attention for their wide range of biological activities such as antimicrobial, antimalarial, antileukemic, cytotoxic, and antifouling activities.

As a result of studies on the marine sponge *Agelas mauritiana*, Yang and coworkers [12] isolated several new alkaloids and determined their structures. We will reexamine the experimental data presented in the article [12] which allowed the authors to elucidate the structure of a new compound Oxo-agelasine D (**4.19**).



4.19

Compound **4.19** was obtained as a white, amorphous solid. The similarity of the UV absorption pattern (λ_{max} 220, 269 nm, MeOH) to those of agelasines suggested that compound **4.19** was a related metabolite. The molecular formula $\text{C}_{26}\text{H}_{39}\text{N}_5\text{O}$ was deduced from the HRESIMS, ^{13}C NMR, and HSQC data. The ^1H NMR, ^{13}C NMR, and 2D NMR data of compound **4.19** are shown in Table 4.10. The MCD (Fig. 4.23) displays the spectrum-structural information graphically.

No edits were made to the MCD. As no contradictions were determined in the 2D NMR data strict structure generation combined with ^{13}C chemical shift prediction was initiated which was completed with the following results: $k = 12 \rightarrow 3 \rightarrow 3$, $t_g = 0.1$ s. The ranked output structural file is presented in Fig. 4.24.

Table 4.10 Oxo-agelazine D: Spectroscopic NMR data

Label	δ_X	C_{calc}	XH_n	δH	M (J)	C HMBC
C1	39.1	37.25	CH ₂	1.71	u	C8, C20, C9, C5, C7, C10, C2
C1	39.1	37.25	CH ₂	0.96	u	–
C2	19.3	26.33	CH ₂	1.44	u	–
C2	19.3	26.33	CH ₂	1.54	u	–
C3	42.1	46.59	CH ₂	1.38	u	C18, C5, C4, C19
C3	42.1	46.59	CH ₂	1.15	u	–
C4	33.5	36.57	C	–	–	–
C5	55.5	56.61	CH	1.04	u	C10, C4, C9, C7, C20, C6, C19
C6	24.4	24.18	CH ₂	1.71	u	–
C6	24.4	24.18	CH ₂	1.3	u	–
C7	38.3	38.84	CH ₂	1.9	u	–
C7	38.3	38.84	CH ₂	2.36	u	C17, C6, C8, C5, C9
C8	148.4	148.9	C	–	–	–
C9	56.2	42.92	CH	1.54	u	C10, C1, C8, C7, C5, C20, C17, C12
C10	39.6	39.24	C	–	–	–
C11	21.6	21.58	CH ₂	1.6	u	C12, C9
C11	21.6	21.58	CH ₂	1.43	u	–
C12	38.3	33.6	CH ₂	2.2	u	C11, C16
C12	38.3	33.6	CH ₂	1.88	u	–
C13	141.9	137.63	C	–	–	–
C14	120.8	122.74	CH	5.32	u	C16, C15, C12
C15	40.2	42.54	CH ₂	4.65	u	C13, C25, C23, C14
C15	40.2	42.54	CH ₂	4.61	u	–
C16	16.8	25.27	CH ₃	1.81	u	C13, C14, C12
C17	106.2	106.63	CH ₂	4.45	u	–
C17	106.2	106.63	CH ₂	4.8	u	C9, C8, C7
C18	33.5	30.99	CH ₃	0.87	u	C4, C5, C3, C19
C19	21.7	24.23	CH ₃	0.79	u	C3, C5, C4
C20	14.4	22.81	CH ₃	0.66	u	C10, C5, C9
C21	151.3	152.42	CH	8.18	u	C22, C24
C22	148.5	159.9	C	–	–	–
C23	106.1	104.23	C	–	–	–
C24	146	147.79	C	–	–	–
C25	153	156.49	C	–	–	–
C26	26.4	30.55	CH ₃	3.45	u	C25, C22
N1	100 ^a	–	NH ₂	5.07	u	C23

^a Fictitious ¹⁵N chemical shift

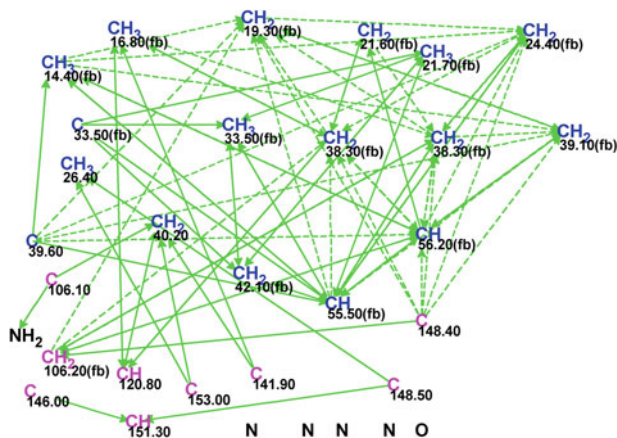
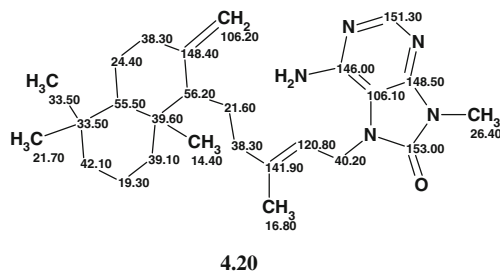


Fig. 4.23 Oxo-agelasine D: Molecular connectivity diagram

1	2	3
$d_A(^{13}\text{C}): 1.719$ $d_N(^{13}\text{C}): 2.016$ $d_I(^{13}\text{C}): 2.313$ $\text{max}_dA(^{13}\text{C}): 8.580$	$d_A(^{13}\text{C}): 3.534$ $d_N(^{13}\text{C}): 4.812$ $d_I(^{13}\text{C}): 3.684$ $\text{max}_dA(^{13}\text{C}): 13.280$	$d_A(^{13}\text{C}): 4.010$ $d_N(^{13}\text{C}): 4.662$ $d_I(^{13}\text{C}): 3.610$ $\text{max}_dA(^{13}\text{C}): 12.230$

Fig. 4.24 Oxo-agelasine D: Ranked output structural file

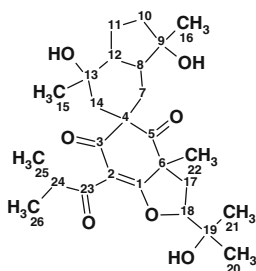
Structure #1 which coincides with the structure of Oxo-agelasine D is unambiguously selected as the best one, and its ^{13}C chemical shift assignment is shown on structure 4.20.



The structure of a new natural product was therefore identified almost instantaneously and fully automatically with the aid of Structure Elucidator.

4.11 Chipericumun A

The plants of the genus *Hypericum* (family Clusiaceae) have been used as traditional remedies in several parts of the world. These plants are known to contain various types of compounds such as naphthodianthrones, xanthenes, flavonoids, and prenylated acylphloroglucinols. Among them, prenylated acylphloroglucinols have attracted much scientific interest because of their *fascinating chemical structures* and intriguing biological activities. Abe et al. [13] undertook a search for structurally interesting compounds from *Hypericum* sp. which resulted in the isolation of four new prenylated acylphloroglucinols, Chipericumins A–D. In this section, we will describe the computer-assisted structure elucidation of Chipericumun A (**4.21**).



4.21

Chipericumun A was obtained as an optically active colorless amorphous solid. The molecular formula of **4.21**, $C_{26}H_{38}O_7$, was established by HRESIMS (m/z 485.2503 $[M+Na]^+$, $\Delta - 0.7$ mmu). The IR spectrum revealed the presence of hydroxyl ($3,422\text{ cm}^{-1}$) and carbonyl ($1,714$ and $1,693\text{ cm}^{-1}$) groups, as well as the supposed presence of conjugated double bonds ($1,618\text{ cm}^{-1}$). The ^{13}C , 1H , and HSQC spectroscopic data, as well as *selected* key COSY and HMBC correlations available graphically from [13] are presented in Table 4.11.

The MCD created from the data included in Table 4.11 is presented in Fig. 4.25.

MCD overview The light blue atoms C 70.40–C 91.90 are most probably oxygenated sp^3 carbons, while hybridization of carbon C 112.5 can be related both to sp^2 and to sp^3 (acetal). In order to be cautious no edits were made for the initial MCD which survived checking for the presence of contradictions in the 2D NMR

Table 4.11 Chipericumun A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	178.5	186.54	C	–	–	–	–
C2	112.5	115.67	C	–	–	–	–
C3	197.9	195.29	C	–	–	–	–
C4	60.2	65.57	C	–	–	–	–
C5	208.2	206.75	C	–	–	–	–
C6	58	58.04	C	–	–	–	–
C7	31.5	28.86	CH ₂	1.63	u	1.88	C14
C7	31.5	28.86	CH ₂	1.71	u	–	–
C8	48.7	47.22	CH	1.88	u	1.91, 1.63	–
C9	79.1	80.97	C	–	–	–	–
C10	40.2	40.74	CH ₂	1.79	u	1.86	–
C11	21.9	24.35	CH ₂	1.86	u	1.79, 1.91	–
C11	21.9	24.35	CH ₂	1.45	u	–	–
C12	51.7	52.02	CH	1.91	u	1.86, 1.88	–
C13	73	71.69	C	–	–	–	–
C14	46.9	47.13	CH ₂	2.1	u	–	–
C14	46.9	47.13	CH ₂	2.2	u	–	C5, C4, C3
C15	21.7	27.75	CH ₃	1.1	s	3.19	C14, C12, C13
C16	26.6	25.72	CH ₃	1.27	s	–	C9, C10, C8
C17	32.1	35.02	CH ₂	1.95	u	–	–
C17	32.1	35.02	CH ₂	2.47	u	4.61	–
C18	91.9	90.92	CH	4.61	u	2.47	–
C19	70.4	71.17	C	–	–	–	–
C20	24.2	26.99	CH ₃	1.16	s	–	C21, C18, C19
C21	26.9	24.35	CH ₃	1.35	s	–	–
C22	26	14.78	CH ₃	1.62	s	–	C1, C5, C17, C6
C23	204	205.05	C	–	–	–	–
C24	39.9	35.41	CH	3.19	u	1.10, 1.05	–
C25	18.2	19.02	CH ₃	1.05	d(6.8)	3.19	C23
C26	18.1	19.02	CH ₃	1.1	d(6.8)	–	–

data. Strict structure generation accompanied by ¹³C chemical shift prediction and structural filtering gave the following results: $k = 6,614 \rightarrow 26 \rightarrow 14$, $t_g = 35$ s. The first three top structures of the ranked output file are presented in Fig. 4.26.

The structure selected as the best one coincided with the structure of Chipericumun A determined by the authors [13] and its chemical shift assignments are presented on structure 4.22:

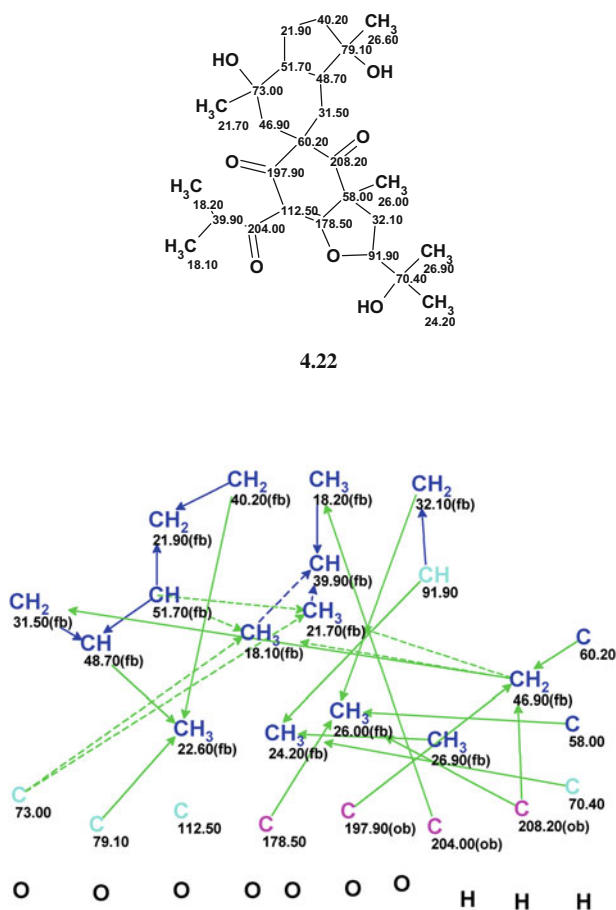


Fig. 4.25 Chipericum A: Molecular connectivity diagram

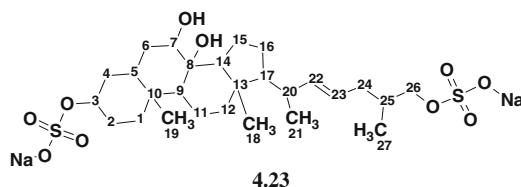
1	2	3
$d_A(^{13}\text{C}): 2.662$ $d_N(^{13}\text{C}): 2.497$ $d_I(^{13}\text{C}): 2.289$ $\text{max}_dA(^{13}\text{C}): 11.220$	$d_A(^{13}\text{C}): 3.512$ $d_N(^{13}\text{C}): 3.615$ $d_I(^{13}\text{C}): 3.503$ $\text{max}_dA(^{13}\text{C}): 15.700$	$d_A(^{13}\text{C}): 3.552$ $d_N(^{13}\text{C}): 3.581$ $d_I(^{13}\text{C}): 3.396$ $\text{max}_dA(^{13}\text{C}): 18.910$

Fig. 4.26 Chipericum A: Top structures of the ranked output file

The complex structure of this natural product was therefore automatically elucidated with the aid of Structure Elucidator from *selected* key COSY and HMBC correlations in several seconds. The difference between the ^{13}C chemical shift deviations calculated for the first and second structures shown in Fig. 4.26 is *ca.* 1 ppm, which confirms the correctness of the solution found for the problem.

4.12 Ascidia SAAF

Sperm activation and chemotaxis that ubiquitously occur in animal species, including humans, play an important role in fertilization. Some examples of a single agent simultaneously inducing sperm activation and attraction were identified. To understand the molecular mechanism underlying the genus specificity of sperm chemotaxis of ascidians, Matsumori et al. [14] have investigated the structure of a novel compound Ascidia-SAAF 2 (**4.23**) isolated from the eggs of the ascidian *Ascidia sydneiensis*.



Ascidia-SAAF 2 was isolated in the amount of 4 nmol (2.6 μg) from the active HPLC fractions. The chemical structure of Ascidia-SAAF 2 was elucidated by performing NMR and MS spectroscopy. The molecular formula of Ascidia-SAAF 2, $\text{C}_{27}\text{H}_{44}\text{O}_{10}\text{S}_2\text{Na}_2$, was obtained from the negative-ion high-resolution MS (m/z 296.1188, $[\text{M}-2\text{Na}]^{2-}$, calculated m/z 296.1193), indicating a dehydrogenated or oxygenated form of the known compound Ciona-SAAF ($\text{C}_{27}\text{H}_{46}\text{O}_{10}\text{S}_2\text{Na}_2$).

The gDQF-COSY and gHMBC spectra obtained using the cold probe technology unambiguously demonstrated the presence of a double bond between C22 and C23. Although the chemical shifts of both H22 and H23 almost overlapped to provide second-order signals, its E-configuration was identified on the basis of a large $^3J_{\text{HH}}$ (17 Hz) coupling between H22 and H23, deduced from a spectral simulation of the signals. The absence of an ROE response between H21 and H24 could lend support for the presence of this configuration.

The spectroscopic NMR data (^1H , ^{13}C , HSQC, COSY, and HMBC) used for the molecular structure elucidation are presented in Table 4.12.

The MCD from which structure generation was initiated is presented in Fig. 4.27.

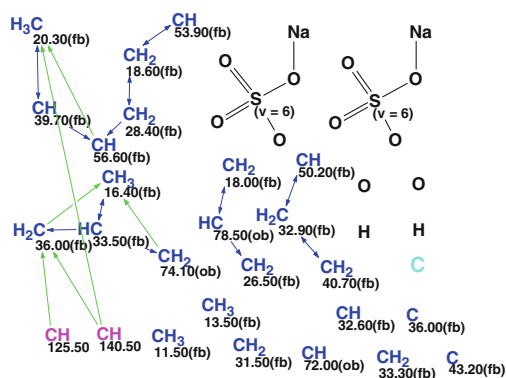
Table 4.12 Ascidia-SAAF 2. Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	33.3	37.13	CH ₂	1.18	u	1.90, 3.56	–
C1	33.3	37.13	CH ₂	<u>1.55</u>	u	1.84	–
C2	26.5	26.54	CH ₂	1.8	u	–	–
C2	26.5	26.54	CH ₂	1.84	u	1.55, 4.65	–
C3	78.5	81.9	CH	4.65	u	1.60, 1.84	–
C4	18.0	34.16	CH ₂	1.6	u	1.90, 4.65	–
C4	18.0	34.16	CH ₂	<u>1.55</u>	u	–	–
C5	32.6	39.37	CH	<u>1.9</u>	u	1.18, 1.60	–
C6	31.5	32.04	CH ₂	<u>1.9</u>	u	–	–
C6	31.5	32.04	CH ₂	1.18	u	–	–
C7	72.0	72.4	CH	3.56	u	1.18	–
C8	N.d.	76.14	C	–	–	–	–
C9	50.2	51.14	CH	1.25	u	1.59	–
C10	36	37.68	C	–	–	–	–
C11	32.9	20.04	CH ₂	1.59	u	1.21, 1.25	–
C12	40.7	37.06	CH ₂	1.21	u	1.59	–
C12	40.7	37.06	CH ₂	1.99	u	–	–
C13	43.2	42.42	C	–	–	–	–
C14	53.9	54.35	CH	1.54	u	1.38	–
C15	18.6	24.3	CH ₂	1.47	u	–	–
C15	18.6	24.3	CH ₂	1.38	u	1.28, 1.54	–
C16	28.4	28.43	CH ₂	1.28	u	1.14, 1.38	–
C16	28.4	28.43	CH ₂	1.73	u	–	–
C17	56.6	55.04	CH	1.14	u	2.07, 1.28	
C18	13.5	20.57	CH ₃	<u>0.9</u>	u	–	C13, C12, C9, C17, C14, C1, C10, C5
C19	11.5	21.58	CH ₃	0.9	u	–	–
C20	39.7	32.67	CH	2.07	u	0.98, 1.14, 5.39	
C21	20.3	20.85	CH ₃	0.98	u	2.07	C22, C20, C17
C22	140.5	140.59	CH	5.39	u	2.02, 2.07	–

(continued)

Table 4.12 (continued)

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C23	125.5	128.38	CH	5.39	u	–	–
C24	36.0	33.32	CH ₂	1.93	u	–	C22, C23
C24	36.0	33.32	CH ₂	2.02	u	5.39, 1.87	–
C25	33.5	34.07	CH	1.87	u	0.92, 2.02, 3.83	–
C26	74.1	76.27	CH ₂	3.83	u	1.87	–
C26	74.1	76.27	CH ₂	3.95	u	–	–
C27	16.4	17.41	CH ₃	0.92	u	1.87	C26, C24, C25

**Fig. 4.27** Ascidia-SAAF 2: Molecular connectivity diagram

MCD overview Two SO_4Na fragments were drawn by hand as their presence was postulated by the authors [14] from previous chemical knowledge and confirmed by MS analysis. The signal from one carbon atom was not observed in the ^{13}C NMR spectrum, so it was introduced in Table 4.12 without any chemical shift (C8). This atom is most probably a quaternary carbon and it was marked manually as “ sp^2 or sp^3 ”-hybridized on the MCD (light blue colored). One can see from Table 4.12 (italic underlined) that there are three pairs of overlapping 1H signals. Nine carbon atoms are not involved in the formation of the COSY and HMBC connectivities, but at the same time all of them are supplied with the labels “fb” defining their admissible neighbors (prohibition of neighboring with heteroatoms), which accelerates structure generation significantly and reduces the size of the output structural file.

No contradictions were detected in the 2D NMR data and strict structure generation combined with ^{13}C chemical shift prediction and utilizing an average deviation

threshold ($d = 4$ ppm) was initiated. The results were: $k = 2,880 \rightarrow 20 \rightarrow 8$, $t_g = 31.5$ s. The two top structures of the ranked output file are presented in Fig. 4.28.

Figure 4.28 shows that structure #1 coincides with the structure of Ascidia-SAAF 2, though the difference $\Delta = d(2) - d(1)$ is small and the values of the deviations are large enough. The reason can be explained by comparing the experimental ^{13}C chemical shifts with the predicted ones. Table 4.12 shows that the largest discrepancy between the experimental and calculated values are observed for atoms C4 and C11. The Chemical Shift Calculation protocols generated for these atoms are presented in Figs. 4.29 and 4.30 correspondingly.

Only one hit was found in the first case and two hits—in the second, but these results suggest that the chemical shifts 18.00 and 32.90 ppm could be exchanged. When the exchange was carried out the average deviations reduced to the following values: $d_A = 2.28$, $d_I = 2.864$, and $d_N = 2.17$ ppm. We believe that this result can be considered as justification of the atom permutation.

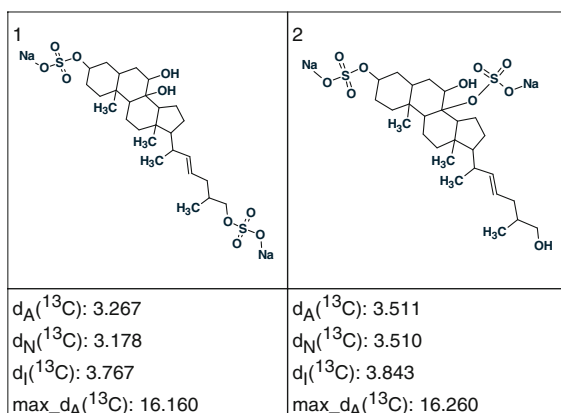


Fig. 4.28 Ascidia-SAAF 2: Two top structures of the ranked output file

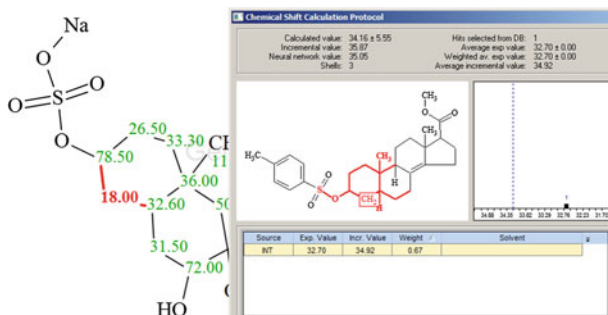


Fig. 4.29 Ascidia-SAAF 2: Chemical Shift Calculation Protocol for the carbon atom C4 (δ 18.0)

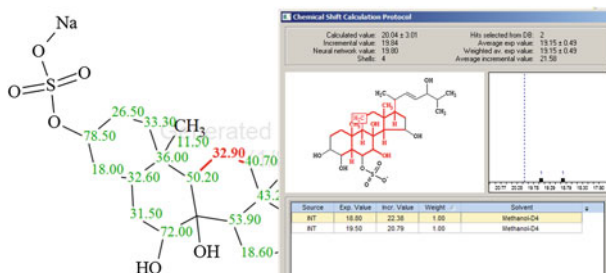
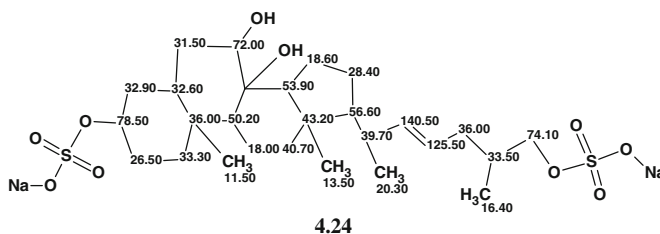


Fig. 4.30 Ascidia-SAAF 2: Chemical Shift Calculation Protocol for the carbon atom C11 (δ 32.9)

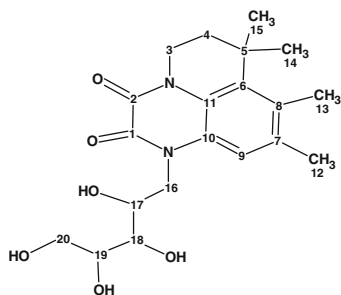
As established by the authors [14] on the basis of a large $^3J_{\text{HH}}$ (17 Hz) between H22 and H23 (see above), the single double bond C=C exists in structure **4.23** in the E-configuration. However, ^{13}C chemical shift prediction using ACD Predictor can also help in distinguishing between E- and Z-configurations. When structure #1 was converted to the Z-configuration by moving substituents, the deviations increased to $d_A = 2.584$, $d_I = 3.08$, and $d_N = 2.42$ ppm. Therefore, utilizing Structure Elucidator allowed for the determination of the correct structure of the unknown and its configuration around a double bond.

The elucidated structure **4.24** with the refined chemical shift assignment is shown below.



4.13 Hunanamycin A

Hunanamycin A (**4.25**), the *first natural product* with a pyrido[1,2,3-de]quinoxaline-2,3-dione core, was isolated by Hu et al. [15] from a marine-derived *Bacillus hunanensis*.



4.25

Hunanamycin A is related to a degradation product of riboflavin but has undergone an N-prenylation and subsequent cyclization. The structure, including stereochemistry, was determined by NMR and MS methods.

Hunanamycin A was isolated as a light yellow glass; high-resolution ESI-MS (HRESIMS) analysis of **4.25** gave an $[M+H]^+$ ion at m/z 393.2018 consistent with a molecular formula of $C_{20}H_{28}N_2O_6$ (calculated for $C_{20}H_{28}N_2O_6$, 393.2025) and eight degrees of unsaturation.

To elucidate the structure of Hunanamycin A, ^{13}C , 1H , HSQC, HMBC, and COSY, NMR spectra were used (Table 4.13).

Slightly edited MCD is presented in Fig. 4.31.

MCD overview The molecule contains six oxygen and two nitrogen atoms as well as four freely exchangeable hydrogens. This means that all possible combinations of OH, NH, and NH_2 groups will need to be taken into account during the structure generation process and this will lead to an increase in the size of the output file. To reduce the dimension of the problem without introducing risky assumptions (“axioms”), two light blue atoms (sp^2 or sp^3) C 123.3 and C 126.2 were marked as sp^2 -hybridized, while the carbon atoms of the methine groups 70.60, 74.40, and 75.00 for which the 1H chemical shift values of the attached hydrogens vary between 3.75 and 4.25 ppm, were marked as having a heteroatom as a neighbor (“ob” label). No nonstandard connectivities were detected, therefore strict structure generation combined with ^{13}C chemical shift prediction and rejection of structures for which the average deviations exceeded 4 ppm was initiated. The results gave $k = 10,281 \rightarrow 2 \rightarrow 1$, $t_g = 18$ s.

The single structure supplied with the average deviations and associated ^{13}C chemical shift assignment is presented in Fig. 4.32.

The resultant structure when compared with structure **4.25** shows that the structure of the new antibiotic Hunanamycin A was correctly and almost automatically identified using minimum assumptions.

For completeness of the investigation structure generation was repeated with the hybridization *not* sp which was automatically assigned to carbon atoms C 123.3 and C 126.2 by the program. The results: $k = 175,621 \rightarrow 2 \rightarrow 1$, $t_g = 5$ min. Comparison of both the solutions obtained shows visually the evident role of constraints imposed by the chemist.

Table 4.13 Hunanamycin A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M (J)	COSY	C HMBC
C1	156.7	155.27	C	–	–	–	–
C2	155.3	154.73	C	–	–	–	–
C3	39.3	40.74	CH ₂	4.12	u	–	C4, C11, C2, C5
C4	40.8	34.56	CH ₂	1.94	u	–	C6, C3, C14, C15, C5
C5	34.5	29.68	C	–	–	–	–
C6	134.3	134.24	C	–	–	–	–
C7	133.8	133.26	C	–	–	–	–
C8	136	127.44	C	–	–	–	–
C9	117.2	120.1	CH	7.47	u	–	C12, C11, C7, C10, C8
C10	126.2	128.99	C	–	–	–	–
C11	123.3	125.12	C	–	–	–	–
C12	21.6	19.69	CH ₃	2.33	u	–	C7, C9, C8
C13	19.6	17.31	CH ₃	2.45	u	–	C8, C6, C7
C14	29.3	30.91	CH ₃	1.555	u	–	C6, C5, C4, C15
C15	29	30.83	CH ₃	1.54	u	–	C6, C14, C5, C4
C16	46.6	45.68	CH ₂	4.79	u	4.25	C1, C10, C18, C17
C16	46.6	45.68	CH ₂	4.28	u	–	–
C17	70.6	70.05	CH	4.25	u	4.79, 3.76	C16
C18	75	73.98	CH	3.76	u	3.77, 4.25	C19, C16
C19	74.4	72.56	CH	3.77	u	3.81, 3.76	C20, C17, C18
C20	65	63.36	CH ₂	3.81	u	3.77	–
C20	65	63.36	CH ₂	3.67	u	–	C18, C19

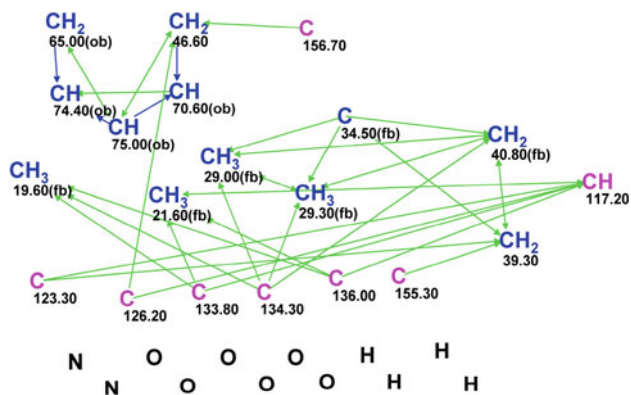
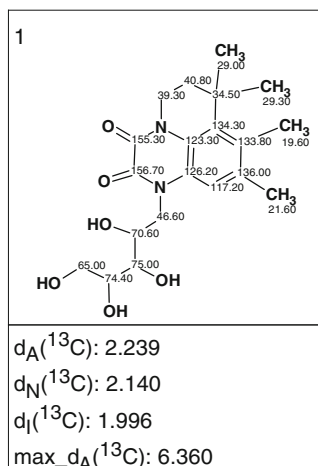
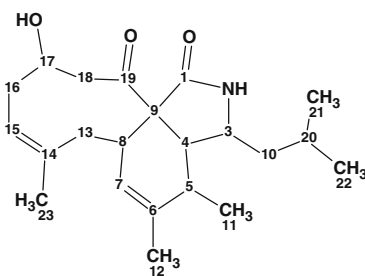
**Fig. 4.31** Hunanamycin A: Slightly edited MCD (see text)

Fig. 4.32 Hunanamycin A:
The resultant structure



4.14 Periconiasin

Cytochalasans are structurally characterized by their tricyclic core which consists of a macrocyclic ring fused to an isoindolone moiety contributed by a highly reduced polyketide backbone and an amino acid (for example, leucine or phenylalanine). Over 80 different cytochalasans have been isolated, and their macrocyclic rings all featured 11- to 13-membered carbocyclic or 12- to 14-membered lactone rings, which were assembled by the condensation of 8/9 acetate/malonate units. Zhang et al. [16] isolated three novel cytochalasans, periconiasins A–C, which featured an *unusual 9/6/5 tricyclic ring system*. Their molecular structures together with the relative configuration were unambiguously established on the basis of extensive spectroscopic data analysis. Here we will carry out the structure elucidation of Periconiasin A (**4.26**) using the spectroscopic data adopted from [16].



4.26

Periconiasin A was obtained as a colorless gum and gave an HRESIMS ion peak at m/z 360.2528 $[M+H]^+$ which corresponded to a molecular formula of $C_{22}H_{33}NO_3$ with seven degrees of unsaturation. The IR absorption bands at 3351, 3270, 3217, and 1690 cm^{-1} allow suggesting the presence of hydroxyl and/or amine, and carbonyl moieties.

^1H NMR, ^{13}C NMR, DEPT, HSQC, HMBC, and ^1H - ^1H COSY spectroscopic data (only key HMBC and COSY correlations were graphically presented in the article) which were used for the structure elucidation are presented in Table 4.14.

The MCD (Fig. 4.33) maps the spectroscopic data graphically.

Table 4.14 Periconiasin A: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M (J)	COSY	C HMBC
C1	175.7	175.96	C	–	–	–	–
C3	49.5	53.75	CH	3.02	u	1.02, 2.11, 8.19	C5, C1
C4	55.3	49.78	CH	2.11	u	2.39, 3.02	C9, C1, C19
C5	34.8	34.04	CH	2.39	u	2.11, 1.11	C9
C6	138.7	139.55	C	–	–	–	–
C7	128.9	122.51	CH	5.4	u	2.41	–
C8	43.6	39.68	CH	2.41	u	5.40, 4.08	C6
C9	65.9	65.51	C	–	–	–	–
C10	48.8	48.59	CH ₂	1.02	u	1.60, 3.02	–
C11	12.9	14.08	CH ₃	1.11	u	2.39	C6, C4
C12	19.4	20.64	CH ₃	1.71	u	–	C5, C7
C13	30.8	40.67	CH ₂	1.58	u	–	–
C13	30.8	40.67	CH ₂	4.08	u	2.41	C8, C9, C14, C7
C14	136.6	135.92	C	–	–	–	–
C15	122.7	124.56	CH	5.08	u	2.69	C17
C16	37.3	35.03	CH ₂	2.69	u	3.75, 5.08	C18
C16	37.3	35.03	CH ₂	1.99	u	–	–
C17	69.3	68	CH	3.75	u	2.69, 3.30, 4.82	C19, C15
C18	46.7	47.1	CH ₂	3.3	u	3.75	–
C18	46.7	47.1	CH ₂	2.35	u	–	–
C19	212.2	210.83	C	–	–	–	–
C20	23.9	24.63	CH	1.6	u	1.02, 0.83	–
C21	21.6	22.65	CH ₃	0.83	u	–	–
C22	23.5	22.65	CH ₃	0.83	u	1.6	–
C23	23.6	22.7	CH ₃	1.54	u	–	C15
N1	100 ^a	–	NH	8.19	u	3.02	C1
O1	150 ^a	–	OH	4.82	u	3.75	C17

^a Fictitious ^{15}N and ^{17}O chemical shifts

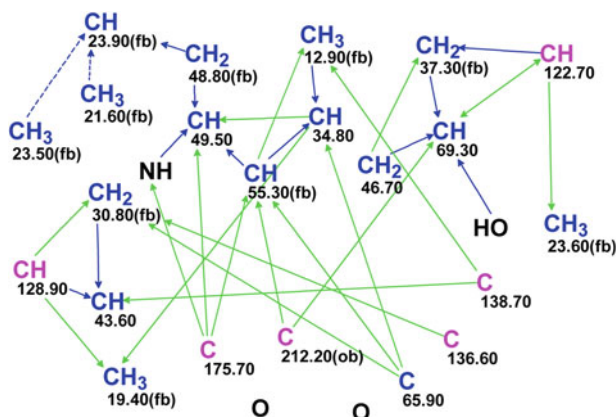
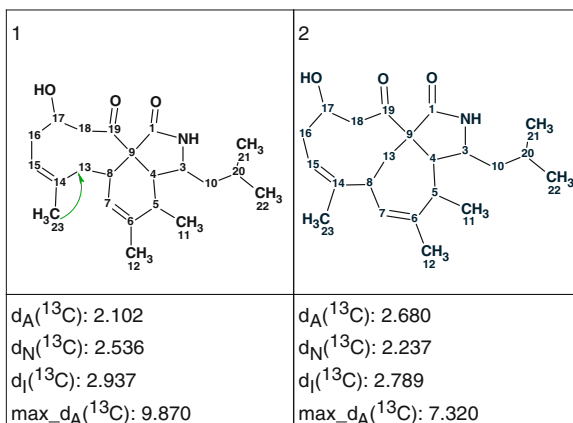


Fig. 4.33 Periconiasin A: Molecular connectivity diagram

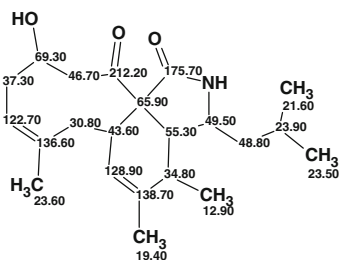
Fig. 4.34 Periconiasin A:
The two output structures
ranked by the average
deviations $d_A(^{13}\text{C})$



No contradictions were detected in the 2D NMR data so strict structure generation was performed. Results: $k = 2 \rightarrow 2 \rightarrow 2$, $t_g = 0.1$ s. The structures obtained are shown in Fig. 4.34.

We see that the average deviations calculated for both structures are fairly close due to the significant similarity of the structures. Nevertheless, structure ranking by average deviation $d_A(^{13}\text{C})$ distinguished the structure of Periconiasin A as the most probable one. At the same time ^{13}C chemical shift predictions using additive rules and neural nets indicate structure #2 as the preferred structure. In this situation, it is desirable to obtain additional confirmation of the preferable structure. Here we should take into account that only key HMBC and COSY correlations were used for structure generation. Analysis of the HMBC pattern available from the Supporting Information associated with article [16] shows that a connectivity of standard length exists from carbon 23 to carbon 13 that must be added to the 2D NMR data. It is

obvious that only structure #1 meets this constraint, which confirms the solution obtained (structure **4.27**).



4.27

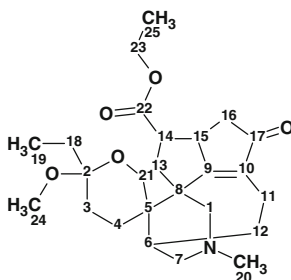
If the full HMBC dataset was input into the program from the very beginning then the competing structure #2 would not be generated at all. As the presence of nonstandard correlations in HMBC data is not excluded a 1,1-ADEQUATE experiment would be crucial for obtaining an unambiguous solution.

Thus the *unprecedented structure* of Periconiasin A was determined instantaneously and fully automatically without any user intervention.

4.15 Daphmacromine A

Daphniphyllum alkaloids are a family of structurally diverse natural products with complex polycyclic systems elaborated by plants of the genus *Daphniphyllum*. They are characterized by *unique* structural features which have attracted great interest from researchers.

Cao and coworkers [17] recently described the isolation and structure elucidation of 10 new alkaloids belonging to this family. Spectral data presented in reference [17] for one of these alkaloids, Daphmacromine A, (**4.28**), were examined using Structure Elucidator.



4.28

The molecular formula, $C_{25}H_{37}NO_5$, was established by positive ion HRESIMS (m/z 432.2758 $[M+H]^+$, calcd 432.2749), corresponding to eight degrees of unsaturation. IR absorptions indicated the presence of ester carbonyl ($1,729\text{ cm}^{-1}$) and likely α,β -unsaturated carbonyl ($1,694$ and $1,662\text{ cm}^{-1}$) groups. ^1H and ^{13}C chemical shifts of this compound were presented by the authors [17] in a tabular form, but COSY and key HMBC correlations, not all, were depicted graphically,

In spite of the absence of a full HMBC dataset, elucidation of the structure from the limited spectrum-structural information was attempted. The molecular formula and available 1D and 2D NMR data were input into Structure Elucidator (Table 4.15).

Figure 4.35 shows the MCD.

MCD overview Using the system's knowledge regarding spectrum-structure correlations, the program automatically set the atom hybridizations (blue atoms— sp^3 , violet atoms— sp^2 , a light blue atom— sp^3 or sp^2) and the possibilities of a heteroatom as a neighbor (“*ob*”—obligatory, “*fb*”—forbidden). Note that the hybridization of atom C(98.90) is marked as sp^3 or sp^2 , because the possibilities of an atom belonging to a C=C double bond and to an acetal O–C–O are taken into account. At the same time the atom C(209.20) was labeled as “*ob*” (a neighborhood with a heteroatom is obligatory), while atoms C(173.20) and C(183.70) have no labels indicating the presence (*ob*) or absence (*fb*) of a heteroatom in the first sphere of the atom environment is allowed. It is most probable that, in general, when the molecule under study is a common (trivial) chemical compound and only NMR spectra are used, a spectroscopist would assign the C(173.20) and C(183.70) atoms to carbonyls from either an ester, amide, or ketone. However, it is possible to see from structure 4.28 and Table 4.15 that atom C(183.70) is included into a C=C double bond and has neither O or N in its closest environment. It turns out that this chemical shift value is specific to all *Daphniphyllum* alkaloids [17].

Since we are modeling a real situation where Structure Elucidator is utilized for the elucidation of an unknown structure, we initially investigated what solution to the problem would be obtained if we relied only on “common sense” and ascribed the label “*ob*” to both of the carbon atoms C(173.20) and C(183.70). Note that when adding the labels we introduced new structural constraints which led to a reduced number of possible structures and a reduced time of structure generation.

The properties of the C(173.20) and C(183.70) atoms were modified in the MCD and strict structure generation was initiated using the common option settings (no NSCs were present in the 2D NMR data according to [17]). The results gave: $k = 2,816 \rightarrow 2,320 \rightarrow 1,056$; generation time $t_g = 8$ s. After ^{13}C chemical shift prediction and structure ranking according to the standard methodology common to the Structure Elucidator system, a ranked output structural file was obtained. The four top structures of the ranked file along with chemical shift deviations calculated are shown in Fig. 4.36:

We see that both the average and maximum deviations are too big for the first (“best”) structure, and consequently the structural constraints introduced by us were likely mistaken. Note that only several minutes were needed to reject the initial

Table 4.15 Daphmacromine A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	HMBC
C1	62.4	58.21	CH ₂	2.47	u	–	–
C1	62.4	58.21	CH ₂	2.38	u	–	–
C2	98.9	101.14	C	–	–	–	–
C3	27.6	31.74	CH ₂	1.62	u	1.59	C2
C4	22.6	24.62	CH ₂	1.59	u	1.62	C2, C6
C4	22.6	24.62	CH ₂	2.02	u	–	–
C5	36.8	39.9	C	–	–	–	–
C6	32.7	39.88	CH	2.35	u	1.76, 2.59	
C7	55.6	56.22	CH ₂	2.71	u	–	–
C7	55.6	56.22	CH ₂	2.59	u	2.35	C1
C8	50.1	48.15	C	–	–	–	–
C9	183.7	179.93	C	–	–	–	–
C10	138.2	136.67	C	–	–	–	–
C11	21	20.44	CH ₂	2.53	u	–	–
C11	21	20.44	CH ₂	2.43	u	1.76	C9
C12	26.1	29.44	CH ₂	1.76	u	2.35, 2.43	
C12	26.1	29.44	CH ₂	2.14	u	–	–
C13	37.9	37.25	CH ₂	2.55	u	–	–
C13	37.9	37.25	CH ₂	1.77	u	3.05	C22, C5, C8
C14	42.4	45.75	CH	3.05	u	3.21, 1.77	C22
C15	44.7	45.71	CH	3.21	u	3.05, 2.46	C22
C16	38.3	42.68	CH ₂	2.46	u	3.21	C17, C9, C10
C16	38.3	42.68	CH ₂	2.55	u	–	–
C17	209.2	208.5	C	–	–	–	–
C18	28.6	30.45	CH ₂	1.41	u	0.84	C2
C18	28.6	30.45	CH ₂	1.7	u	–	–
C19	7.8	7.43	CH ₃	0.84	u	1.41	C2
C20	46.1	45.87	CH ₃	2.14	u	–	C7
C21	62.5	65.26	CH ₂	3.55	u	–	–
C21	62.5	65.26	CH ₂	3.51	u	–	C2, C6
C22	173.2	174.52	C	–	–	–	–
C23	60.4	60.42	CH ₂	4.08	u	1.25	C22
C24	47	47.84	CH ₃	3.11	u	–	C2
C25	14.2	14.71	CH ₃	1.25	u	4.08	–

wrong assumption. The second run was performed from the *initial* MCD that was automatically created by the program. The result gave: $k = 1,254 \rightarrow 1,112 \rightarrow 376$, $t_g = 4$ s.

The top structures of the ranked file are shown in Fig. 4.37.

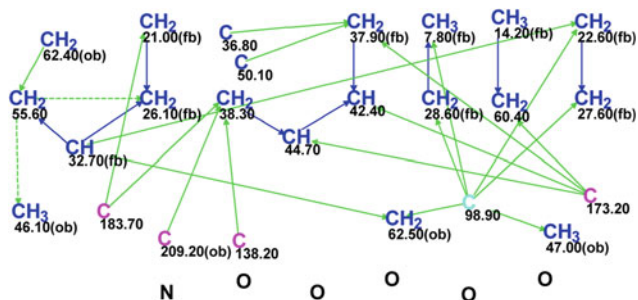


Fig. 4.35 Daphmacromine A: Molecular connectivity diagram

1	2	3	4
$d_A(^{13}\text{C})$: 4.035 $d_N(^{13}\text{C})$: 3.938 $d_I(^{13}\text{C})$: 4.493 $\text{max_}d_A(^{13}\text{C})$: 24.320	$d_A(^{13}\text{C})$: 4.487 $d_N(^{13}\text{C})$: 4.663 $d_I(^{13}\text{C})$: 4.602 $\text{max_}d_A(^{13}\text{C})$: 19.620	$d_A(^{13}\text{C})$: 4.775 $d_N(^{13}\text{C})$: 5.096 $d_I(^{13}\text{C})$: 4.496 $\text{max_}d_A(^{13}\text{C})$: 19.260	$d_A(^{13}\text{C})$: 4.893 $d_N(^{13}\text{C})$: 5.277 $d_I(^{13}\text{C})$: 4.845 $\text{max_}d_A(^{13}\text{C})$: 18.290

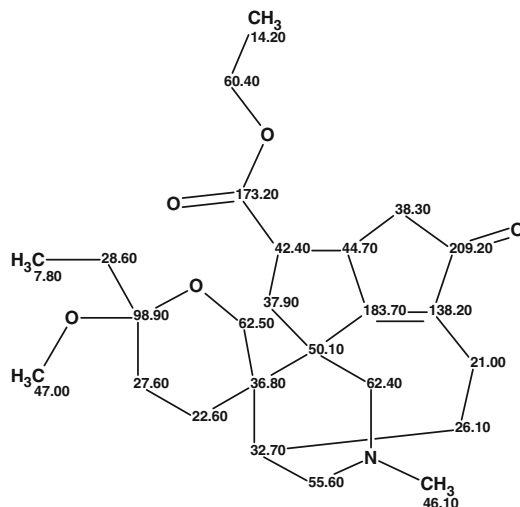
Fig. 4.36 Daphmacromine A: Top ranked structures resulted from the first program run

1	2	3	4
$d_A(^{13}\text{C})$: 1.908 $d_N(^{13}\text{C})$: 2.105 $d_I(^{13}\text{C})$: 2.311 $\text{max_}d_A(^{13}\text{C})$: 9.590	$d_A(^{13}\text{C})$: 2.364 $d_N(^{13}\text{C})$: 2.579 $d_I(^{13}\text{C})$: 2.681 $\text{max_}d_A(^{13}\text{C})$: 10.030	$d_A(^{13}\text{C})$: 2.988 $d_N(^{13}\text{C})$: 3.393 $d_I(^{13}\text{C})$: 3.592 $\text{max_}d_A(^{13}\text{C})$: 14.700	$d_A(^{13}\text{C})$: 3.291 $d_N(^{13}\text{C})$: 3.019 $d_I(^{13}\text{C})$: 3.621 $\text{max_}d_A(^{13}\text{C})$: 11.430

Fig. 4.37 Daphmacromine A: Top ranked structures resulting from the second program run (valid solution)

We see that the best structure coincides with structure **4.28** and the values of the deviations are common for the majority of problems solved using Structure Elucidator. The structure of Daphmacromine A (**4.29**) supplied with the automatically assigned ^{13}C chemical shifts is shown below:

This example clearly shows that even well-known spectrum-structure NMR correlations can be violated in complex polycyclic structures. However, utilizing Structure Elucidator allows a researcher to minimize the amount of expert



4.29

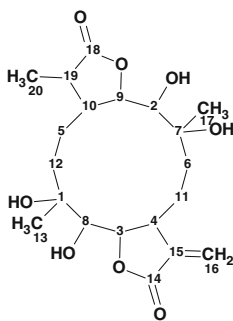
assumptions (“axioms”) and to get the correct solution to a problem in a short time. Note that no user assumptions were used to elucidate the structure of Daphmacromine A.

4.16 Eryngiolide A

Wang et al. [18] separated and identified a novel macrocyclic diterpene Eryngiolide A possessing a previously *undescribed skeleton* originating from a cyclododecane core fused with two γ -lactone units (**4.30**).

The α,β -unsaturated lactone moiety has been regarded as one of the most significant structural characteristics for many biologically active natural products.

The molecular formula of Eryngiolide A was unambiguously established as $C_{20}H_{30}O_8$ on the basis of HRESIMS (found $m/z = 421.1834$ $[M+Na]^+$, calculated:



4.30

$m/z = 421.1833$). A tabular representation is given for the ^{13}C and ^1H NMR spectra in reference [18], while only selected COSY and HMBC correlations are available in graphical form.

This restricted information was nevertheless used for computer-assisted structure elucidation. The data used are presented in Table 4.16.

The following MCD was created by the program from the molecular formula and NMR initial data (Fig. 4.38):

MCD overview The overwhelming majority of COSY and HMBC connectivities are ambiguous (they are marked by dotted lines). This is a result of the presence of four overlapped peaks at δC 20.3 and three pairs of overlapped peaks (26.8, 33.5, 120.2) in the ^{13}C spectrum, as well as one pair (δH 1.61) in the proton spectrum. For the two carbon atoms C(73.1) and C(73.5), hybridization was set as sp^2 or sp^3 and in addition atom C(73.1) has no connectivities. For the latter carbon atom this means that the program will try all possibilities to embed the carbon atom C(73.1) in the structure. The authors of reference [18] distinguished “five

Table 4.16 Eryngiolide A: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	HMBC
C1	73.1	74.43	C	–	–	–	–
C2	78.8	77.15	CH	3.04	d(8.2)	4.63	–
C3	84.7	79.61	CH	4.63	u	3.35, 3.04	C14, C15
C4	41.6	43.27	CH	3.35	u	1.91, 4.63	–
C5	20.3	21.72	CH_2	2.17	u	–	–
C5	20.3	21.72	CH_2	1.91	u	1.47, 3.35	C15, C7
C6	33.5	31.38	CH_2	1.62	u	–	–
C6	33.5	31.38	CH_2	1.47	u	1.91	–
C7	73.5	74.43	C	–	–	–	–
C8	77.4	76.49	CH	3.33	u	4.47	–
C9	84.7	80.75	CH	4.47	u	3.33, 2.40	C19, C18
C10	44.8	42.5	CH	2.4	u	2.08, 2.58, 4.47	C12
C11	20.3	25.39	CH_2	1.61	u	2.08	–
C11	20.3	25.39	CH_2	2.08	u	2.40, 1.61, 1.61	–
C12	33.9	34.54	CH_2	1.61	u	2.08	–
C13	26.8	24.22	CH_3	1.23	s	–	C2, C12
C14	173.1	170.44	C	–	–	–	–
C15	139.5	140.02	C	–	–	–	–
C16	120.2	122.1	CH_2	6.2	u	–	C4
C16	120.2	122.1	CH_2	5.64	u	–	C4, C14
C17	26.8	24.22	CH_3	1.27	s	–	C8, C6
C18	182.6	180.45	C	–	–	–	–
C19	37.7	41.94	CH	2.58	u	1.19, 2.40	–
C20	14	12.58	CH_3	1.19	d(6.8)	2.58	C18, C10

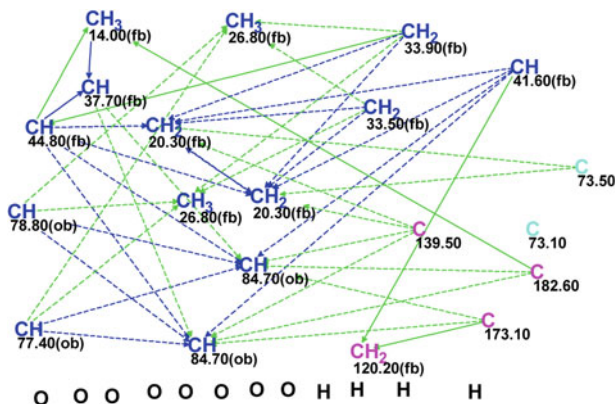


Fig. 4.38 Eryngiolide A: Molecular connectivity diagram

quaternary carbons including two oxygenated carbons (δC 73.1 and 73.5), an olefinic carbon (δC 139.5), and two ester carbonyl carbons (δC 173.1 and 182.0)". These suggestions may seem rather believable (as a contrary instance see structure 2.29), but when StrucEluc is used we can frequently refuse some non-evident assumptions and initiate structure generation from the MCD as it is. All MCD peculiarities listed hamper structure generation and make the generation time relatively long for large molecules. To facilitate solving the problem, defined multiplicities for four of the ^1H NMR signals presented in Table 4.16 were used to set constraints on a number of hydrogen atoms attached to neighboring carbon atoms. For instance, for δH 3.04, a doublet with $J_{\text{HH}} = 8.2$ Hz, this number is equal to one.

Structure generation was combined with ^{13}C chemical shift prediction (structures with $d_i > 4$ ppm were rejected during structure generation), and structural and spectral filtering gave the following results: $k = 52 \rightarrow 26 \rightarrow 9$, $t_g = 6$ min. The top structures of the ranked output file are shown in Fig. 4.39.

The best structure was distinguished by all three methods of ^{13}C chemical shift prediction and coincided with structure 4.30 (see structure 4.31). In spite of the large number of ambiguous correlations the program found a valid solution in a reasonable time.

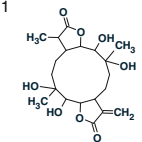
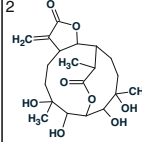
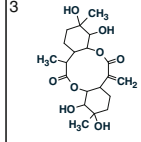
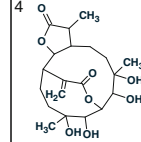
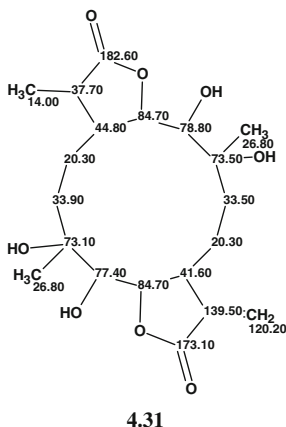
			
$d_A(^{13}\text{C}): 2.378$ $d_N(^{13}\text{C}): 2.184$ $d_l(^{13}\text{C}): 2.098$ $\text{max_}d_A(^{13}\text{C}): 5.690$	$d_A(^{13}\text{C}): 2.676$ $d_N(^{13}\text{C}): 2.962$ $d_l(^{13}\text{C}): 2.588$ $\text{max_}d_A(^{13}\text{C}): 12.070$	$d_A(^{13}\text{C}): 3.049$ $d_N(^{13}\text{C}): 2.865$ $d_l(^{13}\text{C}): 3.497$ $\text{max_}d_A(^{13}\text{C}): 9.230$	$d_A(^{13}\text{C}): 3.299$ $d_N(^{13}\text{C}): 3.163$ $d_l(^{13}\text{C}): 3.031$ $\text{max_}d_A(^{13}\text{C}): 12.010$

Fig. 4.39 Eryngiolide A: The top structures of the ranked output file

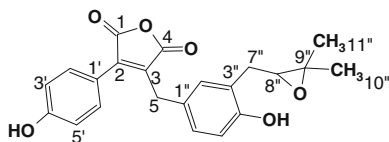


It was interesting to know how the process of structure generation could be accelerated if some authors' [18] assumptions were used. When both "light blue" carbon atoms C(73.10) and C(73.5) were reassigned as sp^3/ob (sp^3 -hybridized and having oxygen atoms as neighbors) the generation time reduced drastically with the result: $k = 52 \rightarrow 26 \rightarrow 9$, $t_g = 0.6$ s, almost 600 times faster! The generated structures were the same as those found during the first run. This example shows that the application of reasonable user assumptions ("axioms") can speed up the program drastically. Of course, the assumptions are helpful only if they turn out to be valid.

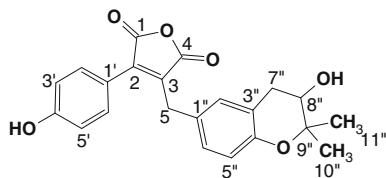
4.17 Asperjinone

During the course of preparing problems for this book we utilized spectroscopic data reported by Liao and coworkers [19] to deduce the structure **4.32** of a new natural product named as Asperjinone (atom numbering is shown as in the original article [19]). This compound was isolated, along with other 12 known compounds,

from *Aspergillus terreus*. As a result of our analysis using Structure Elucidator the structure of **4.32** was revised [20] and we suggested that structure **4.33** is the correct structure (atom numbering is shown as in the original article [19]).



4.32



4.33

The molecular formula $C_{22}H_{20}O_6$ and the NMR data presented in Table 4.17 obtained from the article [19] were used as input into the Structure Elucidator software.

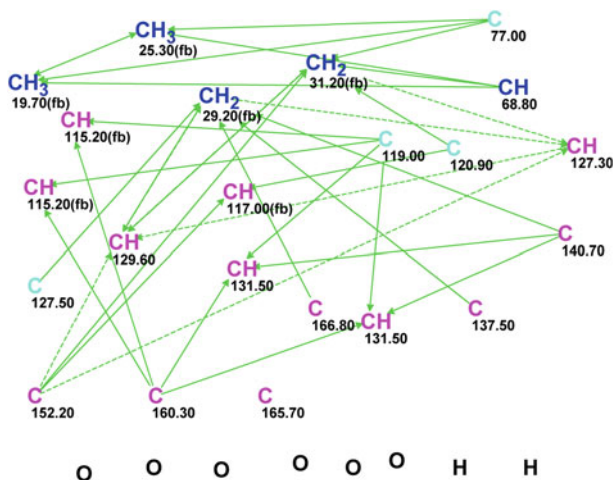
The MCD automatically created by the program is presented in Fig. 4.40.

MCD overview Two atoms (colored in pale blue) in the MCD—C(119.0) and C(120.9)—were classified as having ambiguous hybridization because these chemical shifts are characteristic for both the C=C double bonds (sp^2) and for C(sp^3) atom if it is included into an O—C—O fragment. Carbons with chemical shifts falling into the interval 152–167 ppm are likely connected with at least one oxygen atom. The information presented in the MCD was used by the program for the purpose of structure generation. As a result all structures in agreement with the HMBC correlations and atom properties were produced. No expert considerations common for a traditional approach regarding HMBC correlations were introduced. No structural inputs regarding the presence of aromatic rings or other conceivable rings in the structure were made.

The following results from the structure generation process were obtained: $k = 3,658 \rightarrow 2,641 \rightarrow 1,939$, $t_g = 58$ s. The first eight structures of the ranked file are displayed in Fig. 4.41. Atoms for which $\Delta = |\delta C_{\text{calc}} - \delta C_{\text{exp}}|$ value, the difference between experimental and calculated chemical shifts, is less than 3 ppm are marked by green circles, yellow circles correspond to $\Delta = 3\text{--}15$ ppm, and red to $\Delta > 15$ ppm. The figure shows that the first ranked structure (all green) is characterized by the smallest deviations calculated by HOSE code and neural network-based methods, while the structure proposed by Liao and coworkers [19] was

Table 4.17 Asperjinone: 1D and 2D spectroscopic data (600 MHz, Acetone- d_6)

Position	δC	CH_n	δH	M(J)	HMBC ^a
1	165.7	C	–	–	–
2	140.7	C	–	–	–
3	137.5	C	–	–	–
4	166.8	C	–	–	–
5	29.2	CH ₂	3.97 3.98	d(11.2) d(11.2)	C2, 3, 4, 1'',2''
1'	119.0	C	–	–	–
2',6'	131.5	CH	7.63	d(8.1)	C2, 1',2',4'
3',5'	115.8	CH	7.01	d(8.1)	C1',4'
4'	160.3	C	–	–	–
1''	127.5	C	–	–	–
2''	129.6	CH	6.99	m	C4'',6'',7''
3''	120.9	C	–	–	–
4''	152.2	C	–	–	–
5''	117.0	CH	6.66	d(8.6)	C3'',4''
6''	127.3	CH	6.99	m	C5
7''	31.2	CH ₂	2.67 2.94	dd(16.9, 8.0) dd(16.9, 5.0)	C2'',3'',4'',8'',9''
8''	68.8	CH	3.76	m	–
9''	77.0	C	–	–	–
10''	19.7	CH ₃	1.22	s	C8'',9'',11''
11''	25.3	CH ₃	1.33	s	C8'',9'',10''

**Fig. 4.40** Asperjinone: The MCD extracted from the spectroscopic data

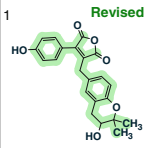
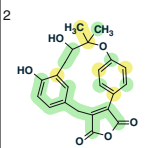
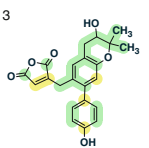
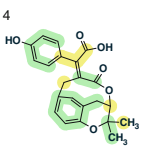
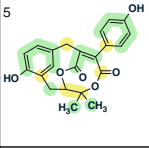
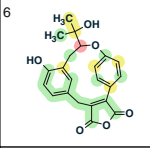
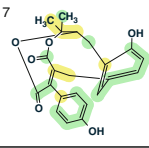
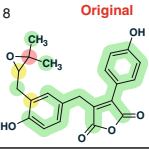
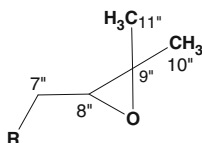
1 	2 	3 	4 
$d_A(^{13}\text{C})$: 1.128 $d_N(^{13}\text{C})$: 1.533 $d_I(^{13}\text{C})$: 1.613 $\text{max_}d_A(^{13}\text{C})$: 2.600	$d_A(^{13}\text{C})$: 2.158 $d_N(^{13}\text{C})$: 2.696 $d_I(^{13}\text{C})$: 2.681 $\text{max_}d_A(^{13}\text{C})$: 8.180	$d_A(^{13}\text{C})$: 2.261 $d_N(^{13}\text{C})$: 2.890 $d_I(^{13}\text{C})$: 2.433 $\text{max_}d_A(^{13}\text{C})$: 12.560	$d_A(^{13}\text{C})$: 2.288 $d_N(^{13}\text{C})$: 2.574 $d_I(^{13}\text{C})$: 2.210 $\text{max_}d_A(^{13}\text{C})$: 6.630
5 	6 	7 	8 
$d_A(^{13}\text{C})$: 2.365 $d_N(^{13}\text{C})$: 2.915 $d_I(^{13}\text{C})$: 2.703 $\text{max_}d_A(^{13}\text{C})$: 11.510	$d_A(^{13}\text{C})$: 2.495 $d_N(^{13}\text{C})$: 2.833 $d_I(^{13}\text{C})$: 2.640 $\text{max_}d_A(^{13}\text{C})$: 16.880	$d_A(^{13}\text{C})$: 2.525 $d_N(^{13}\text{C})$: 2.752 $d_I(^{13}\text{C})$: 2.752 $\text{max_}d_A(^{13}\text{C})$: 10.870	$d_A(^{13}\text{C})$: 2.621 $d_N(^{13}\text{C})$: 2.701 $d_I(^{13}\text{C})$: 2.758 $\text{max_}d_A(^{13}\text{C})$: 17.810

Fig. 4.41 Asperjinone: The first eight structures of the ranked output file

placed in *eighth* position by the ranking procedure. The deviation is almost twice the size of that given for the structure ranked in first position.

To confirm the revised structure, **4.33**, we performed a search for the (3,3-dimethyloxiran-2-yl)methyl fragment, as shown below and with the numbering system included for easy recognition of the substructure in the reference molecule.



A fragment search was carried out in the ACD/NMR Database containing 425,000 structures with assigned ^{13}C and ^1H chemical shifts.

The program selected almost 180 structures from which such *ca.* 150 structures were chosen that exhibit the closest similarity with the environment of the oxirane fragment. For these structures, a scatter plot was created (see Fig. 4.42). Here the ^{13}C chemical shifts related to the C8'' and C9'' atoms of structure **4.32** are presented for all selected structures. The chemical shift values (69 and 77 ppm) assigned to the corresponding atoms C8'' and C9'' in the original structure **4.32** are also shown by their labels on the right side of the graph.

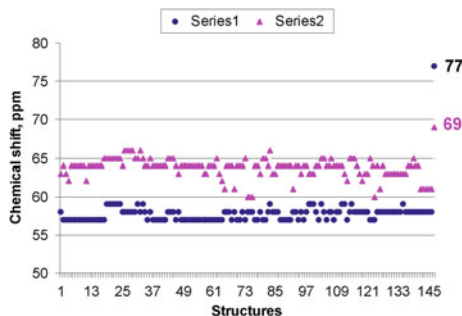


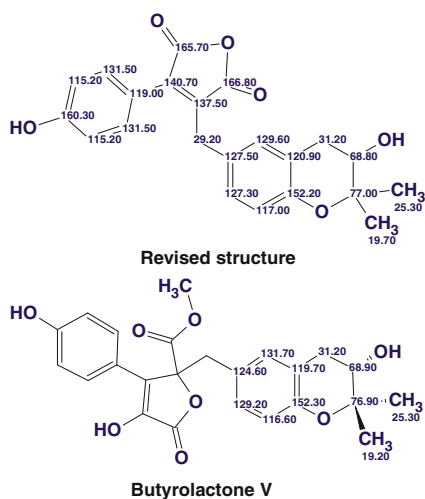
Fig. 4.42 Asperjinone: A scatter plot of the ^{13}C chemical shift values related to atoms 8'' and 9'' of the original structure **4.32**. Series 1 (blue circles) corresponds to atom 9'' (δC 77 ppm in structure **4.32**), series 2 (violet triangles)—to atom 8'' (δC 69 ppm in structure **1**) [20]

Inspection of the scatter plot convincingly confirms the incorrectness of the original structure: the chemical shifts of C8'' (68.8 ppm in structure **4.32**) are observed in the range of 60–65 ppm while for C9'' (77.0 ppm in structure **4.32**) the corresponding range is 57–59 ppm.

On the other hand, corroboration of the revised structure **4.33** was found in the Supporting Information of the original work. One of the compounds separated by the authors along with Asperjinone (designated as Butyrolactone V) was characterized, and ^{13}C and ^1H NMR chemical shifts were assigned to the structure of Butyrolactone V. This compound contains the revised structural component of structure **4.33**. Both structures supplied with the assigned ^{13}C chemical shifts (for butyrolactone V only partial assignment is shown) are presented in Fig. 4.43.

The comparison leaves no doubts regarding the correctness of structure **4.33**. Moreover, oxirane $^1J_{\text{CH}}$ couplings are typically ~ 180 Hz, far larger than other

Fig. 4.43 Asperjinone: Comparison of chemical shifts in the revised part of structure **4.33** with those in Butyrolactone V

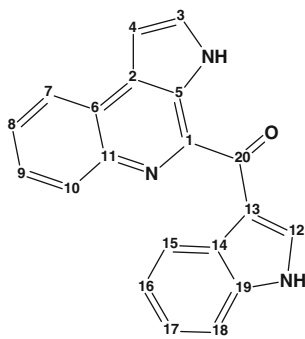


oxygen-bearing aliphatic carbon. If Liao and coworkers had measured this value for the C8" atom they would easily realize that their suggested structure was wrong.

Therefore, the suggestion within [19] regarding the existence of an oxirane ring in the Asperjinone structure proved to be erroneous. We believe that the application of a CASE system to the structure elucidation of this natural product would have allowed the authors to avoid this incorrect structure as an output from their analysis. It should be noted that as far as we know this is the first example when reliable structure revision was performed only with the aid of a CASE system without performing additional experiments or quantum-chemical NMR shift calculations.

4.18 Marinoquinoline F

Okanya and coworkers [21] isolated a new natural product Marinoquinoline F (4.34) along with a series of related compounds from the strain PWU 25. Its structure was determined from 1D NMR, HSQC, HMBC, and COSY spectra (in total 26 correlations were observed in the COSY and HMBC, see Table 4.18).



4.34

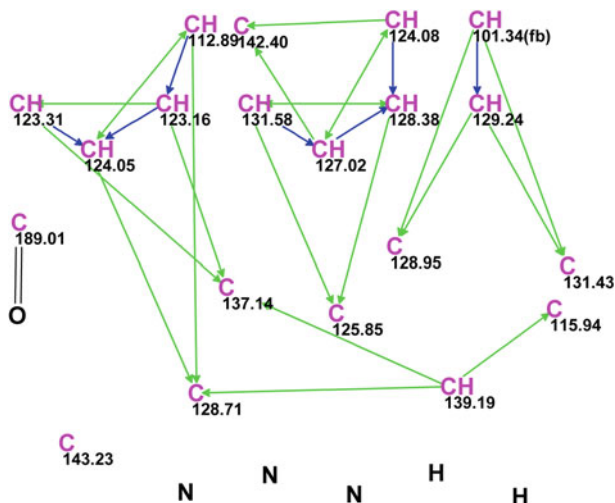
The molecular formula of Marinoquinoline F $C_{20}H_{13}N_3O$ was derived from the HRESIMS analysis of the $[M+H]^+$ ion at m/z 312.1138 (calculated for $C_{20}H_{13}N_3O$, 312.1131). The ratio of skeletal atoms to hydrogens is close to 2, which is a specific feature of challenging problems (see Sect. 1.2.2).

The MCD created by the program is shown in Fig. 4.44.

MCD overview The MCD allows us to see some specific peculiarities of the ^{13}C NMR spectrum: the ^{13}C chemical shifts of all carbons except C 189.01 are observed in the narrow interval between 100 and 140 ppm. This implies many possibilities for the carbon chemical shift assignments. An additional specific trait of the initial data is that all carbon atoms are in the sp^2 hybridization state, while the possibility of connections with neighboring heteroatoms is not defined for any of them except

Table 4.18 Marinoquinoline F: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	143.23	135.76	C	–	–	–	–
C2	131.43	127.32	C	–	–	–	–
C3	129.24	128.64	CH	7.79	u	7.25	C5, C4, C2
C4	101.34	99.62	CH	7.25	u	7.79	C3, C2, C5
C5	128.95	130.22	C	–	–	–	–
C6	125.85	121.67	C	–	–	–	–
C7	124.08	122.84	CH	8.39	dd(7.5, 2.0)	7.7	C9, C11
C8	128.38	127.6	CH	7.7	td(6.6, 1.8)	8.39, 7.68	C7, C10, C6
C9	127.02	126.96	CH	7.68	td(7.5, 1.8)	7.70, 8.30	C7, C11
C10	131.58	129.92	CH	8.3	dd(7.7, 1.5)	7.68	C8, C6
C11	142.4	140.94	C	–	–	–	–
C12	139.19	137.6	CH	9.7	u	–	C13, C14, C19
C13	115.94	114.31	C	–	–	–	–
C14	128.71	127.94	C	–	–	–	–
C15	123.31	122.1	CH	8.65	dd(6.6, 2.5)	7.29	C19, C16
C16	124.05	121.85	CH	7.29	td(7.3, 1.8)	7.31, 8.65	C15, C14, C18
C17	123.16	122.35	CH	7.31	td(7.5, 1.4)	7.29, 7.60	C19, C15
C18	112.89	112.07	CH	7.6	u	7.31	C14, C16, C17
C19	137.14	135.8	C	–	–	–	–
C20	189.01	190.81	C	–	–	–	–

**Fig. 4.44** Marinoquinoline F: The MCD. The C=O bond was drawn by hand

one (C 101.34 *fb*). In addition, carbons 143.23 and 189.01 have no connectivities with any other atoms. These particular qualities of the initial data suggest that the number of generated structures will be large and the generation time will be long. To assist the structure generation process the double bond C=O was drawn on the MCD by hand as carbon C 189.01 obviously belongs to a carbonyl bond. The number of hydrogen atoms present in the first sphere were set for carbon atoms whose attached hydrogens were manifested by definite multiplicities in the ^1H NMR spectrum (see column M(J), Table 4.18).

MCD checking has not revealed any contradictions in the 2D NMR data, hence strict structure generation was initiated. ^{13}C NMR chemical shifts were calculated during the structure generation process and structures characterized with deviation values of $d > 4$ ppm were discarded by the filter. Results: $k = 3,312,184 \rightarrow 32 \rightarrow 12$, $t_g = 40$ min. The top six structures of the ranked output file are shown in Fig. 4.45.

The structure prioritized as structure #1 is identical to that elucidated for Marinoquinoline F and is confirmed by all three methods of ^{13}C chemical shift prediction. Automatic ^{13}C chemical shift assignment is shown on structure 4.35.

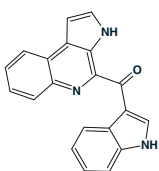
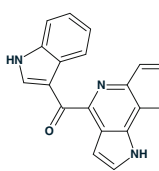
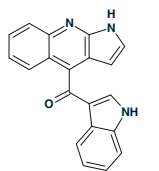
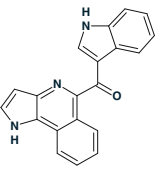
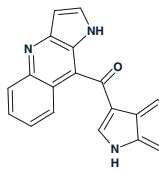
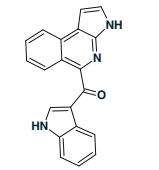
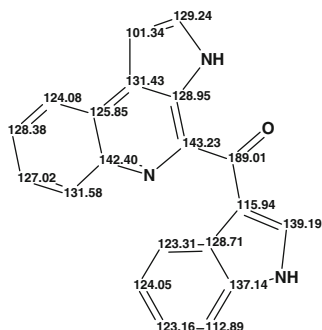
1  1.0000	2  1.0000	3  0.9932
$d_A(^{13}\text{C})$: 1.836 $d_N(^{13}\text{C})$: 2.002 $d_I(^{13}\text{C})$: 1.530 $\text{max}_d(^{13}\text{C})$: 7.470	$d_A(^{13}\text{C})$: 2.870 $d_N(^{13}\text{C})$: 2.509 $d_I(^{13}\text{C})$: 2.582 $\text{max}_d(^{13}\text{C})$: 11.530	$d_A(^{13}\text{C})$: 3.334 $d_N(^{13}\text{C})$: 3.424 $d_I(^{13}\text{C})$: 3.772 $\text{max}_d(^{13}\text{C})$: 14.750
4  0.9829	5  0.9373	6  0.9311
$d_A(^{13}\text{C})$: 3.404 $d_N(^{13}\text{C})$: 3.161 $d_I(^{13}\text{C})$: 2.724 $\text{max}_d(^{13}\text{C})$: 17.470	$d_A(^{13}\text{C})$: 3.835 $d_N(^{13}\text{C})$: 3.502 $d_I(^{13}\text{C})$: 3.612 $\text{max}_d(^{13}\text{C})$: 20.720	$d_A(^{13}\text{C})$: 3.883 $d_N(^{13}\text{C})$: 3.570 $d_I(^{13}\text{C})$: 3.267 $\text{max}_d(^{13}\text{C})$: 14.190

Fig. 4.45 Marinoquinoline F: The top six structures of the ranked output file. The Tanimoto similarity coefficients are shown in the *right upper corners* of the boxes



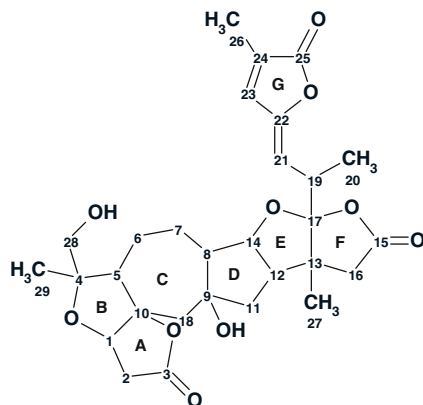
4.35

It is interesting to note that all top ranked six structures are similar (the Tanimoto similarity coefficients are shown in the right upper corners of the individual boxes), however ^{13}C chemical shift calculation allowed the program to confidently select the correct structure. The correct choice was made from among ~3.3 million generated structures thereby demonstrating the high “resolving power” of the CASE methodology.

4.19 Schilancitrilactone A

Schinortriterpenoids (*Schisandra* nortriterpenoids) are a series of naturally occurring polycyclic molecules, which are interesting for study of their structures, bioactivities, and synthesis. Some of them were found to possess anti-HIV-1 and antitumor bioactivities. As a consequence, these *exotic* schinortriterpenoids have drawn widespread attention and, especially, have brought herculean challenges and ambitious targets for organic synthesis endeavors. *Schisandra lancifolia*, belonging to the genus *Schisandra* of the family *Schisandraceae*, denotes excellent sources for the discovery of *architecturally intriguing* nortriterpenoids due to its accumulation of a high number of this class of secondary metabolites.

Chemical research on the stems of *S. lancifolia* conducted by Luo et al. [22] resulted in the isolation of three *unique* nortriterpenoids, Schilancitrilactones A–C. One of them, Schilancitrilactone A (**4.36**), was used to challenge Structure Elucidator. Compound **4.36** possesses a 5/5/7/5/5/5-fused hexacyclic ring system with a C₂₉ backbone. The three *cis*-fused five-membered rings (rings D–F), which possess the entire envelope conformations and involve six contiguous chiral centers (including two quaternary ones), form a structurally rigid tricyclic moiety.



4.36

Compound **4.36** was obtained as optically active colorless needle crystals. The molecular formula $C_{29}H_{36}O_{10}$ was established from the quasimolecular $[M+Na]^+$ ion at m/z 567.2214 in HRESIMS. The IR spectrum showed absorptions at 3,434 and $1,769\text{ cm}^{-1}$, revealing the existence of a hydroxyl group and carbonyl probably included into a small-cycle lactone [11]. ^1H , ^{13}C , and HSQC NMR data as well as selected key COSY and HMBC correlations available from the article [22] are presented in Table 4.19.

The slightly edited MCD is shown in Fig. 4.46.

MCD overview The initial MCD contained six light blue carbon atoms (C 82.7, C 87.1, C 98.4, C 111.3, C 121.9, C 130.7) whose hybridization was marked as “ sp^2 or sp^3 ” and all sp^2 -hybridized carbons did not receive labels indicating the possibility of the presence or absence of heteroatoms in their vicinities. MCD checking did not detect any contradictions in the COSY or the HMBC data. An attempt to perform strict structure generation was undertaken from the initial MCD. It allowed us to quickly (in ~15 min) realize that utilizing a restricted number of key 2D NMR correlations is insufficient for obtaining a solution in a reasonable time, so some additional constraints (assumptions) are needed to be introduced. Therefore, the property “ sp^3/ob ” was set for carbons C 82.7 and C 87.1, and at the same time carbons C 170.3, C 173.4, and C 173.7 were labeled with “ sp^2/ob ” (neighboring with a heteroatom was postulated). The number of hydrogen atoms attached to the nearest skeletal atoms were set in accordance with the ^1H signal multiplicities shown in column M(J) of Table 4.19. Then, the repeated strict structure generation accompanied with ^{13}C chemical shift calculation was initiated from the edited MCD (Fig. 4.46). Results: $k = 269,408 \rightarrow 4,490 \rightarrow 4,490$, $t_g = 18$ min. The three top ranked structures of the output file are shown in Fig. 4.47.

Figure 4.47 leads us to conclude that the first ranked structure #1 coincides with the structure of Schilancitrilactone A suggested by the authors [22]. ^{13}C chemical shift assignment as automatically performed by the program is displayed on structure **4.37**.

Table 4.19 Schilancitrilactone A: The spectroscopic NMR data

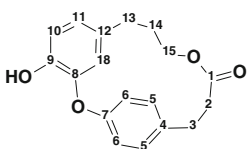
Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	C HMBC
C30	17.4	20.04	CH_3	1.06	S	–	C5, C29, C4
C1	80.8	81.53	CH	4.19	u	2.69	C3, C10
C2	34.9	35.45	CH_2	2.69	u	4.19	C3, C10
C2	34.9	35.45	CH_2	2.76	u	–	–
C3	173.4	176	C	–	–	–	–
C4	87.1	85.45	C	–	–	–	–
C5	52.8	56.67	CH	2.77	u	1.52	C4, C10
C6	21.5	26.88	CH_2	1.4	u	–	–
C6	21.5	26.88	CH_2	1.52	u	2.77, 1.86	–
C7	23.7	28.25	CH_2	1.86	u	1.52, 2.38	–
C8	53.3	56.57	CH	2.38	u	1.86, 4.47	C9
C9	82.7	79.12	C	–	–	–	–
C10	98.4	97.29	C	–	–	–	–
C11	41.9	39.72	CH_2	2.05	u	2.43	C9
C11	41.9	39.72	CH_2	1.7	u	–	–
C12	50.5	50.33	CH	2.43	u	4.47, 2.05	–
C13	50.4	49.93	C	–	–	–	–
C14	84.7	84.83	CH	4.47	u	2.38, 2.43	–
C15	173.7	174.1	C	–	–	–	–
C16	46.7	39.41	CH_2	2.44	u	–	C15
C16	46.7	39.41	CH_2	2.63	u	–	–
C17	121.9	122.1	C	–	–	–	–
C19	41.2	45.38	CH_2	1.91	u	–	C5, C10, C8
C20	36.1	31.46	CH	3.34	u	1.25, 4.97	–
C21	16.0	16.15	CH_3	1.25	d(6.7)	3.34	C20, C17, C22
C22	111.3	111.2	CH	4.97	d(10.5)	3.34	C17, C23, C24
C23	148.4	146.6	C	–	–	–	–
C24	137.5	138.6	CH	6.98	u	–	–
C25	130.7	129	C	–	–	–	–
C26	170.3	170.7	C	–	–	–	–
C27	10.6	10.73	CH_3	2.01	s	–	C25, C24, C26
C28	18.7	19.65	CH_3	1.19	s	–	C16, C17, C13, C12
C29	66.7	65.41	CH_2	3.57	u	–	–
C29	66.7	65.41	CH_2	3.4	u	–	–

The ^{13}C chemical shift assignment is the same as reported in article [22]. It is worth noting that the predicted ^{13}C chemical shifts are in good agreement with the experimental values in spite of the unusual skeleton of the molecule (see column $\delta\text{C}_{\text{calc}}$ in Table 4.19).

4.20 Isocorniculatolide A

Aegeceras corniculatum is a small tree or shrub that grows in the mangrove swamps of Asia and Australia. It has been used traditionally to treat asthma, diabetes, inflammation, and rheumatism. A number of saponins, triterpenes, sterols, and hydroquinones have been previously reported from this plant. The metabolites that were isolated from mangroves often possess unique structural features and incorporate new or unusual assemblages of functional groups.

In the course of search for bioactive metabolites from Indian mangrove plants, Ponnappalli et al. [23] isolated and identified four new *unusual* isomeric macrolides of combretastatin D-2 congeners. This was the *first isolation* of combretastatin D-2 congeners from *A. corniculatum*. The experimental spectra obtained for structure elucidation of one of them, Isocorniculatolide A (**4.38**), were used as initial data for feeding Structure Elucidator.



4.38

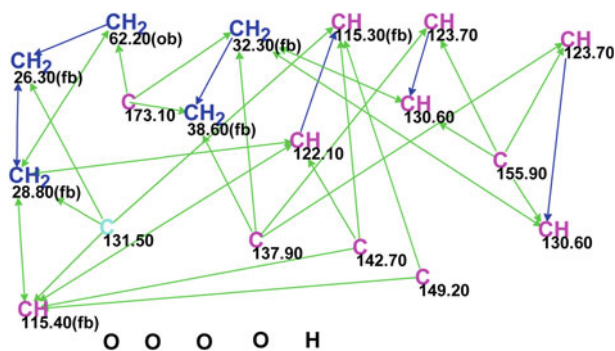
Compound **4.38** was obtained as colorless crystals. The HRESIMS of compound **4.38** displayed a protonated molecular ion $[M+H]^+$ at m/z 299.1283 (calculated 299.1278) to derive a molecular formula of $C_{18}H_{18}O_4$, suggesting 10 degrees of unsaturation. The IR spectrum exhibited absorption bands for both hydroxyl ($3,390\text{ cm}^{-1}$) and carbonyl ($1,714\text{ cm}^{-1}$) functionalities. The NMR spectroscopic data used in article [23] for the Isocorniculatolide A structure elucidation are collected in Table 4.20. The created MCD is presented in Fig. 4.48.

The analyzed molecule is of modest size and the number of connectivities is large enough to avoid any need to edit the MCD to ease problem solving. No contradictions were detected in the 2D NMR data following MCD checking, hence strict structure generation was carried out. The following results were obtained: $k = 5 \rightarrow 3 \rightarrow 3$, $t_g = 0.003$ s. The ranked structures of the output file are presented in Fig. 4.49.

Structures #1 and #2 are very similar, but the priority of structure #1, which is identical to the structure of Isocorniculatolide A, is confirmed by the d_A values (the difference is 0.54 ppm) though $d_I(\#2) < d_I(\#1)$. Unambiguous evidence of the priority of structure #1 was through NOESY correlations between hydrogen atom H18 attached to carbon atom C18 (structure **4.38**) and hydrogens H6, H14, and H15. ^{13}C chemical shift assignment for the Isocorniculatolide A molecule is displayed on structure **4.39**.

Table 4.20 Isocorniculatolide A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M (J)	COSY	C HMBC
C1	173.1	172.82	C	–	–	–	–
C2	38.6	33.02	CH_2	2.52	u	3.03	C4, C1, C3
C3	32.3	32.07	CH_2	3.03	u	2.52	C1, C5, C4, C2
C4	137.9	133.51	C	–	–	–	–
C5	130.6	130.08	CH	7.25	u	7.02	C7, C3, C6
C6	123.7	123.75	CH	7.02	u	7.25	C7, C4, C5
C7	155.9	156.8	C	–	–	–	–
C8	149.2	147.58	C	–	–	–	–
C9	142.7	146.81	C	–	–	–	–
C10	115.3	116.2	CH	6.85	u	6.6	C9, C8, C12
C11	122.1	123.67	CH	6.6	u	6.85	C9, C13, C16
C12	131.5	132.88	C	–	–	–	–
C13	28.8	32.19	CH_2	2.48	u	1.73	C11, C12, C14, C16, C15
C14	26.3	30.26	CH_2	1.73	u	2.48, 3.58	C15, C13, C12
C15	62.2	64.43	CH_2	3.58	u	1.73	C13, C1, C14
C16	115.4	115.94	CH	5.38	u	–	C12, C13, C8, C9, C11

**Fig. 4.48** Isocorniculatolide A: Molecular connectivity diagram

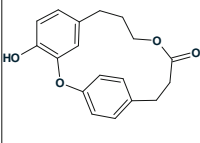
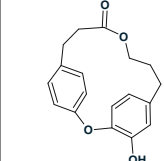
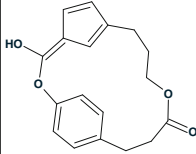
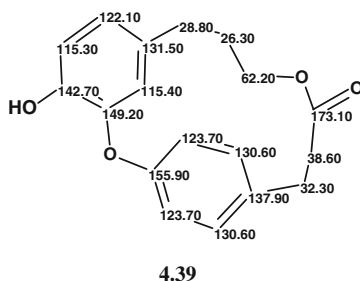
1	2	3
		
$d_A(^{13}\text{C}): 1.790$ $d_N(^{13}\text{C}): 2.453$ $d_I(^{13}\text{C}): 2.596$ $\text{max}_d(^{13}\text{C}): 5.580$	$d_A(^{13}\text{C}): 2.330$ $d_N(^{13}\text{C}): 2.499$ $d_I(^{13}\text{C}): 2.269$ $\text{max}_d(^{13}\text{C}): 8.630$	$d_A(^{13}\text{C}): 6.405$ $d_N(^{13}\text{C}): 5.852$ $d_I(^{13}\text{C}): 5.688$ $\text{max}_d(^{13}\text{C}): 51.610$

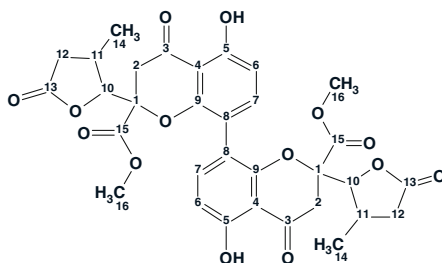
Fig. 4.49 Isocorniculatolide A: Ranked structures of the output file



The ^{13}C chemical shift assignment is identical to that performed by the authors [23].

4.21 Paecilin A

Endophytic fungi have been proven to be a rich source of new biologically active natural products because as a group they represent a relatively untapped ecological environment, and their secondary metabolism is particularly active because of their metabolic interactions with their hosts. Guo et al. [24] have isolated a series of new compounds from endophytic fungi. Particularly, they obtained a novel dimer metabolite having a *new carbon skeleton*, named Paecilin A and its monomer Paecilin B. The spectral data of Paecilin A (**4.40**) were used for the structure elucidation with the aid of Structure Elucidator.



4.40

Paecilin A was obtained as a pale yellow gum whose molecular formula $C_{32}H_{30}O_{14}$ was determined by HREIMS at m/z D 638.1632 (calcd. 638.1630) along with the analysis of the NMR data. The UV spectrum showed bands at 214, 281, and 348 nm while the IR spectrum displayed bands at 3422, 1791, 1739, and 1650 cm^{-1} which can be used for confirmation of the presence of hydroxyl, γ -lactone carbonyl, ester, and conjugated ketone groups, respectively. The molecular formula $C_{32}H_{30}O_{14}$ indicated 18 degrees of unsaturation. Inspection of the ^{13}C NMR spectrum revealed that the spectrum contained 16 double intensity signals, which indicates the presence of two identical parts of a molecule, i.e., the molecule is symmetric. Here we will demonstrate how the Structure Elucidator deals with symmetric structures. It has been shown [25] that the structure generation of symmetric molecules from 2D NMR data is a nontrivial problem and required enhancements to the structure generation algorithm in Structure Elucidator.

The 1D, HSQC, and HMBC NMR data utilized in [24] for the structure elucidation of Paecilin A are presented in Table 4.21, and the corresponding MCD partly modified by the user is displayed in Fig. 4.50.

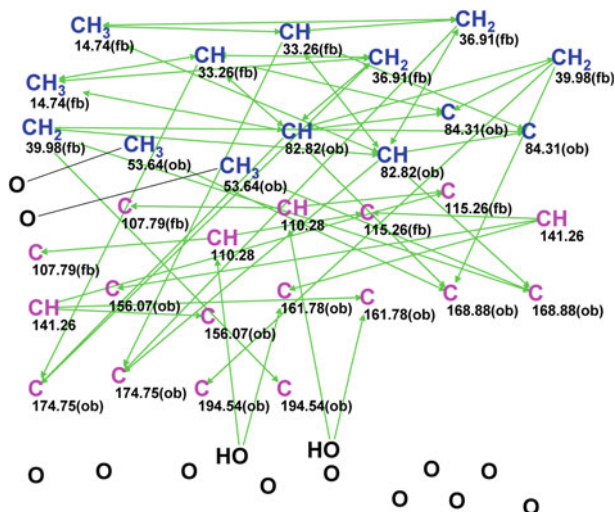
MCD overview It was suggested that the sp^2 -hybridized carbon atoms, characterized with ^{13}C chemical shifts lying in the interval of 156–195 ppm, should be oxygenated and they were marked with the label sp^2/ob . It was also assumed that acetal and ketal functionalities were absent from the molecule, hence the light blue atoms C 107.79–115.26 were marked with the label sp^2/fb . Two light blue carbon atoms C 84.31 were marked with the sp^3/ob label. Two CH_3 (C 53.64) groups were connected to oxygen atoms in accordance with their ^{13}C and ^1H chemical shifts. The number of hydrogen atoms attached to the neighboring carbon atoms were added to the atom properties on the MCD in accordance with the multiplicities and coupling constants presented in column M(J) of Table 4.21.

MCD checking detected no contradictions in the HMBC data therefore strict structure generation combined with ^{13}C chemical shift prediction was performed. Results: $k = 1,880 \rightarrow 45 \rightarrow 45$, $t_g = 1\text{ min } 42\text{ s}$, and the three top ranked structures of the output file are presented in Fig. 4.51.

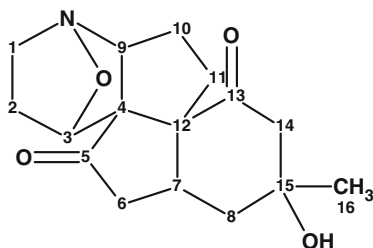
Table 4.21 Paecilin A: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M(J)	C HMBC
C1	84.31	84.24	C	–	–	–
C2	39.98	44.17	CH ₂	3.21	u	–
C2	39.98	44.17	CH ₂	3.34	u	C1, C15, C3, C10
C3	194.54	201.2	C	–	–	–
C4	107.79	113.6	C	–	–	–
C5	161.78	160.84	C	–	–	–
C6	110.28	112.78	CH	6.63	d(8.5)	C4, C8
C7	141.26	133.26	CH	7.46	d(9.0)	C9, C8, C5
C8	115.26	122.58	C	–	–	–
C9	156.07	156.09	C	–	–	–
C10	82.82	82.4	CH	4.86	d(6.5)	C15, C13, C14, C1, C11, C2, C12
C11	33.26	33.93	CH	2.95	u	C10, C1, C13, C12, C14
C12	36.91	36.89	CH ₂	2.43	u	–
C12	36.91	36.89	CH ₂	2.74	d(7.0)	C10, C11, C13, C14
C13	174.75	174.4	C	–	–	–
C14	14.74	14.61	CH ₃	1.3	d(7.0)	C10, C11, C12
C15	168.88	169.68	C	–	–	–
C16	53.64	53.37	CH ₃	3.78	s	C15
O1	100 ^a	–	OH	11.57	s	C6, C5, C4

^a Fictitious ¹⁷O NMR chemical shift

**Fig. 4.50** Paecilin A: Modified MCD

detected in Lycopodium alkaloids. Furthermore, the nitrogen atom in compound **4.42** was attached to C3 through an oxygen atom to form a 1-aza-7-oxabicyclo [2.2.1]heptane moiety, which was *first reported in natural products*. The structure was elucidated by spectroscopic methods and X-ray diffraction analysis.



4.42

The molecular formula, $C_{16}H_{21}NO_4$, was obtained from the $[M+H]^+$, $[M+Na]^+$, and $[M+K]^+$ ions at m/z 292.1545 (calculated 292.1543), 314.1362 (calculated 314.1363), and 330.1099 (calculated 330.1102), respectively, in HRESIMS analysis, requiring seven degrees of unsaturation. The IR spectrum showed absorptions for OH or NH and carbonyl groups, respectively, at 3371, 1736, and 1697 cm^{-1} .

The 1H and ^{13}C NMR chemical shifts extracted from the HSQC data were tabulated in reference [6], while the key COSY and HMBC correlations were graphically depicted on structure **4.42**. The combined NMR spectroscopic data are presented in Table 4.22.

From a spectroscopic point of view, the structure of this unusual alkaloid contains two quaternary carbon atoms—C(75.00) and C(66.60)—whose chemical shifts are characteristic for carbon atoms having a neighboring heteroatom in the first environment sphere rather than for atoms with a pure carbon environment surrounding it [11]. In addition, a methine carbon chemical shift of 78.9 pm (δH 4.9 ppm) is more typical for a carbon atom containing an oxygen in the first sphere. It is expected that all possible environments must be tried for the mentioned carbon atoms if the structure elucidation is performed manually.

The MCD shown in Fig. 4.52 was created by the program. The option “Allow sp carbons” was switched off as in the IR spectrum no features associated with potential triple bonds were observed.

MCD overview Figure 4.52 shows that carbon atoms with chemical shifts of 66.6, 75.0, 77.5, and 88.00 are colored in a light blue color indicating that their hybridization states can be either sp^3 or sp^2 while the possibility of neighboring with heteroatoms is undefined.

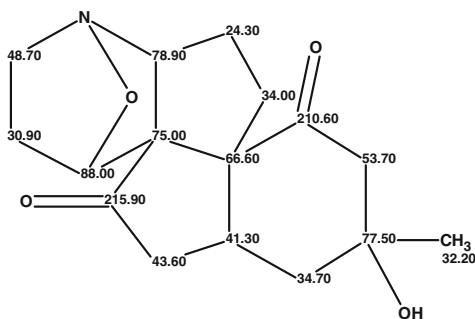
For the first run, bonds between heteroatoms were forbidden as such bonds are relatively rare in natural products. Experience shows that when chemical bonds between heteroatoms are allowed, then the number of generated structures increases (sometimes dramatically). Therefore, the first attempt to solve a problem is usually made under the supposition that bonds between heteroatoms are absent. The following results were obtained for the first run: $k = 7,144 \rightarrow 3,307 \rightarrow 2,889$, $t_g = 4.3$ s.

^{13}C chemical shift calculation and structure ranking revealed that the “best” structure was characterized by large deviation values of 5–6 ppm, which suggests, based on experience, that the solution is most probably wrong.

The following reasons may result in an incorrect solution: (a) the presence of nonstandard correlations in the 2D NMR data, (b) using wrong suggestions (“axioms”) for the atom properties, (c) the correct structure did not pass spectral or structural filtering, (d) suggestions regarding the absence of chemical bonds between heteroatoms are incorrect. In our case, 2D NMR data were presented in the article [6] as arrows on structure **4.42** and all of them were of a standard length. No changes to the MCD were made by the user so wrong suggestions regarding the atom properties are excluded. Rejection of the correct structure by filtering happens very rarely as the filter library was adjusted to include the chemical shifts of ca. 200,000 compounds collected in the ACD/NMR DB. Therefore, a second attempt at structure generation was made using options where chemical bonds between heteroatoms are allowed. ^{13}C chemical shift prediction and $d < 4$ ppm as a threshold was used during structure generation. The results gave: $k = 16,246 \rightarrow 1$, $t_g = 26$ s. A single structure identical to **4.42** was obtained with a deviation value of 3.8–3.9 ppm.

Taking into account that the average deviation for empirical methods of ^{13}C chemical shift prediction is ca. 1.8 ppm, the deviations calculated for structure **4.42** are relatively large, which can be accounted for the very unusual structure of the new natural product. Repeated structure generation with a switched off filter confirmed that structure **4.42** has the smallest deviations and is indeed the best one.

To ensure that the structure identified is correct structure, generation was repeated with a switched off Atomic Properties Correlation Table (APCT). This table is used not only for automated setting of the atom properties during MCD creation but for preventing incorrect structures (those that contain carbons with improbable environments) in the process of structure generation. The following result was obtained: $k = 2,011,349 \rightarrow 1,941,788 \rightarrow 1$, $t_g = 34$ min and the single output structure agreed with structure **4.42**. This example clearly shows the high efficiency of utilizing APCT during structure generation—it turned out that the t_g value was ca. 4,000 times larger and the number of generated structures—125 times more, when the APCT table was switched off. The Lycojaponicum A structure **4.43** with assigned ^{13}C chemical shifts is shown below:



4.43

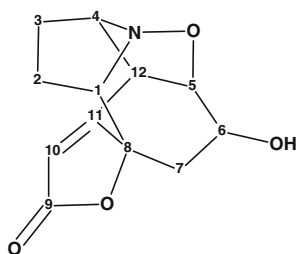
It was interesting to see what would happen if we suggested that carbons C(75.00) and C(66.60) have a heteroatom as a neighbor. When the property sp^3/ob was set for the atom C(75.00) structure generation accompanied with ^{13}C chemical shift prediction gave the following result: $k = 6,399 \rightarrow 0$, $t_g = 23$ s. A similar result was obtained for C(66.6): $k = 9,798 \rightarrow 0$, $t_g = 33$ s. Thus, both wrong hypotheses (“axioms”) were immediately rejected by the program. Note that it would take a lot of time to check these hypotheses if a manual approach to structure elucidation was used.

In conclusion, the structure of a novel alkaloid with an unprecedented structure was elucidated by the system almost automatically. No user influence was needed. A hint to set more freedom in the options for structure generation, when the existence of chemical bonds between heteroatoms is allowed, was given by the system itself.

4.23 Virosaine

Zhao and co-workers [26] isolated two new *Securinega* alkaloids with an *unprecedented skeleton*, virosaines A and B from the twigs and leaves of *Flueggea virosa* and elucidated their structures.

We used the spectroscopic data presented in [26] to establish the structure of Virosaine A (4.44) which was confirmed by X-ray diffraction in [26].



4.44

The molecular formula of **4.44** was established as $C_{12}H_{13}NO_4$ by its HRESIMS (m/z 258.0736 $[M+Na]^+$, calcd for $C_{12}H_{13}NO_4Na$: 258.0737). The authors [26] proceed with the following considerations: “The UV absorption maximum at 238 nm and IR bands at 3422, 1725, 1647 cm^{-1} implied the presence of an α,β -unsaturated γ -lactone ring and a hydroxyl group. The NMR spectra revealed that **4.44** possessed 12 carbons including an α,β -unsaturated γ -lactone ring [δH 5.84 (1H, br s); δC 175.5, 171.9, 110.8, and 85.4]. The spectral data suggested that **4.44** could be a norsecurinine-type alkaloid”. We think that such strict conclusions can be made by chemists only if they work with compounds common to a given laboratory. Following the methodology used in this book, we will try to elucidate the structure of the unknown with an “ab initio” approach.

^{13}C and 1H chemical shifts assigned to atoms of the molecule using HSQC were tabulated in the article [26], while only key 1H - 1H COSY and HMBC correlations were displayed graphically on the structure. The data which were utilized for computer-assisted structure elucidation are collected in Table 4.23 and the MCD created by the program is shown in Fig. 4.53.

MCD overview Four methine groups characterized by ^{13}C chemical shifts in the range of 50.20–74.1 ppm have 1H chemical shifts falling in the chemical shift interval 3.90–4.30 ppm. Combining these data we can suggest that all carbon atoms have heteroatoms (O or N) as neighbors. Because the IR spectrum clearly indicates the presence of a carbonyl group (1,725 cm^{-1}) it would be rational to assign C 175.50 as belonging to C=O group, however, the C=O bond was not drawn by hand and only the “ob” label was ascribed to this carbon. The chemical shift $\delta C = 171.9$ ppm is also rather typical of an ester or amide group but at this stage we will leave this atom without any “ob” label (the consequences of this decision will be considered later). The properties of the three “light blue” carbon atoms

Table 4.23 Virosaine A: Spectroscopic data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	74.1	64.6	CH	3.88	u	1.2	C7, C11, C3
C2	22.1	26.12	CH ₂	1.2	u	3.88, 1.48	–
C2	22.1	26.12	CH ₂	1.7	u	–	–
C3	20.8	27.88	CH ₂	1.48	u	1.20, 4.01	–
C3	20.8	27.88	CH ₂	1.91	u	–	–
C4	70.9	70.65	CH	4.01	u	3.89, 1.48	C5, C11
C5	88	89.76	CH	4.67	u	4.29, 3.89	
C6	65.4	68.89	CH	4.29	u	4.67, 1.83	C8
C7	44.7	39.32	CH ₂	2.92	u	–	–
C7	44.7	39.32	CH ₂	1.83	u	4.29	–
C8	85.4	90.33	C	–	–	–	–
C9	175.5	172.98	C	–	–	–	–
C10	110.8	111.81	CH	5.84	u	–	–
C11	171.9	172.04	C	–	–	–	–
C12	50.2	47.15	CH	3.89	u	4.01, 4.67	C3, C10

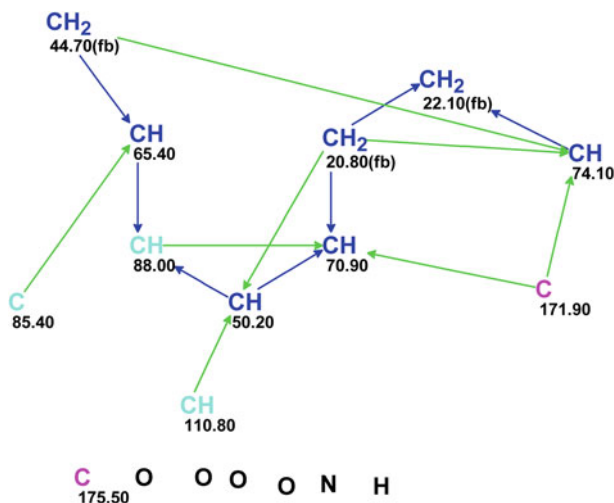


Fig. 4.53 Virosaine A: Molecular connectivity diagram

C 85.40–C 110.80 (sp^2 or sp^3) were not edited, as both carbon double bond C=C and O–C–O or O–C–N atomic groupings are possible in the molecule according to the molecular formula. The IR absorption band at $3,422\text{ cm}^{-1}$ may belong to stretching vibrations of both OH and NH groups so neither of these groups could be drawn by hand on the MCD. The MCD edited in accordance with the above-mentioned considerations is shown in Fig. 4.54.

MCD checking did not revealed the presence of contradictions in the 2D NMR data and structure generation was run. Because chemical bonds between heteroatoms are observed relatively rarely in natural products, then to prevent generation of a huge number of nonrealistic structures these bonds are usually marked in Structure Generation Options as not allowed for the first run. This option was used from the very beginning. ^{13}C chemical shift calculations during structure generation were performed with the threshold parameters $d_l > 4\text{ ppm}$, $d(\text{max}) > 20\text{ ppm}$: all structures which meet this criterion are rejected. Results: $k = 2,756 \rightarrow 0$, $t_g = 3\text{ s}$.

Experience showed that in such a situation the failure most frequently can be accounted for the presence of nonstandard correlations that were not detected during the MCD checking. Therefore, the next step was FSG with the initial parameters $m = 1$, $a = 16$, that is we introduce the possibility that at least one NSC of unknown length exists in the 2D NMR data. Results: $k = 75,790 \rightarrow 0$, $t_g = 2\text{ min } 34\text{ s}$, 15 from 15 possible connectivity combinations used during generation. The repeated failure is a hint to an attempt of changing options governing possibility of chemical bonds between heteroatoms. The next program run was initiated with an option that allowed chemical bonds between heteroatoms of different kinds as shown below in the dialog window of the structure generator (Fig. 4.55):

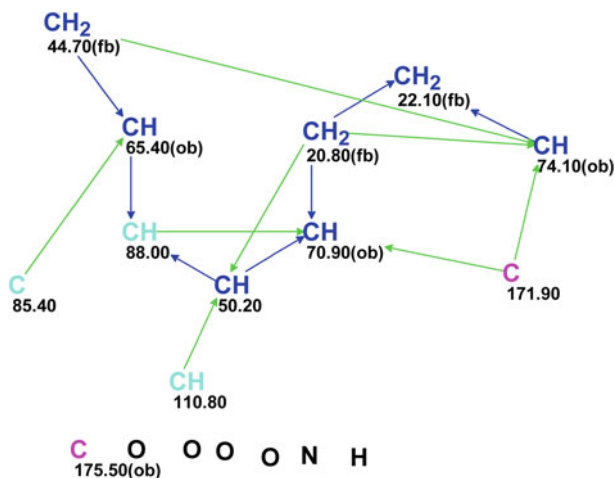


Fig. 4.54 Virosaine A: Edited MCD

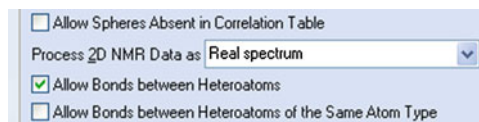


Fig. 4.55 Virosaine A: A part of the Dialog Window where chemical bonds between heteroatoms are allowed

Results of the strict structure generation: $k = 5,667 \rightarrow 3 \rightarrow 3$, $t_g = 11$ s, and the best structure selected by HOSE and incremental chemical shift predictions ($d_A = 3.76$, $d_I = 3.13$ ppm) coincided with structure of Virosaine A published in the article [26]. The ranked output file is shown in Fig. 4.56 and structure 4.45 displays the ^{13}C chemical shift assignments.

The values of deviations are relatively large, which are explained by the unprecedented structure of the molecule, but the values of the maximum deviations also support the priority of #1 for the structure. Nevertheless, the priority is not fully convincing—the difference $d(\#2) - d(\#1)$ is small, and this happens when CASE results should be confirmed by additional experiments (recollect that the structure of Virosaine A was confirmed by X-ray crystallography in [26]).

Structure 4.45 shows that carbon C 171.9 belonging to a carbon double bond $(\text{C})_2\text{C}=\text{C}(\text{C}_2)$ has an unusual ^{13}C chemical shift. It was interesting to see what will happen if this atom was supplied with an “ob” label which forces the atom to form $-\text{C}=\text{O}$ or $\text{C}=\text{N}$ fragments. Such a suggestion is acceptable from the point of view of an unbiased expert who performs the structure elucidation “ab initio”.

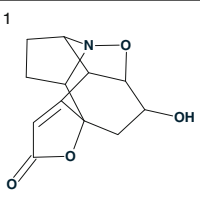
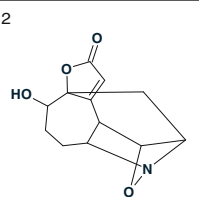
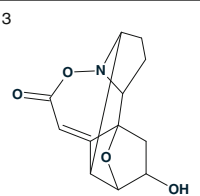
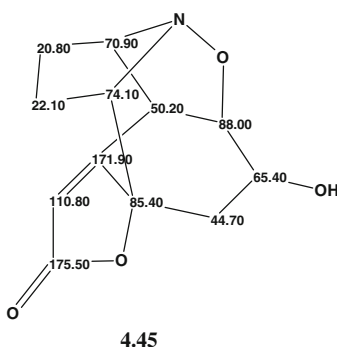
1	2	3
		
$d_A(^{13}\text{C}): 3.763$ $d_N(^{13}\text{C}): 3.594$ $d_l(^{13}\text{C}): 3.132$ $\text{max_}d_A(^{13}\text{C}): 9.670$	$d_A(^{13}\text{C}): 4.607$ $d_N(^{13}\text{C}): 3.503$ $d_l(^{13}\text{C}): 3.599$ $\text{max_}d_A(^{13}\text{C}): 15.210$	$d_A(^{13}\text{C}): 4.724$ $d_N(^{13}\text{C}): 5.086$ $d_l(^{13}\text{C}): 3.930$ $\text{max_}d_A(^{13}\text{C}): 10.070$

Fig. 4.56 Virosaine A: The ranked output structure file



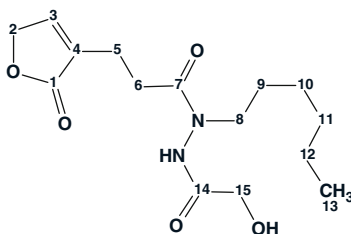
The next (and last) run of structure generation was repeated from the MCD where carbon C 171.9 was labeled as sp^2/ob . Results: $k = 4,593 \rightarrow 0$, $t_g = 8$ s. We see that the wrong hypothesis was immediately refused by the program.

4.24 Geralcin A

Hydrazines and hydrazides are structural motifs often found in synthetic therapeutics that are in clinical use. An important example is iproniazid, a hydrazine monoamine oxidase inhibitor, used as an antidepressant. In contrast, natural hydrazide compounds are notably scarce and only four such structures have been reported to date.

Le Goff et al. [27] reported the structural characterization of two novel alkylhydrazides produced by the bacterial strain *Streptomyces* sp. LMA-545. The structures were elucidated using both 1D and 2D ^1H and ^{13}C NMR spectroscopic analysis and high-resolution mass spectrometry. ^1H - ^{15}N NMR experiments were required for full structural elucidation. Here we will describe the structure

elucidation of Gercalcin A (**4.46**), one of the three isolated novel compounds. This compound consists of a *novel natural scaffold* that connects an alkylhydrazide to an α,β -unsaturated γ -lactone.



4.46

Compound **4.46** was obtained as yellowish oil. The HRESIMS analysis gave the molecular formula $C_{15}H_{24}N_2O_5$: m/z $[M+Na]^+$ 335.1575 (calculated for $C_{15}H_{24}N_2O_5Na$, 335.1583). IR absorption bands observed at 3,430 and 3,275 cm^{-1} indicate the presence of both NH and OH functionalities. An absorption band at 1,741 cm^{-1} is characteristic of a carbonyl group. The NMR data acquired for the structure elucidation are listed in Table 4.24.

To assist in the structure elucidation of this unknown the authors [27] also acquired an 1H - ^{15}N HMBC spectrum revealing two nitrogen atoms at δ_N 133.5 (N 1, **N**) and 137.0 (N 2, **NH**). Table 4.24 shows that the proton at δ 10.44 attached to a nitrogen atom produces correlations to carbons C15, C14, and C7, and to N 1. Three correlations to nitrogen atoms are presented in the column N HMBC. The correlations of H-N (δ_H 10.44, s, **NH**) to N1 (δ_N 133.5, **N**) and (δ_H 3.58, m) to N 1 (δ_N 133.5, **N**) allow us to suggest that a N-N bond corresponding to a hydrazide group was present. The spectrum-structure information contained in Table 4.24 is presented graphically on the MCD (Fig. 4.57).

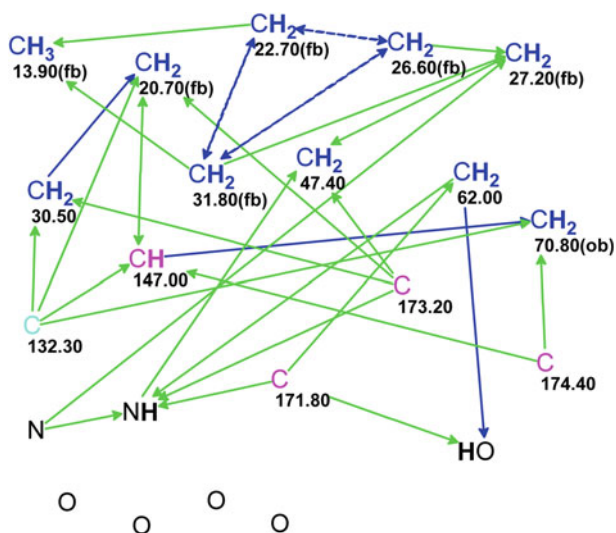
MCD overview The authors [27] declare that there are three carbonyl atoms at δ_C 174.4 (C1), 173.2 (C7), and 171.8 (C14), but the properties of these atoms were not edited on the MCD (i.e., C=O bonds were not drawn by hand) because the same ^{13}C chemical shifts can be characteristic also for sp^2 -hybridized carbons connected to an oxygen by an ordinary bond (C=C-O- or N=C-O). Ambiguous (dotted lines) COSY correlations are due to the overlapping three 1H resonances at 1.28 ppm related to the protons attached to atoms C10, C11, and C12. The numbers of hydrogen atoms attached to neighboring skeletal atoms were set on the MCD in accordance with the multiplicities listed in Table 4.24, column M(J). Note that the multiplicities of the 1H signals produced by the NH and OH groups were also taken into account.

First run For the first run, it was assumed that there were no hints as to the existence of an N-N chemical bond in the molecule. MCD checking under the condition that chemical bonds between heteroatoms are forbidden revealed contradictions in the 2D NMR data. FSG run with the checkbox “Determine Options

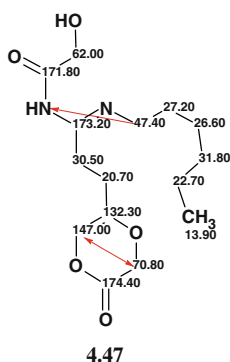
Table 4.24 Geralcin A: Spectroscopic NMR data

Label	δ_X	δC_{calc}	XH _n	δH	M(J)	COSY	C HMBC	N HMBC
C1	174.4	173.71	C	–	–	–	–	–
C2	70.8	70.08	CH ₂	4.85	u	7.48	C1, C3, C4	–
C3	147	146	CH	7.48	u	4.85	C2, C5, C1, C4	–
C4	132.3	132.86	C	–	–	–	–	–
C5	20.7	20.18	CH ₂	2.61	u	2.47	C7, C4, C6, C3	–
C6	30.5	30.7	CH ₂	2.47	u	2.61	C5, C4, C7	–
C7	173.2	171.63	C	–	–	–	–	–
C8	47.4	49.09	CH ₂	3.58	u	–	C9, C7	N2
C9	27.2	26.21	CH ₂	1.52	u	–	C8, C11, C10	N1
C10	26.6	26.92	CH ₂	1.28	u	–	–	–
C11	31.8	31.43	CH ₂	1.28	u	–	–	–
C12	22.7	22.56	CH ₂	1.28	u	–	–	–
C13	13.9	13.95	CH ₃	0.86	t(6.9)	–	C12, C11	–
C14	171.8	171.18	C	–	–	–	–	–
C15	62	62.18	CH ₂	4.14	d(5.9)	5.79	C14	–
N1	133.5	139.53	N	–	–	–	–	–
N2	137	129.66	NH	10.44	s	–	C15, C14, C7	N1
O1	100 ^a	–	OH	5.79	t(5.9)	4.14	C15, C14	–

^a Fictitious ¹⁷O chemical shift

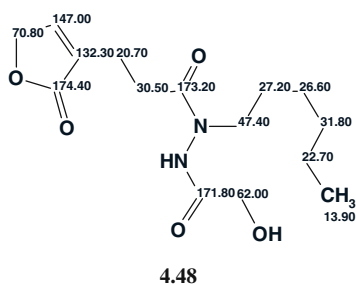
**Fig. 4.57** Geralcin A: Molecular connectivity diagram

Automatically” was selected. The results gave: $k = 1$, $t_g = 1.5$ s, (see structure **4.47**), 2 from 26 connectivities have been extended and 325 from 325 connectivity combinations were used. All deviations calculated for structure **4.47** (arrows show nonstandard connectivities) were too large to accept the solution obtained without reexamination ($d \sim 5.5$ ppm).



The results obtained hinted at the need to try structure generation with options allowing chemical bonds between heteroatoms of the same kind.

Second run The corresponding checkboxes were selected in the window **CSB Generator Options** and structure generation was carried out with the following results: $k = 1$, $t_g = 0.001$ s, all deviations are less than 1 ppm. The structure obtained, **4.48**, which coincided with the structure of Gercalcin A is shown below.



A correct solution was therefore found instantly and almost automatically using the spectral data presented in Table 4.24. For completeness of analysis, it was interesting to know if Structure Elucidator is capable of solving this problem without the need for H–N HMBC data.

Third run Structure generation was repeated with the options shown in Fig. 4.58.

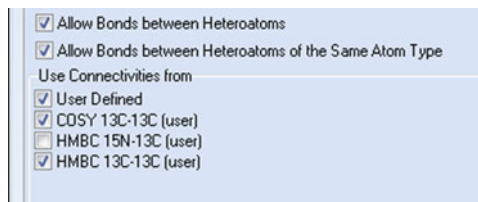


Fig. 4.58 Geralcin A: Options for structure generation with H–N NMBC data switched off

The following results were obtained: $k = 49 \rightarrow 31 \rightarrow 9$, $t_g = 0.03$ s, and the first six structures of the ranked output file are presented in Fig. 4.59.

The best structure, #1, which is characterized with very small deviations, is identical to the suggested structure for Geralcin A. The deviation values of the ^{13}C chemical shifts predicted for structure #2 are greater than those found for structure #1. Although these values are also acceptable, the priority of structure #1 is evident. Note that structure #2 which is characterized with small deviations (1.3–1.5 ppm) contains only two carbonyl groups (not three as in structure #1), and its generation became possible since the carbonyl bonds at carbons C 174.4, C 173.2, and C 171.8 were not forced by adding them manually. So, the utilization of soft constraints at the analysis stage and editing of the MCD allow consideration of all plausible structural hypotheses. This example shows that a compound containing a *novel*

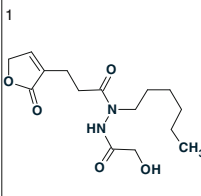
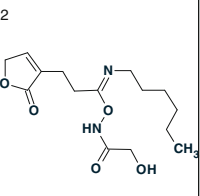
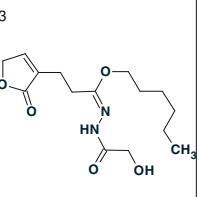
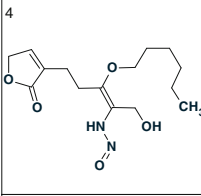
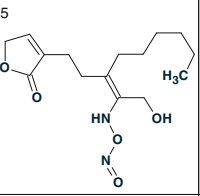
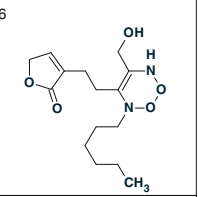
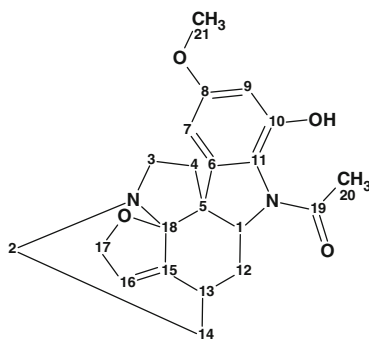
		
$d_A(^{13}\text{C}): 0.599$ $d_N(^{13}\text{C}): 0.802$ $d_I(^{13}\text{C}): 0.859$ $\text{max}_dA(^{13}\text{C}): 1.690$	$d_A(^{13}\text{C}): 1.431$ $d_N(^{13}\text{C}): 1.327$ $d_I(^{13}\text{C}): 1.478$ $\text{max}_dA(^{13}\text{C}): 6.550$	$d_A(^{13}\text{C}): 3.305$ $d_N(^{13}\text{C}): 3.168$ $d_I(^{13}\text{C}): 3.250$ $\text{max}_dA(^{13}\text{C}): 18.310$
		
$d_A(^{13}\text{C}): 7.880$ $d_N(^{13}\text{C}): 7.821$ $d_I(^{13}\text{C}): 8.497$ $\text{max}_dA(^{13}\text{C}): 46.390$	$d_A(^{13}\text{C}): 7.977$ $d_N(^{13}\text{C}): 8.047$ $d_I(^{13}\text{C}): 7.361$ $\text{max}_dA(^{13}\text{C}): 57.580$	$d_A(^{13}\text{C}): 8.239$ $d_N(^{13}\text{C}): 8.120$ $d_I(^{13}\text{C}): 8.023$ $\text{max}_dA(^{13}\text{C}): 49.320$

Fig. 4.59 Geralcin A: The first six structures of the ranked output file obtained without employing the H–N HMBC data

natural scaffold was identified with the aids of Structure Elucidator instantaneously and without having to use the ^1H - ^{15}N HMBC data.

4.25 Methoxygeissospermidine

Reina et al. [28] reported on the isolation and structure elucidation of 10 indole alkaloids extracted from the leaves and bark of *Geissospermum reticulatum* A. The structure of one of them, 10-Methoxygeissospermidine (**4.49**), was established by the analysis of 1D and 2D NMR data (COSY, HSQC, HMBC, NOESY) and confirmed by a single-crystal X-ray diffraction study.



4.49

Alkaloid **4.49** is a dihydroindole derivative with the extended fused ring system, which is a *new skeleton*. Compound **4.49** was classified as an indole alkaloid of the aspidospermatan type with a 1-oxa-3-cyclopentene group between C15 and C18, thus comprising a new subtype. HRESIMS of **4.49** provided the molecular formula $\text{C}_{21}\text{H}_{24}\text{N}_2\text{O}_4$ (M^+ 368.1736). The NMR spectroscopic data available from the work [28] and used for CASE analysis are presented in Table 4.25 (HMBC data were displayed in the article only in graphical form).

IR spectra were not presented in the article, but the authors [28] noted that absorption bands were observed at 3480, 1600, 1580 cm^{-1} , which suggests that the molecule contains OH or NH groups and a benzene ring. The absorption band at 1,630 cm^{-1} can be assigned either to conjugated carbon double bonds or to conjugated carbonyl functional groups. The MCD is presented in Fig. 4.60.

MCD overview MCD analysis shows that there is definitely a lack of structural information. Particularly, atoms CH_3 22.4, CH 28.7, OCH_3 55.6, C 147.4, and C 160.0 are “free” of associated correlations. Some carbon atoms (CH 102.9, CH 103.4, and C 106.2) have close chemical shifts and, judging by chemical shift values only, can be assigned either to sp^2 or to sp^3 carbons connected with two heteroatoms. A series of carbon atoms (CH_2 25.70, CH 28.70, CH 102.90, etc.)

Table 4.25 10-Methoxygeissospermidine: NMR spectroscopic data

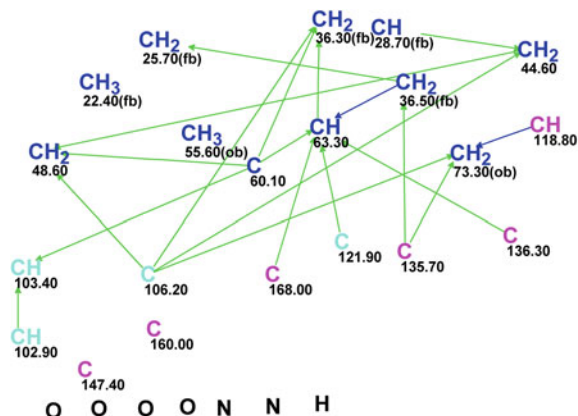
Label	δC	δC_{calc}	CH_n	δH	$M(J)^a$	COSY	HMBC (H \rightarrow C)
C1	63.3	64.62	CH	4.15	u	1.56	C11, C12, C5, C19, C6
C2	44.6	44.39	CH ₂	2.62	u	–	C13, C18
C2	44.6	44.39	CH ₂	3.06	u	–	–
C3	48.6	51.08	CH ₂	2.9	u	–	C2, C5, C18
C3	48.6	51.08	CH ₂	3.4	u	–	–
C4	36.3	36.45	CH ₂	1.95	u	–	–
C4	36.3	36.45	CH ₂	2.26	u	–	C5, C1, C18
C5	60.1	64.06	C	–	–	–	–
C6	136.3	135.36	C	–	–	–	–
C7	103.4	102.32	CH	6.91	S	–	C5, C9
C8	160.0	156.34	C	–	–	–	–
C9	102.9	98.37	CH	6.38	S	–	–
C10	147.4	147.02	C	–	–	–	–
C11	121.9	122.46	C	–	–	–	–
C12	36.5	37.37	CH ₂	2.17	u	–	–
C12	36.5	37.37	CH ₂	1.56	u	4.15	C15
C13	28.7	30.19	CH	3.06	u	–	–
C14	25.7	31.84	CH ₂	1.83	u	–	C12
C15	135.7	147.75	C	–	–	–	–
C16	118.8	121.82	CH	5.62	u	4.73	–
C17	73.3	72.54	CH ₂	4.51	u	–	–
C17	73.3	72.54	CH ₂	4.73	u	5.62	C15, C18
C18	106.2	106.91	C	–	–	–	–
C19	168.0	169.7	C	–	–	–	–
C20	22.4	23.56	CH ₃	2.4	S	–	–
C21	55.6	55.59	CH ₃	3.74	u	–	–

^a Only multiplicities which were used for the CASE are shown

have only a single connectivity to another series of carbon atoms. Hence it should be expected that structure generation will be too time-consuming and, therefore, to accelerate it, additional structural constraints (“axioms”) must be imposed by the user.

Taking into account the ¹H chemical shift values of H7 and H9 (between 6 and 7 ppm, see Table 4.25) we can suppose that the corresponding carbon atoms C7 and C9 are in the *sp*² hybridization state. There exist nine carbon atoms with chemical shifts observed in the range between 100 and 160 ppm (the interval is characteristic of *sp*² hybridized carbons forming double bonds). The total number of *sp*² carbons included in the benzene ring and in double bond (see IR data) is equal to eight, so one of the nine carbons should be assigned as a *sp*³ hybridized carbon bonded with two heteroatoms. Having this in mind, carbon C 106.20 was supplied with the property *sp*³/*2ob*. The carbon with $\delta C = 160$ ppm was allocated the property *sp*²/*1ob*

Fig. 4.60 10-Methoxygeissospermidine:
Molecular connectivity
diagram



according to [11], while the quaternary carbon resonating at 168 ppm was assigned to a carbonyl group, and the bond C=O was drawn on the MCD by hand. As CH₃ 55.6 (δ H 3.74) is obviously connected to an oxygen atom [11] the chemical bond CH₃–O was also drawn on the MCD. A signal from one exchangeable proton was observed in the ¹H NMR spectrum at 10.7 ppm, and one can conclude that this proton belongs to a hydroxyl group (N–H shows resonances in the region of 0.5–5.0 ppm [11]). Hence the O–H bond was added to the MCD. Finally, structure generation followed by ¹³C chemical shift prediction was initiated from the edited MCD shown in Fig. 4.61. To accelerate the structure generation, three- and four-membered cycles were forbidden in the filter options (see comments below).

The results are: $k = 1,344,936 \rightarrow 46 \rightarrow 33$, $t_g = 2$ h 23 min. The four top ranked structures of the output file are shown in Fig. 4.62.

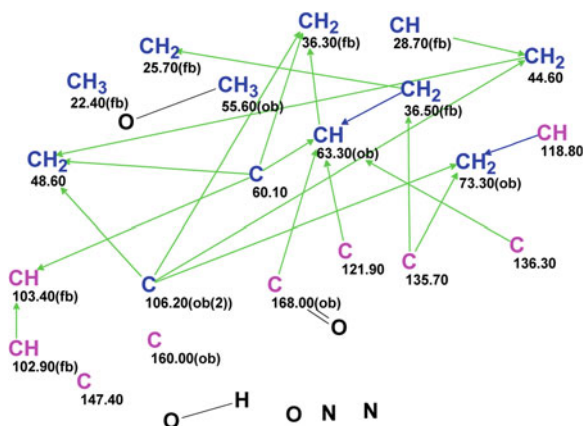


Fig. 4.61 10-Methoxygeissospermidine: Edited MCD

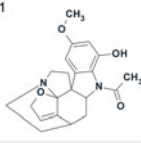
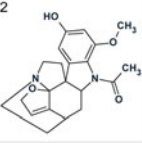
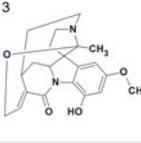
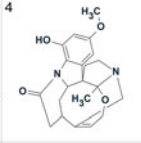
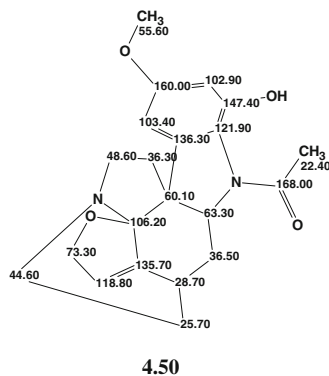
1	2	3	4
			
$d_A(^{13}\text{C}): 2.722$ $d_N(^{13}\text{C}): 2.247$ $d_I(^{13}\text{C}): 2.621$ $\text{max}_dA(^{13}\text{C}): 12.510$	$d_A(^{13}\text{C}): 3.172$ $d_N(^{13}\text{C}): 2.703$ $d_I(^{13}\text{C}): 2.638$ $\text{max}_dA(^{13}\text{C}): 12.510$	$d_A(^{13}\text{C}): 3.653$ $d_N(^{13}\text{C}): 3.420$ $d_I(^{13}\text{C}): 3.845$ $\text{max}_dA(^{13}\text{C}): 18.440$	$d_A(^{13}\text{C}): 3.710$ $d_N(^{13}\text{C}): 3.970$ $d_I(^{13}\text{C}): 3.804$ $\text{max}_dA(^{13}\text{C}): 12.810$

Fig. 4.62 10-Methoxygeissospermidine: The four top ranked structures of the output file

It is evident that the best structure #1 coincides with the structure of Methoxygeissospermidine (**4.49**). Differences between deviations calculated for structures #1 and #2 are small, which is accounted for by small differences between the corresponding structures—substituents on the benzene ring are simply exchanged. Structure **4.50** shows the Methoxygeissospermidine molecule with assigned ^{13}C chemical shifts:



Chemical shift assignment is in agreement with that suggested by the authors [28]. It is interesting to note that carbon C5 implies a ^{13}C signal at 60.1 ppm, which is characteristic for a carbon whose closest environment contains a heteroatom. The maximum deviation of the predicted ^{13}C chemical shift (12 ppm) has a place for the C15 carbon (δC 135.7), while predicted chemical shifts are rather close to experimental ones for other atoms in the molecule (see Table 4.25).

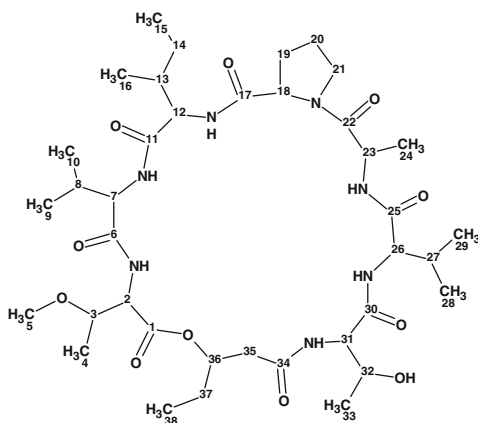
To confirm the results obtained, structure generation was repeated with an option which allowed generation of structures containing 3–4-membered cycles. The results are: $k = 2,752$, $646 \rightarrow 54 \rightarrow 41$, $t_g = 3$ h 25 min, and the same molecules that are displayed in Fig. 4.62 were chosen as preferable. Comparison of results obtained in the two runs shows that the number of generated structures and time consumed are two and half times larger when small ring cycles are allowed. This is accounted for by a specific feature of the algorithm for structure generation (see

Chap.1) which is implemented in StrucEluc: if small cycles are forbidden then the program skips the structures that could contain the mentioned cycles. Note that constraints on the cycles of larger dimensions are imposed posteriori in the process of structural filtering, i.e., structures are generated first and then checked for constraints.

4.26 Peptidolipin B

Marine-derived actinomycetes have emerged as a rich source of secondary metabolites with therapeutic relevance. Marine *Nocardia* spp., while not investigated as extensively as other marine actinomycetes, have been a source of antibacterial compounds. Among the antibacterial compounds derived from terrestrial isolates of *Nocardia* spp. is the lipopeptide Peptidolipin NA. Peptidolipin NA belongs to a rare class of lipopeptides that are characterized by a peptide cyclized via an ester to a lipophilic tail.

Wyche et al. [29] isolated five new lipopeptides, peptidolipins B–F from a marine *Nocardia* sp. Peptidolipins B–F were deemed similar to the L-Val(6) analog of peptidolipin NA. Each peptide contained a lipid chain. One of them, Peptidolipin B (**4.51**), was used to challenge the StrucEluc software.



4.51

The molecular formula $C_{59}H_{107}N_7O_{11}$ (HREIMS $[M+H]^+$ m/z 1090.8087, calculated 1090.8101) was determined for Peptidolipin B. An infrared absorption band observed at $1,742\text{ cm}^{-1}$ can be considered as evidence for the presence of an ester group. The absorption maximum at 213 nm in a UV spectrum indicates the absence of conjugated bonds in the molecule. The authors [29] noted a peak corresponding to 18 CH_2 groups in the region 1.2–1.4 ppm of the 1H NMR spectrum belonging to

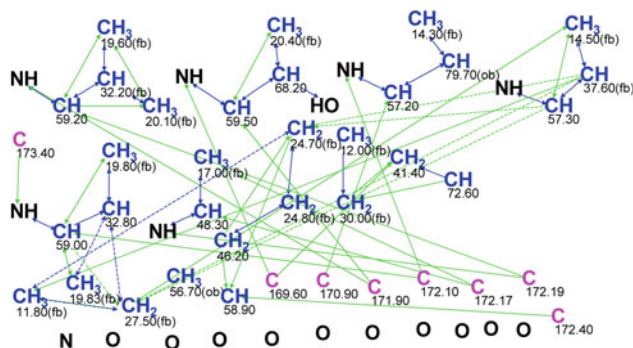


Fig. 4.63 Peptidolipin B: Initial MCD

the lipid chain. To avoid the generation of this obvious chain by the program, the chain was replaced by an isopropyl group (atoms 8–10) in the *truncated* structure of Peptidolipin B (**4.51**) and the molecular formula was reduced to $C_{38}H_{65}N_7O_{11}$.

The NMR data used for the computer-assisted structure elucidation are presented in Table 4.26, and the MCD created from these data is shown in Fig. 4.63. The data include 46 COSY and 51 HMBC correlations inferred from the 2D NMR spectra.

MCD overview The number of connectivities is large enough but the great number of heteroatoms (18 in total) and their diversity (7 nitrogens and 11 oxygens) allow for a huge number of carbon–heteroatom combinations during structure generation. For instance, according to known spectrum–structure correlations, all carbons with ^{13}C chemical shifts lying in the range of 169–173 ppm can be connected by double bonds either to oxygen or to nitrogen atoms. Having in mind a priori that the unknown compound is a peptide that allows us to draw double bonds from all relevant carbon atoms to oxygen atoms. A property “*ob*” can be assigned to all carbon atoms with chemical shifts in the range of 46.2–173.4 ppm, which follows from the characteristic ^{13}C and 1H NMR chemical shifts (see the corresponding δH values in Table 4.26). The MCD was edited in accordance with the considerations given above. The edited MCD is presented in Fig. 4.64.

MCD checking showed that all 2D NMR data are consistent and strict structure generation accompanied with ^{13}C chemical shift prediction was initiated. The results are: $k = 5,117 \rightarrow 482 \rightarrow 458$, $t_g = 20$ s. The top ranked structures of the output file are presented in Fig. 4.65.

The best structure #1 selected by HOSE code-based prediction differs from the truncated structure **4.51** of Peptidolipin B while structure #2 coincides with it. We see that d_A deviations calculated for both structures are almost the same, but that the average deviations found by other methods for ^{13}C and 1H chemical shifts “vote” for the priority of structure #2. All of the top ranked structures are very similar and the differences between deviations of the first and next structures are small. In such a case additional confirmation of structure #2 is necessary. The truncated structure of Peptidolipin B (**4.52**) supplied with the ^{13}C chemical shift assignments is shown below.

Table 4.26 Peptidolipin B: Spectroscopic NMR data

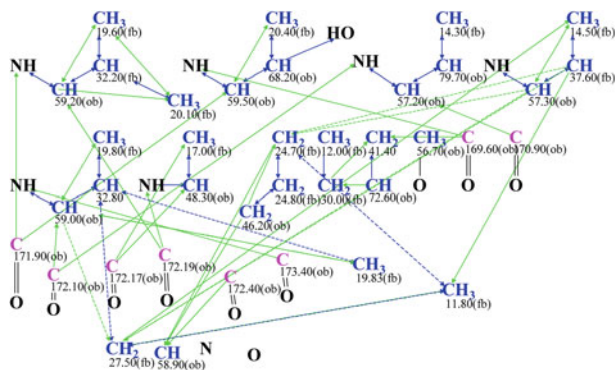
Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C1	170.9	170.05	C	–	–	–	–
C2	57.2	57.68	CH	5.35	u	4.16, 9.46	C1
C3	79.7	76.56	CH	4.16	u	1.09, 5.35	–
C4	14.3	17.28	CH ₃	1.09	d(6.0)	4.16	–
C5	56.7	58.25	CH ₃	3.38	S	–	–
C6	172.1	171.17	C	–	–	–	–
C7	59	58.69	CH	5.55	u	2.48, 9.31	C6, C9, C8, C10
C8	32.8	30.82	CH	2.48	u	1.28, 1.31, 1.31, 5.55	–
C9	19.83	19.43	CH ₃	1.31	u	2.48	–
C10	19.8	18.71	CH ₃	1.28	u	2.48	C7
C11	173.4	173.16	C	–	–	–	–
C12	57.3	58.47	CH	5.07	u	2.13, 8.31	C16
C13	37.6	37.3	CH	2.13	u	1.15, 5.07	–
C14	27.5	25.91	CH ₂	1.54	u	0.8	C13, C12
C14	27.5	25.91	CH ₂	1.31	u	2.48	C7
C15	11.8	11.75	CH ₃	0.8	t(7.5)	1.54, 1.54	C14, C13
C16	14.5	15.58	CH ₃	1.15	d(6.8)	2.13, 8.31	C12, C13, C14
C17	172.4	171.83	C	–	–	–	–
C18	58.9	63.18	CH	5.05	u	–	C19, C21, C17
C19	24.7	29.71	CH ₂	2.52	u	2.01	C21
C19	24.7	29.71	CH ₂	1.54	u	0.8	–
C20	24.8	25.76	CH ₂	1.7	u	–	–
C20	24.8	25.76	CH ₂	2.01	u	2.52, 3.45	–
C21	46.2	47.19	CH ₂	3.4	u	–	–
C21	46.2	47.19	CH ₂	3.45	u	2.01	C20
C22	172.17	170.76	C	–	–	–	–
C23	48.3	46.86	CH	5.04	u	1.68, 8.85	C22, C24
C24	17	17.08	CH ₃	1.68	d(6.4)	5.04	C23, C22
C25	172.19	171.73	C	–	–	–	–
C26	59.2	58.14	CH	5.25	u	2.31, 10.15	C28, C25, C27
C27	32.2	30.39	CH	2.31	u	1.19, 1.24, 5.25	C26, C28
C28	19.6	18.97	CH ₃	1.19	d(6.8)	2.31	C29, C26, C27
C29	20.1	19.56	CH ₃	1.24	u	2.31	C27, C26, C28
C30	171.9	172.13	C	–	–	–	–
C31	59.5	62.71	CH	5.38	u	4.45, 7.72	C30, C32, C33
C32	68.2	67.49	CH	4.45	u	1.46, 5.38, 6.79	–
C33	20.4	20.09	CH ₃	1.46	d(6.8)	4.45	C31, C32
C34	169.6	170.95	C	–	–	–	–
C35	41.4	40.64	CH ₂	2.59	u	–	–
C35	41.4	40.64	CH ₂	2.89	u	5.4	C36, C34, C37

(continued)

Table 4.26 (continued)

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C36	72.6	73.45	CH	5.4	u	2.89	–
C37	30	26.14	CH ₂	1.87	u	0.9	C38, C36
C37	30	26.14	CH ₂	1.82	u	–	–
C38	12	9.6	CH ₃	0.9	u	1.87	–
N1	100 ^a	–	NH	9.46	u	5.35	C6
N2	110 ^a	–	NH	9.31	u	5.55	C11
N3	120 ^a	–	NH	8.31	u	5.07	–
N4	130 ^a	–	NH	8.85	u	5.04	C25, C23
N5	140 ^a	–	NH	10.15	u	5.25	C30
N6	150 ^a	–	NH	7.72	u	5.38	C34
O1	200 ^a	–	OH	6.79	u	4.45	–

^a Fictitious ¹⁵N and ¹⁷O chemical shifts

**Fig. 4.64** Peptidolipin B: Edited MCD

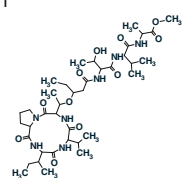
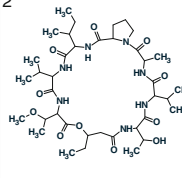
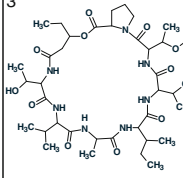
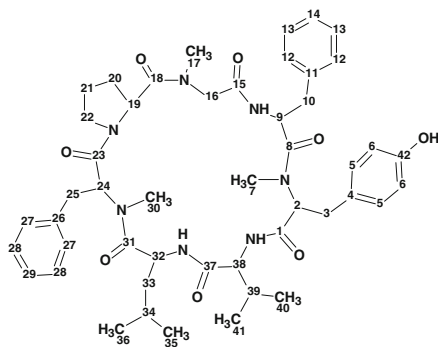
		
$d_A(^{13}\text{C})$: 1.327 $d_N(^{13}\text{C})$: 1.516 $d_I(^{13}\text{C})$: 1.396 $\text{max}_dA(^{13}\text{C})$: 6.840 $d_A(^1\text{H})$: 0.371 $d_N(^1\text{H})$: 0.371 $d_I(^1\text{H})$: 0.339	$d_A(^{13}\text{C})$: 1.344 $d_N(^{13}\text{C})$: 1.258 $d_I(^{13}\text{C})$: 1.058 $\text{max}_dA(^{13}\text{C})$: 5.010 $d_A(^1\text{H})$: 0.320 $d_N(^1\text{H})$: 0.331 $d_I(^1\text{H})$: 0.312	$d_A(^{13}\text{C})$: 1.352 $d_N(^{13}\text{C})$: 1.276 $d_I(^{13}\text{C})$: 1.183 $\text{max}_dA(^{13}\text{C})$: 6.430 $d_A(^1\text{H})$: 0.319 $d_N(^1\text{H})$: 0.328 $d_I(^1\text{H})$: 0.316

Fig. 4.65 Peptidolipin B: The top ranked structures of the output file

and bioactive metabolites of fungi derived from the South China Sea. They obtained a fermentation extract of strain SCSIO 115, identified as *Acremonium persicinum*, showing strong lethality against brine shrimp and cytotoxicity toward a panel of tumor cell lines.

Chemical investigation of the *A. persicinum* SCSIO 115 extract led to the identification of three new cycloheptapeptides, designated cordyheptapeptides C–E. Here we consider the structure elucidation of Cordyheptapeptide C (**4.54**).



4.54

Compound **4.54** was isolated as colorless crystals. Its molecular formula, $C_{48}H_{63}N_7O_8$, was established using HRESIMS, indicating 21 degrees of unsaturation. The 1H and ^{13}C NMR spectra of **4.54** exhibited characteristic signals for a heptapeptide. The authors [30] deduced the following considerations based on spectral data analysis. “Seven carbonyl resonances at δC 174.1, 172.3, 170.9, 170.5, 170.4, 168.4, and 168.2, together with the seven α -amino acid carbon resonances between δC 69.2 and 47.6 in the ^{13}C NMR spectrum, indicated the presence of seven amino acid residues. The 1H NMR spectrum of **4.54** displayed seven methyl signals in the upfield region, three of which were assigned as N–Me groups at δH 3.03, 2.61, and 2.91. Comprehensive analysis of the 1D (1H and ^{13}C , DEPT) and 2D (COSY, HMQC, and HMBC) NMR spectroscopic data revealed that **4.54** was a heptapeptide.”

Only ^{13}C , 1H , and HSQC spectroscopic data were tabulated in the article [30], while only key HMBC and COSY correlations were presented graphically on the structure **4.54**. Figure 4.66 shows that COSY and HMBC correlations produce an almost continuous chain around the molecular skeleton, which eases the structure elucidation process.

All available spectrum-structure information is summarized in Table 4.27, and the MCD is presented in Fig. 4.67.

MCD overview In the right lower part of the MCD, a series of ambiguous COSY and HMBC correlations are displayed (dotted lines). The cause of ambiguity is the presence of overlapping signals in the ^{13}C NMR spectrum: four CH(128.6)

Table 4.27 Cordyheptapeptide C: Spectroscopic NMR data

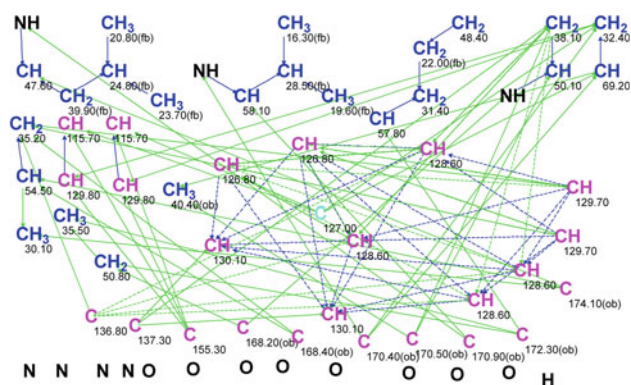
Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	C HMBC
C1	170.40	170.130	C	–	–	–	–
C2	69.20	55.680	CH	3.400	u	3.16	C1, C4
C3	32.40	32.390	CH ₂	3.160	u	3.40	C1, C4, C5
C4	127.00	127.430	C	–	–	–	–
C5	129.80	130.100	CH	6.220	u	6.53	C3, C42
C6	115.70	115.000	CH	6.530	u	6.22	C42, C4
C7	40.40	30.050	CH ₃	2.610	u	–	C2, C8
C8	170.50	171.010	C	–	–	–	–
C9	50.10	50.220	CH	5.360	u	8.61, 3.05	C8
C10	38.10	37.470	CH ₂	2.820	u	–	–
C10	38.10	37.470	CH ₂	3.050	u	5.36	C8, C11, C12
C11	137.30	137.180	C	–	–	–	–
C12	128.60	128.350	CH	7.310	u	7.40	C14, C10
C13	130.10	128.200	CH	7.400	u	7.31, 7.36	C11
C14	126.80	126.710	CH	7.360	u	7.40	C12
C15	168.20	169.700	C	–	–	–	–
C16	50.80	50.800	CH ₂	5.400	u	–	C18, C15
C16	50.80	50.800	CH ₂	3.330	u	–	–
C17	35.50	36.700	CH ₃	2.910	u	–	C18, C16
C18	172.30	172.460	C	–	–	–	–
C19	57.80	57.570	CH	4.390	u	2.42	C18
C20	31.40	28.740	CH ₂	2.420	u	4.39, 1.78	–
C20	31.40	28.740	CH ₂	2.040	u	–	–
C21	22.00	24.860	CH ₂	1.850	u	–	–
C21	22.00	24.860	CH ₂	1.780	u	2.42, 3.60	–
C22	48.40	47.360	CH ₂	3.600	u	1.78	–
C22	48.40	47.360	CH ₂	3.770	u	–	–
C23	168.40	170.340	C	–	–	–	–
C24	54.50	53.290	CH	5.560	u	3.28	C23, C31
C25	35.20	35.420	CH ₂	3.280	u	5.56	C23, C26, C27
C25	35.20	35.420	CH ₂	3.020	u	–	–
C26	136.80	136.860	C	–	–	–	–
C27	129.70	130.100	CH	7.150	u	7.12	C29
C28	128.60	129.270	CH	7.120	u	7.15, 7.04	C26
C29	126.80	127.360	CH	7.040	u	7.12	C27
C30	30.10	30.660	CH ₃	3.030	u	–	C31, C24
C31	174.10	172.980	C	–	–	–	–
C32	47.60	49.680	CH	4.920	u	8.20, 1.34	C31
C33	39.90	40.610	CH ₂	1.340	u	1.55, 4.92	–
C33	39.90	40.610	CH ₂	0.120	u	–	–
C34	24.80	25.140	CH	1.550	u	0.91, 0.84, 1.34	–

(continued)

Table 4.27 (continued)

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	C HMBC
C35	23.70	22.520	CH ₃	0.910	u	1.55	–
C36	20.80	22.520	CH ₃	0.840	u	1.55	–
C37	170.90	170.450	C	–	–	–	–
C38	58.10	58.290	CH	4.400	u	2.63, 5.88	C37
C39	28.50	30.580	CH	2.630	u	0.89, 0.80, 4.40	–
C40	16.30	19.200	CH ₃	0.800	u	2.63	–
C41	19.60	18.630	CH ₃	0.890	u	2.63	–
C42	155.30	156.000	C	–	–	–	–
N1	100.00 ^a	–	NH	8.610	u	5.36	C15
N2	110.00 ^a	–	NH	8.200	u	4.92	C37
N3	120.00 ^a	–	NH	5.880	u	4.40	C1

^a Fictitious ¹⁵N chemical shifts

**Fig. 4.67** Cordyheptapeptide C: Molecular connectivity diagram

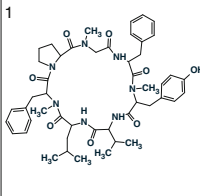
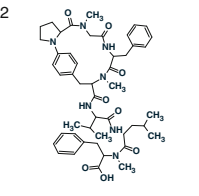
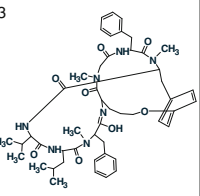
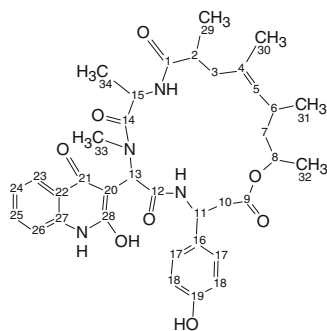
		
$d_A(^{13}\text{C})$: 1.435 $d_N(^{13}\text{C})$: 1.362 $d_I(^{13}\text{C})$: 1.353 $\text{max}_d A(^{13}\text{C})$: 13.520 $d_A(^1\text{H})$: 0.155 $d_N(^1\text{H})$: 0.230 $d_I(^1\text{H})$: 0.251	$d_A(^{13}\text{C})$: 1.729 $d_N(^{13}\text{C})$: 1.616 $d_I(^{13}\text{C})$: 1.806 $\text{max}_d A(^{13}\text{C})$: 11.470 $d_A(^1\text{H})$: 0.254 $d_N(^1\text{H})$: 0.252 $d_I(^1\text{H})$: 0.254	$d_A(^{13}\text{C})$: 1.807 $d_N(^{13}\text{C})$: 1.933 $d_I(^{13}\text{C})$: 1.831 $\text{max}_d A(^{13}\text{C})$: 22.130 $d_A(^1\text{H})$: 0.232 $d_N(^1\text{H})$: 0.267 $d_I(^1\text{H})$: 0.270

Fig. 4.68 Cordyheptapeptide C: Three top ranked structures

authors [30]. Note that a large enough structure containing 63 skeletal atoms was quickly elucidated without introduction of any user suggestions, some of which may be questionable and risky.

4.28 Pipestelide C

Investigation of the Indo-Pacific marine sponge *Pipetela candelabra* led Sorres et al. [31] to the discovery of three new *unusual derivatives* of jaspamide. These are pipestelide A containing an ortho-brominated β -tyrosine residue, Pipestelide B, with the *first* natural *Z*-configuration of the double bond in the polypropionate region, and Pipestelide C, with an *unprecedented* hydroxyquinolinone. In this section, we will discuss the structure elucidation of Pipestelide C (**4.56**).



4.56

The ESIMS spectrum of compound **4.56** displayed a pseudomolecular ion peak at m/z 661.3 $[M+H]^+$. The molecular formula of **4.56** was determined to be $C_{36}H_{44}N_4O_8$ by HRESIMS, $RDBE = 17$. Due to the very small amount of the isolated compound (100 μ g), the NMR analysis and the structure elucidation of **4.56** required intensive use of the advanced 600 MHz 1.7 mm microprobe.

The spectroscopic NMR data that were used for the structure elucidation of Pipestelide C (1H , ^{13}C , HSQC, COSY, and HMBC) are collected in Table 4.28, and the initial MCD (Fig. 4.69) represents the spectrum-structural information graphically.

MCD overview The molecule contains 48 skeletal atoms among which there are eight oxygens and four nitrogens, while the degree of unsaturation is equal to 17. One can conclude that the molecule is not only quite large but its molecular formula shows some deficit of hydrogen atoms and a diversity of heteroatoms. Due to the presence of nitrogen atoms whose influence on ^{13}C chemical shifts of neighboring carbons is difficult to predict on the basis of characteristic spectral features only (for sp^2 -hybridized carbons which are neighboring with a nitrogen atom the range of

Table 4.28 Pipestelide C: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	177.6	173.13	C	–	–	–	–
C2	38.4	41.38	CH	2.83	u	1.21, 2.48	–
C3	41.1	42.42	CH ₂	2.48	u	2.83	–
C3	41.1	42.42	CH ₂	1.87	u	–	–
C4	134	132.23	C	–	–	–	–
C5	128.4	128.3	CH	4.78	d(9.3)	2.27	C31, C3
C6	30.3	29.31	CH	2.27	u	0.80, 4.78, 0.90	–
C7	44.4	43.47	CH ₂	0.9	u	2.27, 4.69	–
C8	71.4	70.88	CH	4.69	u	0.90, 1.03	–
C9	173.6	170.11	C	–	–	–	–
C10	41.9	40.19	CH ₂	2.86	u	5.18	C11, C16, C9
C10	41.9	40.19	CH ₂	2.68	u		
C11	50.5	50.12	CH	5.18	u	2.86	C17, C9
C12	173.7	168.88	C	–	–	–	–
C13	58	55.48	CH	6.46	s	–	C28, C21, C12, C20
C14	174.2	173.26	C	–	–	–	–
C15	48.3	46.32	CH	5.13	u	1.37	C34, C14
C16	131.8	133.24	C	–	–	–	–
C17	128.7	127.21	CH	7.1	u	6.67	C11, C19
C18	117	115.07	CH	6.67	u	7.1	C16
C19	158.4	156.28	C	–	–	–	–
C20	101.9	98.68	C	–	–	–	–
C21	176.4	176.56	C	–	–	–	–
C22	123	121.71	C	–	–	–	–
C23	126.2	125.14	CH	8.04	u	7.07	C25, C27, C21
C24	121.9	123.39	CH	7.07	u	8.04, 7.37	C22, C26
C25	131.2	131.74	CH	7.37	u	7.07, 7.18	C27, C23
C26	116.3	114.28	CH	7.18	u	7.37	C24, C22
C27	140.1	138.17	C	–	–	–	–
C28	167.4	158.98	C	–	–	–	–
C29	21.5	19.77	CH ₃	1.21	d(7.1)	2.83	C2, C3, C1
C30	18.6	18.57	CH ₃	1.62	S		C4, C5, C3
C31	22.8	22.09	CH ₃	0.8	d(6.4)	2.27	C7, C5, C6
C32	19.2	19.28	CH ₃	1.03	d(6.1)	4.69	C8, C7
C33	32.3	31.56	CH ₃	3.2	s	–	C14, C13
C34	18.6	17.96	CH ₃	1.37	d(6.7)	5.13	C15, C14

chemical shifts variation is broad enough—approximately between 100 and 175 ppm), many carbon atoms did not receive automatically corresponding labels. The hybridization states of the three carbon atoms (101.9, 123.0, and 131.8) colored in light blue color were not definitively suggested by the program. Five “free”

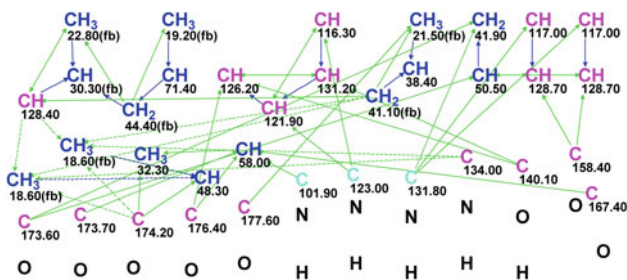


Fig. 4.69 Pipestelide C: The initial MCD

exchangeable hydrogen atoms may, in principle, produce all possible combinations with oxygen and nitrogen atoms to form NH, NH₂, and OH groups. Experience shows that in such a situation an attempt to perform structure generation from the MCD “as is” will lead to a huge number of possible structures and to the unpredictable consumption of processor time. To ease and accelerate problem solving it is necessary to use as much additional structural information as possible and input it into the initial MCD using a priori knowledge obtained from the sample origin and from other available sources. In our case the most important source of additional constraints is the fact that the molecule is a member of a family of cyclic peptides.

With this in mind the MCD was modified as shown in Fig. 4.70.

The following MCD edits were made: (a) carbons C 173.6–C 177.6 were connected with oxygen atoms to produce carbonyl groups; (b) the CH₃ (32.3) was connected to a nitrogen atom; (c) the labels “ob” were assigned to some CH and CH₂ groups on the basis of their characteristic ¹³C and ¹H NMR chemical shifts; (d) two groups of *sp*²-hybridized carbons were distinguished for which the existence of two benzene rings was suggested on the basis of their typical HMBC and COSY correlations, and 1–1 connectivities of one-bond length were drawn by hand to close the benzene rings; (e) carbons C 123.0 and C 131.8 were marked as *sp*²

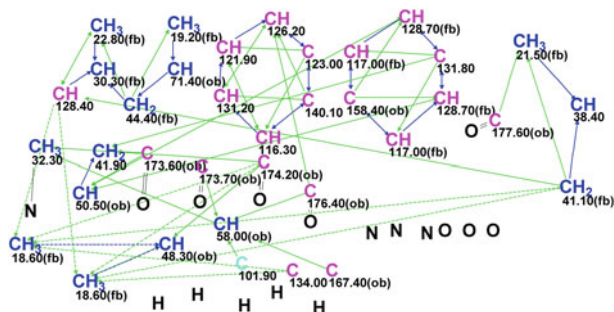


Fig. 4.70 Pipestelide C: Modified MCD displaying two benzene rings defined by drawing their structures with one-bond length connectivities

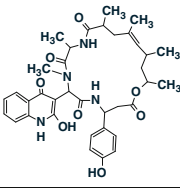
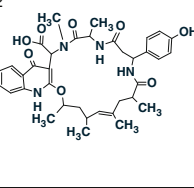
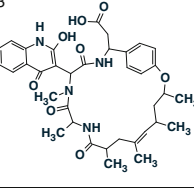
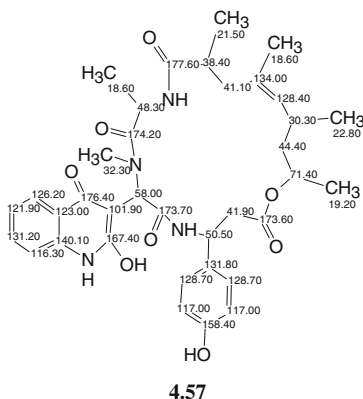
		
$d_A(^{13}\text{C}): 1.761$ $d_N(^{13}\text{C}): 2.049$ $d_I(^{13}\text{C}): 1.642$ $\text{max_}d_A(^{13}\text{C}): 8.420$ $d_A(^1\text{H}): 0.182$ $d_N(^1\text{H}): 0.206$ $d_I(^1\text{H}): 0.211$	$d_A(^{13}\text{C}): 1.909$ $d_N(^{13}\text{C}): 2.123$ $d_I(^{13}\text{C}): 1.651$ $\text{max_}d_A(^{13}\text{C}): 8.920$ $d_A(^1\text{H}): 0.200$ $d_N(^1\text{H}): 0.217$ $d_I(^1\text{H}): 0.175$	$d_A(^{13}\text{C}): 2.050$ $d_N(^{13}\text{C}): 2.056$ $d_I(^{13}\text{C}): 1.764$ $\text{max_}d_A(^{13}\text{C}): 8.420$ $d_A(^1\text{H}): 0.229$ $d_N(^1\text{H}): 0.249$ $d_I(^1\text{H}): 0.236$

Fig. 4.71 Pipestelide C: The three top ranked structures of the output file

hybridized and finally, (f) the numbers of hydrogen atoms attached to the neighboring carbons were set in accordance with Table 4.28, column M(J).

No contradictions were detected in the 2D NMR data and strict structure generation was initiated from the modified MCD, which gave the following results: $k = 2,665,136 \rightarrow 11,814 \rightarrow 11,797$, $t_g = 1 \text{ h } 6 \text{ min}$. The three top ranked structures of the output file are presented in Fig. 4.71.

The best structure is identical to the structure of Pipestelide C, though the differences between the deviations of the first and second structures are very small ($d_A(^1\text{H})$ and $d_N(^1\text{H})$ chemical shift deviations in some measure confirm the selection). Therefore, additional confirmation is desirable for which NOESY/ROESY spectra and X-ray analysis could be very useful. Structure 4.57 displays the ^{13}C chemical shift assignment automatically performed by the program.



The *Tautomers* routine incorporated into ChemSketch allowed us to establish that the structure of Pipestelide C can exhibit three tautomeric forms of the

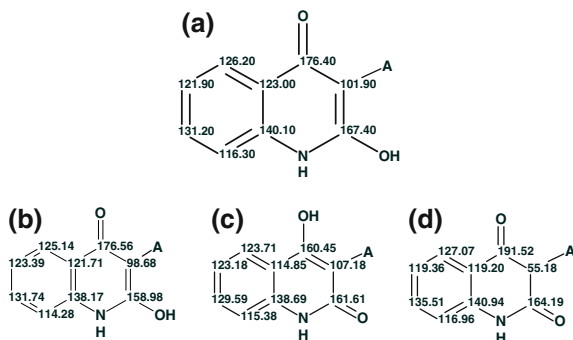


Fig. 4.72 Comparison of experimental ^{13}C chemical shifts (a) with predicted ones for the three possible tautomeric forms b, c, and d of the structural fragment available for tautomerism

2-hydroxy-4-quinolone fragment. ^{13}C chemical shifts of all forms were predicted by the HOSE code-based methods and the predicted chemical shifts are shown in comparison with experimental ones (Fig. 4.72).

Figure 4.72 readily demonstrates that the conceivable forms c and d should be rejected, and structure 4.57 indeed represents the most probable structure of the analyzed molecule.

The problem exemplifies a case when a structure containing an *unprecedented* hydroxyquinolinone fragment was elucidated in a manageable time only by involving additionally a priori information based on spectroscopic experience, characteristic spectral features, and knowing that the unknown belongs to a specific family of natural products. ^{13}C chemical shift prediction allowed selection of the correct structure among ~12,000 candidates which possess very similar structures.

4.29 Tetrabrominated Diphenyl Ether

Dai et al. [32] initiated research to discover potent and selective small molecule inhibitors of hypoxia-mediated tumor cell adaptation, survival, and metastatic spread. They found that the crude extract of the sponge *Lendenfeldia* sp. (Spongiiidae) inhibited hypoxia-induced HIF-1 activation (99 % inhibition at $5\ \mu\text{g mL}^{-1}$). The extract (4 g) was purified to yield two new compounds and three known homoscalarane sesterterpenes. The new compound 4.58 in particular was identified by comparison of its spectroscopic data with those reported in the literature and by interpretation of ^1H , ^{13}C , ^1H - ^1H COSY, ^1H - ^{13}C HMQC, and ^1H - ^{13}C HMBC spectra.

The molecular formula $\text{C}_{22}\text{H}_{18}\text{O}_6$ was deduced from analysis of the HRESIMS data [m/z 378.1105 (calculated for $\text{C}_{22}\text{H}_{18}\text{O}_6$:378.1103)]. While the HRESIMS suggested that the structure of 4.58 contains 22 carbons, only 11 carbon resonances

Table 4.29 Tetrabrominated diphenyl ether: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	C HMBC
C1	150.9	151.1	C	–	–	–
C2	119.1	116.2	CH	6.83	d(2.3)	C4, C1, C6, C3
C3	117.5	116.3	C	–	–	–
C4	130.8	128.3	CH	7.46	d(2.2)	C6, C3
C5	119.2	118.5	C	–	–	–
C6	146.2	145.1	C	–	–	–
C7	61.8	60.4	CH_3	4.03	s	C6
C8	139.3	137.7	C	–	–	–
C9	150.6	151.8	C	–	–	–
C10	120.2	119.7	CH	7.18	d(2.2)	C8, C9, C11, C12
C11	120.1	118.9	C	–	–	–
C12	127.6	125.2	CH	7.35	d(2.2)	C10, C8, C13
C13	117.4	118	C	–	–	–

closely associated with the peaks at 117.4 and 150.5 ppm. These inconsistencies prompted the authors [33] to conduct their own examination of the natural product isolate. The (–)-HRESIMS spectrum of the metabolite exhibited a cluster of peaks indicating four bromines are present and a molecular formula of $C_{13}H_8Br_4O_3$ ($[M-H]^-$ m/z 526.7119) with DBE = 8. Note that the number of skeletal atoms is two and half times larger than the number of hydrogens, which suggests a challenging structure elucidation for a molecule which seems very simple at first glance (see Sect. 1.2.2, Crews's rule).

1D and 2D NMR spectroscopic data used for inferring structure **4.59a** are shown in Table 4.29.

The 1H NMR spectrum also revealed small coupling constants consistent with meta-protons on an aromatic ring (see Table 4.29). The COSY correlation data showed that two spin systems were present, with H2 coupled to H4 and H10 coupled to H12 (see structure **4.59a** and Table 4.29).

Interpretation of the HSQC and HMBC data (Table 4.29) revealed a tetrabrominated diphenyl ether with the OMe and OH on separate rings and two bromines on each ring. In the article [33], eight possible structures (see Fig. 4.73) were proposed based on the data available.

To select the best structure ^{13}C chemical shifts were predicted using the CambridgeSoft ChemBioDraw program. Comparison of the calculated chemical shifts with the observed ^{13}C NMR data of the metabolite (Table 4.29) highlighted structure #1 as the closest match and therefore the possible identity of the natural product.

We will show how the problem could be solved with Structure Elucidator. The molecular formula $C_{13}H_8Br_4O_3$ and the spectroscopic data presented in Table 4.29 were input into StrucEluc and a MCD was created. A value of DBE equal to eight allows us to suggest that all downfield ^{13}C chemical shifts are related to sp^2 -

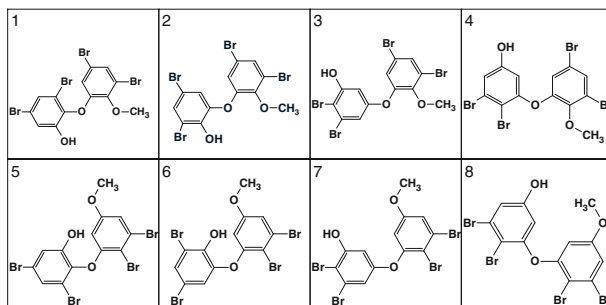


Fig. 4.73 Tetrabrominated diphenyl ether: The eight candidate structures proposed by the authors [33]

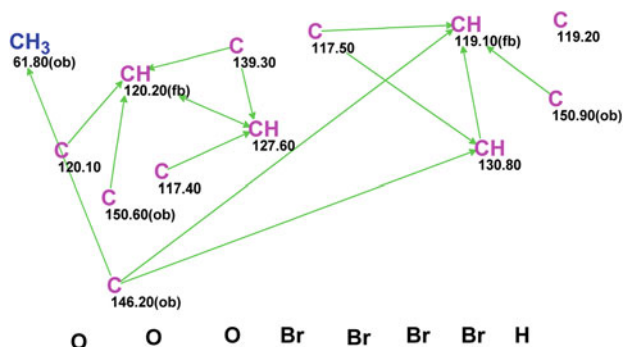


Fig. 4.74 Tetrabrominated diphenyl ether: The edited MCD

hybridized benzene carbon atoms, while the methyl group is definitely connected to an oxygen atom. A label “ob” (the neighbor atom is a heteroatom) was assigned to carbons C 146.2, C 150.6, and C 150.9 ppm. The signal multiplicities in the ^1H NMR spectrum presented in Table 4.29 were used to set the numbers of hydrogens attached to neighboring carbon atoms (all zeros in a given case). The edited MCD is shown in Fig. 4.74.

Strict structure generation was performed and ^{13}C chemical shift prediction was carried out during the structure generation to eliminate structures with deliberately high average deviation values. Results: $k = 26,597 \rightarrow 7 \rightarrow 5$, $t_g = 1$ min 18 s. The four top ranked structures are shown in Fig. 4.75.

Structure #1 characterized by rather small average deviations is confidently selected as the most probable one in agreement with revised structure 4.59a. The assignments of the ^{13}C chemical shifts coincide with those shown on structure 4.59b.

As described above, Podlesny and Kozlowski checked eight isomers containing two benzene rings with isolated hydrogens attached to the rings in the meta-position. They assumed that these isomers make up a full set of the isomers possessing the

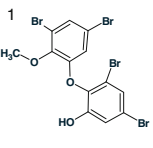
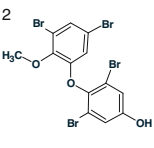
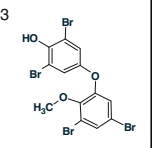
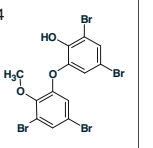
			
$d_A(^{13}\text{C}): 1.091$ $d_N(^{13}\text{C}): 1.299$ $d_I(^{13}\text{C}): 1.302$ $\text{max_}d_A(^{13}\text{C}): 2.910$	$d_A(^{13}\text{C}): 2.027$ $d_N(^{13}\text{C}): 2.719$ $d_I(^{13}\text{C}): 2.271$ $\text{max_}d_A(^{13}\text{C}): 5.440$	$d_A(^{13}\text{C}): 2.893$ $d_N(^{13}\text{C}): 3.547$ $d_I(^{13}\text{C}): 3.228$ $\text{max_}d_A(^{13}\text{C}): 9.590$	$d_A(^{13}\text{C}): 3.037$ $d_N(^{13}\text{C}): 3.145$ $d_I(^{13}\text{C}): 2.472$ $\text{max_}d_A(^{13}\text{C}): 10.170$

Fig. 4.75 Tetrabrominated diphenyl ether: The four top ranked structures of the output file

mentioned properties. Given that the molecule includes two benzene rings, a question arises: How many isomers of $\text{C}_{13}\text{H}_8\text{Br}_4\text{O}_3$ composition with isolated hydrogens attached to the rings in the meta-position should be checked to select the best structure? We repeated the structure generation from MCD shown in Fig. 4.74 under the following constraint: a benzene ring was placed in the GOODLIST (the window **GOOD**) and the number of benzene fragments included in each generated structure was set as two (right click on **Quantity** in the **Properties**, then left click on **Edit Property**). As a result 12 structures were generated. We found that in total 12 (not 8!) possible structures should be verified; hence the additional 4 structures containing hydrogens in meta-positions (see Fig. 4.76) were missed by the authors [33].

Therefore, the authors [33] did not check all candidate structures. Note that structure #1 (Fig. 4.76) is characterized by average deviations which show that this structure should still be considered as a probable candidate. StrucEluc automatically produced and verified a full set of structural hypotheses, which cannot be provided when traditional methods of structure elucidation are used. If the multiplicities in the ^1H NMR spectrum were ambiguous, it would be necessary to check 44 isomers containing two benzene rings, and it is obvious that “manual generation” of 44 isomers would be a rather difficult task when a traditional approach is used.

The described example is especially educational. It shows visually how a set of false “axioms” led the authors [32] to a false structure. The origin of these axioms remains very mysterious. For example, it is difficult to understand how an ion m/z 378.1105 was mistakenly accepted as a molecular ion.

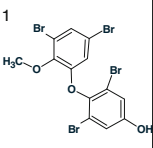
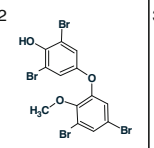
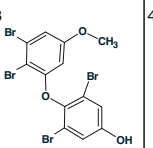
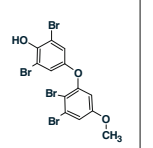
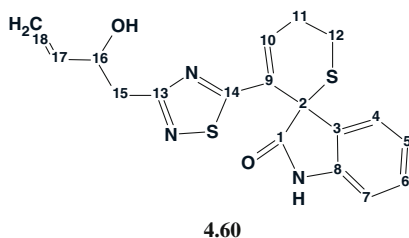
			
$d_A(^{13}\text{C}): 2.027$ $d_N(^{13}\text{C}): 2.719$ $d_I(^{13}\text{C}): 2.271$ $\text{max_}d_A(^{13}\text{C}): 5.440$	$d_A(^{13}\text{C}): 3.201$ $d_N(^{13}\text{C}): 3.755$ $d_I(^{13}\text{C}): 3.703$ $\text{max_}d_A(^{13}\text{C}): 10.590$	$d_A(^{13}\text{C}): 7.534$ $d_N(^{13}\text{C}): 7.901$ $d_I(^{13}\text{C}): 7.365$ $\text{max_}d_A(^{13}\text{C}): 18.330$	$d_A(^{13}\text{C}): 8.317$ $d_N(^{13}\text{C}): 8.251$ $d_I(^{13}\text{C}): 8.854$ $\text{max_}d_A(^{13}\text{C}): 18.500$

Fig. 4.76 Tetrabrominated diphenyl ether: The additional four structures containing hydrogens in meta-positions. Structure #1 coincides with structure #2 of the solution presented in Fig. 4.75

4.30 Indole Alkaloid

Dried roots and leaves of the plant *Isatis indigotica* are used in traditional Chinese medicine for the treatment of various diseases. Diverse structures and significant biological activities from extracts of this plant have attracted considerable interest. Chemical and pharmacological studies have resulted in the characterization of constituents with different structural features and biological activities. As part of a program to assess the chemical and biological diversity of traditional Chinese medicines, an aqueous extract of the roots of *I. indigotica* was investigated by Chen and coworkers [34]. Specifically, the authors isolated and structurally characterized an indole alkaloid containing *unusual* dihydrothiopyran and 1,2,4-thiadiazole rings (**4.60**).



Here we will discuss the structure elucidation of compound **4.60**. The IR spectrum of **4.60** showed absorption bands for hydroxy and/or amino groups ($3,256\text{ cm}^{-1}$), carbonyl ($1,715\text{ cm}^{-1}$), and aromatic ring ($1,619$ and $1,474\text{ cm}^{-1}$) functionalities. The positive mode ESIMS of **4.60** exhibited quasimolecular ion peaks at m/z 372 $[M+H]^+$, 394 $[M+Na]^+$, and 410 $[M+K]^+$. The molecular formula of $C_{18}H_{17}N_3O_2S_2$, with 12 degrees of unsaturation, was determined from HRESIMS at m/z 372.0844 $[M+H]^+$ (calculated for $C_{18}H_{17}N_3O_2S_2$, 372.0835) and 394.0659 $[M+Na]^+$ (calculated for $C_{18}H_{17}N_3O_2S_2Na$, 394.0654), combined with the NMR data which are presented in Table 4.30.

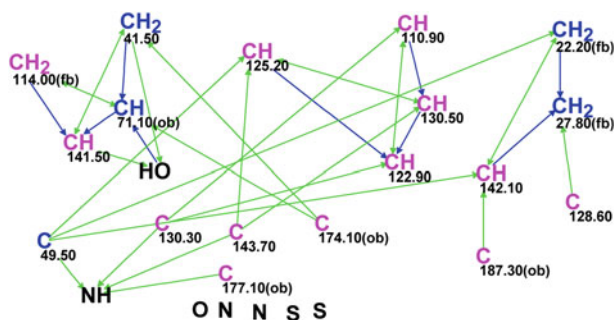
Figure 4.77 shows the corresponding MCD which was slightly edited as described below.

MCD overview Three carbon atoms—C 110.9, C 128.6, and C 130.3 marked by the program as “ sp^2 or sp^3 ” were defined as “ sp^2 ” hybridized. The methine group CH 71.1(4.31), as well as quaternary carbons C 174.1–C 187.3 were supplied with the label “*ob*”. Two hydrogen atoms were allowed to be attached to skeletal atoms that are neighbors of carbon C 142.1 (see Table 4.30). MCD checking resulted in a program message that at least one nonstandard connectivity exists in the collective

Table 4.30 Indole alkaloid: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	C HMBC
C1	177.1	174.5	C	–	–	–	–
C2	49.5	55.32	C	–	–	–	–
C3	130.3	132.07	C	–	–	–	–
C4	125.2	125.35	CH	7.07	u	6.9	C2, C6, C8
C5	122.9	121.64	CH	6.9	u	7.24, 7.07	C7, C3
C6	130.5	128.42	CH	7.24	u	6.90, 6.99	C8, C4
C7	110.9	109.05	CH	6.99	u	7.24	C3, C5
C8	143.7	142.13	C	–	–	–	–
C9	128.6	129.41	C	–	–	–	–
C10	142.1	134.7	CH	7.4	dd(4.8, 4.2)	2.84	C2, C12, C14
C11	27.8	26.07	CH ₂	2.84	u	7.40, 3.76	C9
C12	22.2	23.86	CH ₂	2.74	u	–	–
C12	22.2	23.86	CH ₂	3.76	u	2.84	C2, C10
C13	174.1	165.95	C	–	–	–	–
C14	187.3	183.41	C	–	–	–	–
C15	41.5	39.11	CH ₂	2.88	u	4.31	C17, C13
C15	41.5	39.11	CH ₂	2.79	u	–	–
C16	71.1	71.65	CH	4.31	u	5.63, 2.88, 3.95	C18, C13
C17	141.5	139.27	CH	5.63	u	4.98, 4.31	C15
C18	114	115.4	CH ₂	4.98	u	5.63	C16
N1	100 ^a	–	NH	9.68	u	–	C1, C2, C3, C8
O1	150 ^a	–	OH	3.95	u	4.31	C15, C17

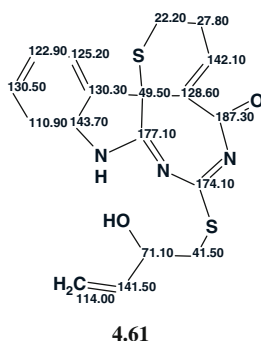
^a Fictitious ¹⁵N and ¹⁷O NMR chemical shifts

**Fig. 4.77** Indole alkaloid: The slightly edited MCD

2D NMR data. The presence of chemical bonds between heteroatoms in the molecule was not suggested during the MCD checking process.

FSG in the mode “Choose Option Automatically” accompanied by ¹³C chemical shift calculation was performed with the following result: $k = 163 \rightarrow 1$, $t_g = 2$ s, one

connectivity was extended ($m = 1$). The single structure **4.61** characterized with deviations $d_A = 5.65$, $d_I = 3.13$, $d_N = 3.58$, $d_A(\max) = 30.21$ ppm is shown below:



Deviations of 3.1–3.6 ppm are not uncommon when the analyzed structure is unusual, but the value of $d_A = 5.65$ ppm, together with the maximum deviation $d_A(\max) = 30.21$ ppm associated with the chemical shift 177.1 ppm should be considered as a warning and hints to the necessity of additional computational experiments. Therefore, the FSG was repeated with the options $m = 2$, $a = 1$ to provide the results: $k = 1,731 \rightarrow 2$, $t_g = 52$ s, 465 from 465 possible connectivity combinations were checked. During this run, structure **4.61** was generated again along with a second structure for which the deviations turned out to be significantly larger.

The absence of any signs indicating that deviations become smaller with enlarging m value suggests that it is desirable to check the possibility of forming chemical bonds between heteroatoms (up to this point we supposed that this type of bond is impossible in the molecule). With this in mind the “Check MCDs options” were set as shown in Fig. 4.78.

No contradictions were found in the MCD in this case, and strict structure generation was performed with the following results: $k = 780 \rightarrow 313 \rightarrow 313$, $t_g = 0.4$ s. The three top ranked structures of the output file are presented in Fig. 4.79.

The best structure coincides with structure **4.60**, and the ^{13}C chemical shift assignments are displayed on structure **4.62**:

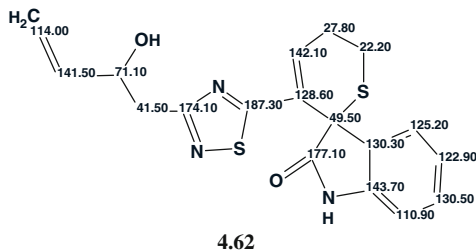




Fig. 4.78 Indole alkaloid: The options of the MCD checking process. The checkbox “allow bonds between heteroatoms” is selected

<p>1</p>	<p>2</p>	<p>3</p>
<p>$d_A(^{13}\text{C})$: 2.628 $d_N(^{13}\text{C})$: 2.894 $d_I(^{13}\text{C})$: 2.119 $\text{max}_dA(^{13}\text{C})$: 8.150</p>	<p>$d_A(^{13}\text{C})$: 4.267 $d_N(^{13}\text{C})$: 4.906 $d_I(^{13}\text{C})$: 4.240 $\text{max}_dA(^{13}\text{C})$: 25.370</p>	<p>$d_A(^{13}\text{C})$: 4.384 $d_N(^{13}\text{C})$: 3.104 $d_I(^{13}\text{C})$: 3.603 $\text{max}_dA(^{13}\text{C})$: 17.240</p>

Fig. 4.79 Indole alkaloid: The three top ranked structures of the output file

Thus, using the problem solving strategy inherent for Structure Elucidator a structure possessing a fairly unusual skeleton was quickly and reliably ($\Delta = d(2) - d(1) \cong 2$ ppm) identified with the assistance of the Structure Elucidator software. The options allowing chemical bonds between heteroatoms have to always be selected if a researcher works with chemical classes for which bonds of this type are common.

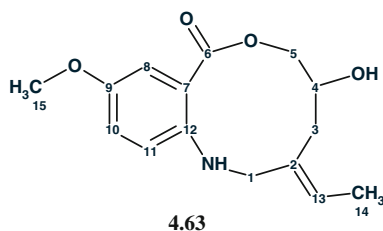
4.31 Barmumycin

Natural products from terrestrial plants and microorganisms have long been a traditional source of drugs. However, over the past few years marine organisms have garnered ever-increasing attention as a rich bank of new bioactive compounds.

Marine actinomycetes have also proven to be an important source of biologically active compounds.

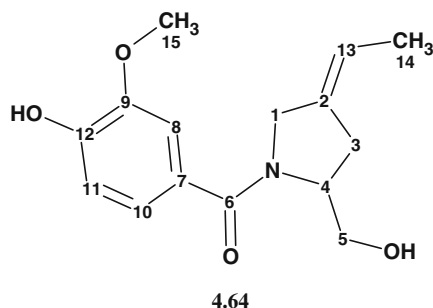
Among the marine actinomycetes those of the genus *Streptomyces* have clearly shown the most pharmacological potential; however, in many bioactive cultures they have yielded only compounds that are already known. During ongoing research efforts to explore the biosynthetic potential of rare marine microorganisms, Lorente and coworkers [35] isolated two known compounds, pretomaymycin and oxotomaymycin, plus the previously unknown compound barmumycin from the culture broth of the marine actinomycete *Streptomyces* sp. BOSC-022A, isolated from a tunicate collected off the Scottish coast. Barmumycin and its diacetate show antitumor activity at micromolar concentrations. The authors reported the isolation, total synthesis, and structure elucidation of Barmumycin [35].

The molecular formula of Barmumycin was determined to be $C_{15}H_{19}NO_4$ by HRMS MALDI-TOF; it gave an $(M+H)^+$ ion at m/z 278.13840 (calculated m/z 278.13869 for $C_{15}H_{20}NO_4$). The IR absorption band at $3,288\text{ cm}^{-1}$ gives a hint to the presence of OH/NH groups, while the bands at $1,600$ and $1,585\text{ cm}^{-1}$ readily indicate the presence of a benzene ring. Reaction of barmumycin with acetic anhydride and pyridine gave a diacetyl derivative, which was confirmed by MS, indicating the presence of two OH and/or NH protons. On the basis of the MS findings and data from one-dimensional ^1H and ^{13}C NMR and two-dimensional (COSY, HMBC, and NOESY) experiments, the authors [35] initially proposed that the structure of the isolated natural product was that of benzomacrolactone (**4.63**), derived from 5-methoxy-2-aminobenzoic acid with an exocyclic (E)-ethylidene and one alcohol function.



To confirm this assignment compound **4.63** was synthesized via two different strategies starting from an o-aminobenzoic ester.

However, comparison of the NMR spectra for **4.63** with those for isolated Barmumycin showed dramatic differences. The structure of Barmumycin was reassessed, and the most probable option conceived was compound **4.64**.



Compound **4.64** was subsequently prepared (in five steps and 18 % overall yield) for comparison with the natural compound. The spectroscopic data for **4.64** fully coincided with that for Barmumycin, thereby confirming that the two structures are equivalent.

We investigated how the structure elucidation of Barmumycin could be carried out using StrucEluc. The data (^{13}C , ^1H , HSQC, and HMBC) available from the article [35] are collected in Table 4.31:

The edited MCD (MCD) created from the spectroscopic data from Table 4.31 is shown in Fig. 4.80.

Table 4.31 Barmumycin: Spectroscopic NMR data

Label	δC	CH_n	δH	M(J)	C HMBC
C1	55.1	CH_2	4.11	u	–
C2	134.7	C	–	–	–
C3	30.1	CH_2	2.72	u	–
C3	30.1	CH_2	2.27	u	–
C4	60.6	CH	4.67	u	–
C5	67.1	CH_2	3.75	u	–
C6	172.2	C	–	–	–
C7	128.3	C	–	–	–
C8	110.6	CH	7.09	s	C6, C12, C10, C9
C9	146.7	C	–	–	–
C10	121	CH	7.04	d	C8, C12
C11	114.1	CH	6.91	d	C9, C7
C12	147.8	C	–	–	–
C13	117.6	CH	5.34	u	–
C14	14.5	CH_3	1.62	u	C2, C13
C15	56.3	CH_3	3.9	u	C9

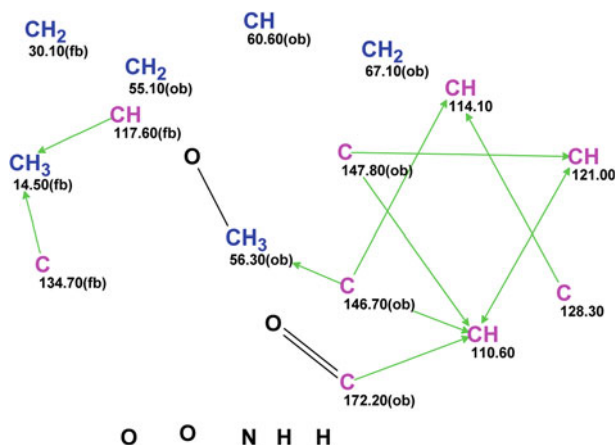


Fig. 4.80 Barmumycin: Edited MCD

MCD Overview The number of HMBC connectivities is small and, in addition, four carbon atoms have no connectivities at all. As a result we expect that the structure generation will be quite time-consuming. Therefore, some atom properties which follow from the spectrum-structure empirical correlations [11] were manually added to the MCD. The presence of an obvious carbonyl double bond and a methoxy group were also manually added. The ¹H NMR signal multiplicities shown in Table 4.31 were used to specify the numbers of hydrogen atoms on neighboring carbons.

Structure generation accompanied by ¹³C chemical shift prediction and spectral filtering (structures for which the deviation value was $d > 5$ ppm were rejected) gave the following results: $k = 73,373 \rightarrow 1,089 \rightarrow 778$, $t_g = 3$ min 17 s.

The three top structures of the output file were ranked in ascending order of the deviation value and are shown in Fig. 4.81.

Figure 4.81 shows that the first ranked structure coincides with the revised structure of Barmumycin 4.64. The original structure 4.63 was only ranked as 45th.

1	2	3
$d_A(^{13}\text{C}): 1.559$ $d_N(^{13}\text{C}): 1.776$ $d_I(^{13}\text{C}): 1.959$ $\text{max}_dA(^{13}\text{C}): 2.970$	$d_A(^{13}\text{C}): 2.385$ $d_N(^{13}\text{C}): 2.535$ $d_I(^{13}\text{C}): 2.432$ $\text{max}_dA(^{13}\text{C}): 5.420$	$d_A(^{13}\text{C}): 2.536$ $d_N(^{13}\text{C}): 2.625$ $d_I(^{13}\text{C}): 3.160$ $\text{max}_dA(^{13}\text{C}): 5.030$

Fig. 4.81 Barmumycin: Three top structures of the ranked output file

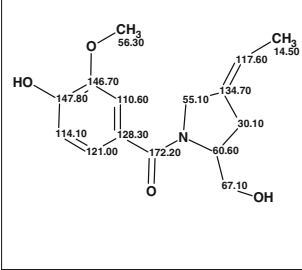
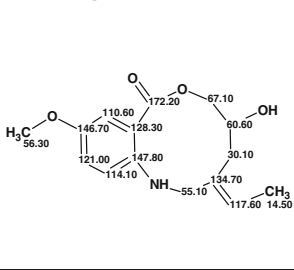
1 Revised	45 Original
	
$d_A(^{13}\text{C}): 1.559$ $d_f(^{13}\text{C}): 1.959$ $d_N(^{13}\text{C}): 1.776$ $\text{max}_dA(^{13}\text{C}): 2.970$	$d_A(^{13}\text{C}): 5.165$ $d_f(^{13}\text{C}): 4.712$ $d_N(^{13}\text{C}): 4.922$ $\text{max}_dA(^{13}\text{C}): 13.120$

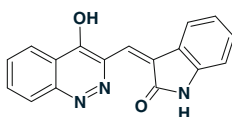
Fig. 4.82 Barmumycin: Original and revised structures. A comparison of the ^{13}C chemical shift assignments and average deviations calculated for both structures

The original and revised structures supplied with the ^{13}C chemical shift assignments automatically generated by the program are shown for comparison in Fig. 4.82.

At the end of the article the authors [35] conclude that “this work is a new example of the importance of total synthesis for structural characterization and confirmation of natural products.” We do not contest this statement but we do wish to add some comments reflecting our point of view regarding the methodology of structure revision. This example shows that if StrucEluc software was employed for elucidation of the Barmumycin structure from the very beginning the right structure would be unambiguously found in several minutes. The wrong original structure which occupied 45th position in the ranked output file would never have been considered and consequently it would not be necessary to carry out its total synthesis. Finally, if at least the empirical ^{13}C NMR chemical shift prediction was performed for the originally suggested structure then the erroneous structural hypothesis would be immediately rejected.

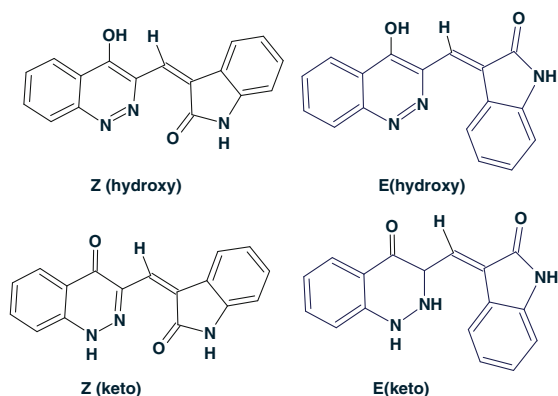
4.32 Schizocommunin

Schizocommunin was first reported in 1999 from the liquid culture medium of *Schizophyllum commune*, strain IFM 46788 (monokaryon). The following structure was suggested for Schizocommunin:



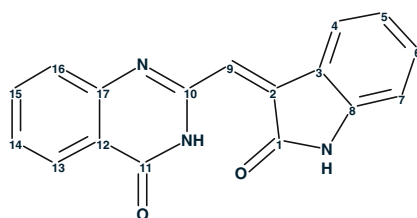
4.65

Fig. 4.83 Schizocommunin:
Possible configurations of
keto and hydroxy forms



Biological studies on schizocommunin were prevented by the very limited supply of the compound available from natural sources, and there have been no reports of the total synthesis. Although a spectroscopic analysis revealed that the structure of schizocommunin included both 4-hydroxycinnoline and oxindole skeletons that were connected by exomethylene, as shown in Z (hydroxy) in Fig. 4.83, the configuration of the olefin double bond was not discussed in detail. It is possible that **4.65** has either geometric structures Z- and E (hydroxy) or tautomeric structures Z- and E (keto).

Recently, Uehata and coworkers [36] performed the total synthesis of the putative structures of schizocommunin Z (hydroxy) and its geometric isomer E (hydroxy), which both exist in a keto form. However, the ^1H NMR spectra for synthetic Z- and E (keto) were not identical to those reported for natural Schizocommunin. A reinvestigation of the NMR and IR results for natural schizocommunin led the researchers to propose a revised structure, quinazolinone **4.66**, which was synthesized in a single step. All of the spectral data of Z-**4.66** were identical to those reported for natural schizocommunin.



4.66

The authors [36] performed a careful investigation of the problem, as a result of which methods of synthesis of the corresponding compounds were developed and the originally determined structure of Schizocommunin was revised.

Table 4.32 Schizocommunin: NMR spectroscopic data

Label	δC	$\delta\text{C}_{\text{calc}}$	XH_n	δH	M(J)	COSY	C HMBC
C1	168.8	167.84	C	–	–	–	–
C2	134.3	144.21	C	–	–	–	–
C3	123.3	123.34	C	–	–	–	–
C4	121.9	122.24	CH	7.94	u	7.1	C2, C6, C8
C5	122.6	121.41	CH	7.1	u	7.94, 7.36	C3
C6	131.6	129.37	CH	7.36	u	7.10, 6.93	C8, C4
C7	110.6	110.36	CH	6.93	u	7.36	C8
C8	141.6	141.29	C	–	–	–	–
C9	130	121.59	CH	7.58	u	–	C1, C3, C2
C10	150.4	154.38	C	–	–	–	–
C11	160.8	162.8	C	–	–	–	–
C12	121.5	120.16	C	–	–	–	–
C13	126	125.92	CH	8.17	u	7.6	C11, C15, C17
C14	128	128.3	CH	7.6	u	7.88, 8.17	C12
C15	134.7	135.01	CH	7.88	u	7.60, 7.79	C13, C17
C16	128	124.54	CH	7.79	u	7.88	C12, C14
C17	148.9	148.45	C	–	–	–	–
N1 ^a	150	–	NH	11.49	u	–	–
N2 ^a	170	–	NH	13.00	u	–	–

^aFictitious ¹⁵N chemical shifts

We suggested that if a CASE approach were used to elucidate the structure when this compound was firstly isolated, the structure revision would not be necessary: for final structure confirmation X-ray crystallographic analysis would be sufficient.

With this in mind, the molecular formula C₁₇H₁₁N₃O₂, the 1D and 2D NMR data of Schizocommunin extracted from the Supporting Information [36] were input into Structure Elucidator. Note that $n(\text{skeletal})/n(\text{hydrogens}) = 2$, hence the molecule is extremely hydrogen deficient, and the problem is challenging (see Sect. 1.2.2). The data are shown in Table 4.32 where atom numbers correspond to those which are displayed in the structure 4.66.

An MCD was created (Fig. 4.84), where two isolated groups of carbon atoms were selected, allowing the recognition of two benzene rings. The carbons included in the benzene rings were connected by COSY-like connectivities. As the presence of two carbonyl absorption bands at 1,675 and 1,685 cm⁻¹ are evident from the IR spectrum, carbons C 160.8 and 168.8 were connected to oxygens with double bonds.

Structure generation was performed with the option “Allow Bonds between Heteroatoms of the Same Atom Type” selected. ¹³C NMR chemical shifts were calculated during the structure generation, and the filtering procedure was used (structures characterized with $d > 5$ ppm were rejected by the filter). The results were: $k = 3,614 \rightarrow 44 \rightarrow 22$, $t_g = 2$ s. The eight top structures of the output file

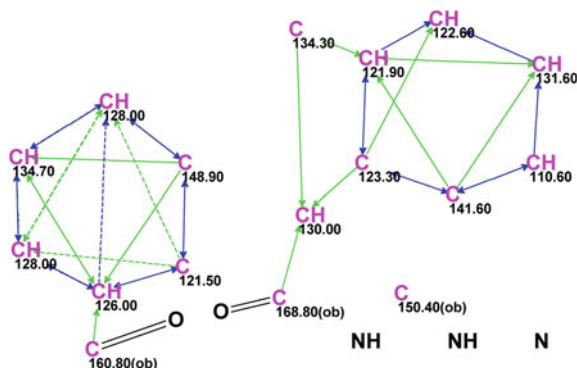


Fig. 4.84 Schizocommunin: Molecular connectivity diagram

ranked by d_A deviations (HOSE code-based spectrum prediction) are shown in Fig. 4.85.

The figure shows that the true structure of schizocommunin was selected as the most probable one by all three methods of chemical shift prediction. The *Z*- and *E*-configurations of the molecule supplied with the ^{13}C chemical shift assignment are displayed in Fig. 4.86.

It was also interesting to see how both the original and revised structures would be ranked if both of them are generated from the MCD and with $\text{N}=\text{N}$ bonds allowed. The corresponding MCD which contains free nitrogen atoms is shown in Fig. 4.87.

1	2	3	4
$d_A(^{13}\text{C}): 2.091$ $d_N(^{13}\text{C}): 3.108$ $d_I(^{13}\text{C}): 2.403$ $\text{max}_A d_A(^{13}\text{C}): 9.910$	$d_A(^{13}\text{C}): 2.521$ $d_N(^{13}\text{C}): 3.305$ $d_I(^{13}\text{C}): 2.543$ $\text{max}_A d_A(^{13}\text{C}): 6.080$	$d_A(^{13}\text{C}): 3.090$ $d_N(^{13}\text{C}): 4.739$ $d_I(^{13}\text{C}): 4.246$ $\text{max}_A d_A(^{13}\text{C}): 8.640$	$d_A(^{13}\text{C}): 3.241$ $d_N(^{13}\text{C}): 3.518$ $d_I(^{13}\text{C}): 3.271$ $\text{max}_A d_A(^{13}\text{C}): 10.210$
5	6	7	8
$d_A(^{13}\text{C}): 3.346$ $d_N(^{13}\text{C}): 3.225$ $d_I(^{13}\text{C}): 3.317$ $\text{max}_A d_A(^{13}\text{C}): 16.650$	$d_A(^{13}\text{C}): 3.458$ $d_N(^{13}\text{C}): 4.834$ $d_I(^{13}\text{C}): 4.957$ $\text{max}_A d_A(^{13}\text{C}): 11.090$	$d_A(^{13}\text{C}): 3.608$ $d_N(^{13}\text{C}): 4.007$ $d_I(^{13}\text{C}): 4.944$ $\text{max}_A d_A(^{13}\text{C}): 9.670$	$d_A(^{13}\text{C}): 3.915$ $d_N(^{13}\text{C}): 4.178$ $d_I(^{13}\text{C}): 4.157$ $\text{max}_A d_A(^{13}\text{C}): 15.820$

Fig. 4.85 Schizocommunin: Eight top structures of the ranked output file of the first run

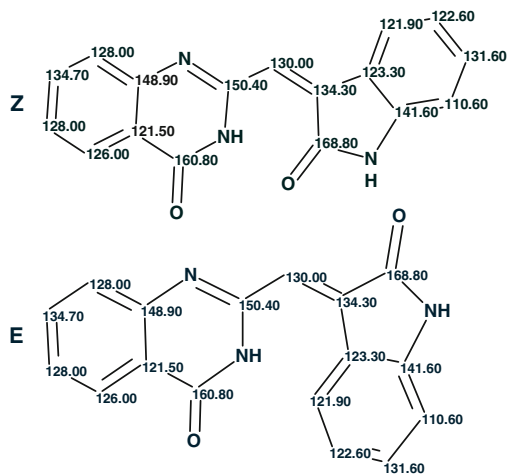


Fig. 4.86 Schizocommunin: Z- and E-configurations of Schizocommunin

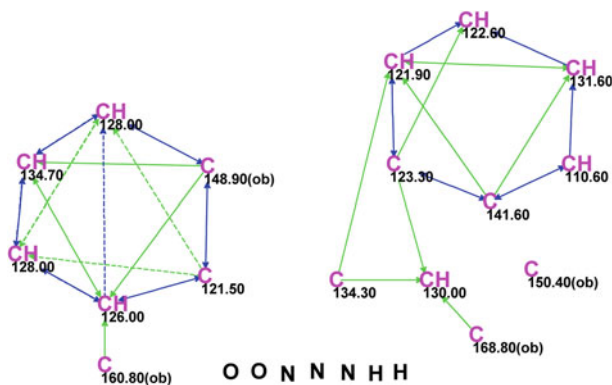


Fig. 4.87 Schizocommunin: Molecular connectivity diagram when N=N bonds are allowed

Structure generation under conditions similar to those used for the first run was carried out from this MCD, and the following results were obtained: $k = 176,535 \rightarrow 344 \rightarrow 297$, $t_g = 2$ min 10 s. The eight top structures of the output file ranked by d_A deviations are shown in Fig. 4.88.

Figure 4.88 shows that the original structure was placed in eighth position by the ranking procedure, which evidently confirmed the priority of the revised structure. Taking into account that all structures are very similar and that the difference between the deviations $d_A(2)$ and $d_A(1)$ is less than 1 ppm, confirmation of the revised structure by X-ray analysis is desirable. This was performed by the authors of article [36].

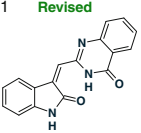
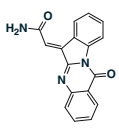
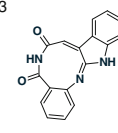
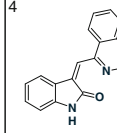
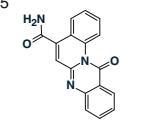
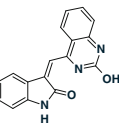
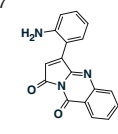
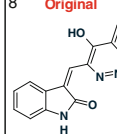
<p>1 Revised</p> 	<p>2</p> 	<p>3</p> 	<p>4</p> 
<p>$d_A(^{13}\text{C})$: 2.220 $d_N(^{13}\text{C})$: 3.116 $d_I(^{13}\text{C})$: 2.403 $\text{max_}d_A(^{13}\text{C})$: 9.910</p>	<p>$d_A(^{13}\text{C})$: 2.496 $d_N(^{13}\text{C})$: 2.884 $d_I(^{13}\text{C})$: 3.499 $\text{max_}d_A(^{13}\text{C})$: 11.240</p>	<p>$d_A(^{13}\text{C})$: 2.521 $d_N(^{13}\text{C})$: 3.305 $d_I(^{13}\text{C})$: 2.543 $\text{max_}d_A(^{13}\text{C})$: 6.080</p>	<p>$d_A(^{13}\text{C})$: 2.651 $d_N(^{13}\text{C})$: 2.454 $d_I(^{13}\text{C})$: 3.235 $\text{max_}d_A(^{13}\text{C})$: 10.810</p>
<p>5</p> 	<p>6</p> 	<p>7</p> 	<p>8 Original</p> 
<p>$d_A(^{13}\text{C})$: 2.840 $d_N(^{13}\text{C})$: 4.464 $d_I(^{13}\text{C})$: 3.243 $\text{max_}d_A(^{13}\text{C})$: 16.910</p>	<p>$d_A(^{13}\text{C})$: 2.922 $d_N(^{13}\text{C})$: 2.576 $d_I(^{13}\text{C})$: 2.346 $\text{max_}d_A(^{13}\text{C})$: 9.550</p>	<p>$d_A(^{13}\text{C})$: 2.952 $d_N(^{13}\text{C})$: 3.443 $d_I(^{13}\text{C})$: 3.332 $\text{max_}d_A(^{13}\text{C})$: 7.900</p>	<p>$d_A(^{13}\text{C})$: 3.078 $d_N(^{13}\text{C})$: 3.258 $d_I(^{13}\text{C})$: 2.457 $\text{max_}d_A(^{13}\text{C})$: 8.310</p>

Fig. 4.88 Schizocommunicin: Eight top structures of the ranked output file from the second run

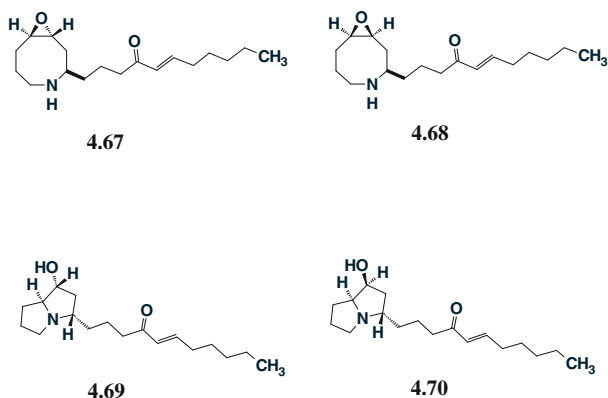
This example shows that application of the StrucEluc system is capable of significantly influencing the strategy of structure revision based on chemical synthesis.

4.33 Epophelmin A

In an earlier review article [37], a wide variety of examples of structure revision utilizing StrucEluc were described. Here we will consider one of them which demonstrates how utilizing the software allows researchers to avoid complex multistaged synthesis to refute a wrong proposed structure and to prove the revised one.

Sakano et al. [38] reported the isolation of the novel lanosterol synthase inhibitors *epohelmins* A (**4.67**) and B (**4.68**). The structures were determined by detailed spectroscopic analysis (including 2D NMR spectra) and proposed to be novel 9-oxa-4-azabicyclo[6.1.0]-nonanes. However, these structure assignments raised doubts based on both chemical and spectroscopic grounds [39].

Snider and Gao [39] comprehensively analyzed both the spectral and chemical aspects of the study of *Epophelmins* A and B and suggested structures **4.69** and **4.70** as being more appropriate hypotheses.



To validate their suggestions, the authors [39] developed an *eight step* synthesis of *Epohelmin A* (**4.67**) and a *11 step* synthesis of *Epohelmin B* (**4.68**).

The ^1H and ^{13}C NMR spectra of **4.69** and **4.70** were identical to those reported for *Epohelmin A* (**4.67**) and *Epohelmin B* (**4.68**), and the revised structures of these compounds were therefore unambiguously established via chemical synthesis.

The 2D NMR spectra of the investigated compounds were not available from [38, 39], so it was only possible to predict and compare the ^{13}C NMR spectra of the competing structures **4.67** and **4.69**, along with a review of the discrepancies between the predicted and experimental data (see Table 4.33).

Table 4.33 unambiguously shows that structure **4.69** is superior to structure **4.67**. It is likely that if 2D NMR data were available to the researchers then application of StrucEluc would deliver the correct structure very quickly and structure **4.67** would immediately be rejected by the program due to the very large deviations, especially with a $d_{(\text{max})}$ value of 21.4 ppm. Multistep syntheses would also not be necessary to resolve the structural problem. However, at the same time the method of synthesizing *Epohelmin A* and *Epohelmin B* would not be developed! This contradictory peculiarity of structural revision work was emphasized strongly in a subsequent review article [40] where a number of striking examples were given.

We will demonstrate how Structure Elucidator would allow a researcher to not only reject the erroneous structure from the very beginning, but to easily find the correct one even if the 2D NMR spectra are not available.

Table 4.33 Epohelmin A: A comparison of the deviations and R^2 (squared correlation coefficient) values calculated for competing structures **4.67** and **4.69**

Structure	$d_{(\text{HOSE})}$ (ppm)	$d_{(\text{NN})}$ (ppm)	$d_{(\text{max})}$	$R^2_{(\text{HOSE})}$	$R^2_{(\text{NN})}$
4.67	4.00	4.17	21.4	0.978	0.980
4.69	1.23	1.25	4.84	0.999	0.999

Stereobonds were taken into account

Table 4.34 Epohelmin A:
¹³C NMR spectrum

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n
C1	51.9	53.71	CH ₂
C2	24.2	24.86	CH ₂
C3	28.4	26.16	CH ₂
C4	73.9	69.06	CH
C5	74.3	74.16	CH
C6	40.9	36.94	CH ₂
C7	67.3	65.31	CH
C8	31.2	32.61	CH ₂
C9	21.1	20.98	CH ₂
C10	39.1	41.03	CH ₂
C11	199.6	201.21	C
C12	130.1	130.36	CH
C13	148.2	148.62	CH
C14	32.5	32.67	CH ₂
C15	27.7	27.87	CH ₂
C16	31.3	31.38	CH ₂
C17	22.4	22.33	CH ₂
C18	13.9	13.9	CH ₃

The User ¹³C NMR spectrum (carbon atoms with attached hydrogens and the associated ¹³C chemical shifts) published in [38] was input into Structure Elucidator (Table 4.34), and the **Molecular Connectivity Diagram (MCD)** was created. The carbon atoms included into the side chains of both the original and revised structures have very precisely predicted ¹³C chemical shifts, therefore it is clear that the differences in calculated deviations for these structures must be due to differences in their cyclic fragments.

With this in mind the atoms included into the side chain were connected by hand on the MCD to complete the common part of both structures (Fig. 4.89).

Structure generation from the MCD gave the following results: $k = 10,134 \rightarrow 4,180 \rightarrow 167$, $t_g = 5$ s. Then ¹³C chemical shifts were predicted for the structures in the output file, and the structures were ranked by chemical shift deviations. The most interesting structures of the ranked file are presented in Fig. 4.90.

Figure 4.90 shows that the correct (revised) structure was selected as the most probable, and it is also distinguished from the other isomers of the family (structures ranked as 2, 3, 24, and 61). Three similar isomers of the wrong structure family occupy positions 86, 112, and 119, while structure **4.67** originally proposed by authors [38] is 112th (!) in the ranked file. The cyclic parts of the original (**4.71**)

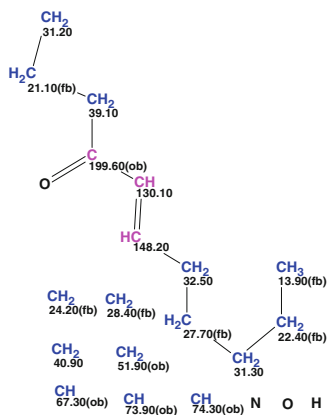
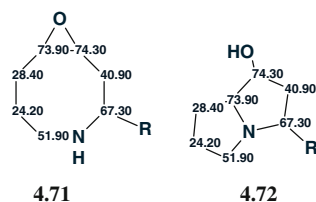


Fig. 4.89 Epohelmin A: The MCD, including the manually modified hydrocarbon chain common to both revised and original structures

1 Revised 	2 	3 	24
$d_A(^{13}\text{C})$: 1.216 $d_N(^{13}\text{C})$: 1.186 $d_I(^{13}\text{C})$: 1.345 $\text{max}_dA(^{13}\text{C})$: 4.840	$d_A(^{13}\text{C})$: 1.466 $d_N(^{13}\text{C})$: 1.814 $d_I(^{13}\text{C})$: 2.150 $\text{max}_dA(^{13}\text{C})$: 7.370	$d_A(^{13}\text{C})$: 1.513 $d_N(^{13}\text{C})$: 1.559 $d_I(^{13}\text{C})$: 1.590 $\text{max}_dA(^{13}\text{C})$: 5.810	$d_A(^{13}\text{C})$: 2.181 $d_N(^{13}\text{C})$: 2.681 $d_I(^{13}\text{C})$: 2.380 $\text{max}_dA(^{13}\text{C})$: 8.900
61 	86 	112 Original 	119
$d_A(^{13}\text{C})$: 2.818 $d_N(^{13}\text{C})$: 3.231 $d_I(^{13}\text{C})$: 3.358 $\text{max}_dA(^{13}\text{C})$: 8.470	$d_A(^{13}\text{C})$: 3.463 $d_N(^{13}\text{C})$: 3.968 $d_I(^{13}\text{C})$: 3.582 $\text{max}_dA(^{13}\text{C})$: 20.050	$d_A(^{13}\text{C})$: 4.117 $d_N(^{13}\text{C})$: 4.174 $d_I(^{13}\text{C})$: 3.968 $\text{max}_dA(^{13}\text{C})$: 20.890	$d_A(^{13}\text{C})$: 4.493 $d_N(^{13}\text{C})$: 4.480 $d_I(^{13}\text{C})$: 4.211 $\text{max}_dA(^{13}\text{C})$: 19.310

Fig. 4.90 Epohelmin A: Portions of the ranked structure list following structure generation

and revised (**4.72**) molecules along with the carbon chemical shift assignments are shown below for comparison.



This example clearly demonstrates that even with only ^{13}C NMR data, Structure Elucidator can be used to refute incorrect structures and confirm the correct one. This method also avoids the need for time-consuming and costly multistep synthesis to confirm the correct structure. Although it is obviously eventually beneficial to confirm a feasible synthesis pathway, this information is superfluous in terms of the work of quickly obtaining the correct structure.

References

- Motti CA, Thomas-Hall P, Hagiwara KA, Simmons CJ, Willis R, Wright AD (2014) Accelerated identification of halogenated monoterpenes from Australian specimens of the red algae *plocamium hamatum* and *plocamium costatum*. *J Nat Prod* 77:1193–1120. doi:[10.1021/np500059h](https://doi.org/10.1021/np500059h)
- Wang CQ, Wang L, Fan CL, Zhang DM, Huang XJ, Jiang RW, Bai LL, Shi JM, Wang Y, Ye WC (2012) Ilelic acids A and B, two unusual triterpenes with a seven-membered ring from *Ilex latifolia*. *Org Lett* 14(16):4102–4105. doi:[10.1021/ol301745b](https://doi.org/10.1021/ol301745b)
- Che Q, Zhu T, Qi X, Mandi A, Kurtan T, Mo X, Li J, Gu Q, Li D (2012) Hybrid isoprenoids from a reeds rhizosphere soil derived actinomycete *Streptomyces* sp. CHQ-64. *Org Lett* 14(13):3438–3441. doi:[10.1021/ol301396h](https://doi.org/10.1021/ol301396h)
- Ding JH, Feng T, Li ZH, Yang XY, Guo H, Yin X, Wang GQ, Liu JK (2012) Trefolane A, a sesquiterpenoid with a new skeleton from cultures of the basidiomycete *Tremella foliacea*. *Org Lett* 14(18):4976–4978. doi:[10.1021/ol302184r](https://doi.org/10.1021/ol302184r)
- Park HB, Kim YJ, Lee JK, Lee KR, Kwon HC (2012) Spirobacillenes A and B, unusual spirocyclopentenones from *Lysinibacillus fusiformis* KMC003. *Org Lett* 14(19):5002–5005. doi:[10.1021/ol302115z](https://doi.org/10.1021/ol302115z)
- Wang XJ, Zhang GJ, Zhuang PY, Zhang Y, Yu SS, Bao XQ, Zhang D, Yuan YH, Chen NH, Ma SG, Qu J, Li Y (2012) Lycojaponicumins A-C, three alkaloids with an unprecedented skeleton from *Lycopodium japonicum*. *Org Lett* 14(10):2614–2617. doi:[10.1021/ol3009478](https://doi.org/10.1021/ol3009478)
- Wang XJ, Liu YB, Li L, Yu SS, Lv HN, Ma SG, Bao XQ, Zhang D, Qu J, Li Y (2012) Lycojaponicumins D and E: two new alkaloids from *Lycopodium japonicum*. *Org Lett* 14(22):5688–5691. doi:[10.1021/ol302701y](https://doi.org/10.1021/ol302701y)
- San Feliciano A, Medarde M, Miguel del Corral JM, Aramburu A, Gordaliza M, Barrero AF (1989) *Tetrahedron Lett* 30:2851
- Lodewyk MW, Soldi C, Jones PB, Olmstead MM, Rita J, Shaw JT, Tantillo DJ (2012) The correct structure of aquatolide-experimental validation of a theoretically-predicted structural revision. *J Am Chem Soc* 134(45):18550–18553. doi:[10.1021/ja3089394](https://doi.org/10.1021/ja3089394)

10. Zou J, Pan L, Li Q, Zhao J, Pu J, Yao P, Gong N, Lu Y, Kondratyuk TP, Pezzuto JM, Fong HH, Zhang H, Sun H (2011) Rubesanolides A and B: diterpenoids from *Isodon rubescens*. *Org Lett* 13(6):1406–1409. doi:[10.1021/ol200086k](https://doi.org/10.1021/ol200086k)
11. Pretsch E, Bühlmann P, Affolter C (2000) Structure determination of organic compounds—tables of spectral data. Springer, Berlin
12. Yang F, Hamann MT, Zou Y, Zhang MY, Gong XB, Xiao JR, Chen WS, Lin HW (2012) Antimicrobial metabolites from the Paracel Islands sponge *Agelas mauritiana*. *J Nat Prod* 75(4):774–778. doi:[10.1021/np2009016](https://doi.org/10.1021/np2009016)
13. Abe S, Tanaka N, Kobayashi J (2012) Prenylated acylphloroglucinols, chipericumins A-D, from *Hypericum chinense*. *J Nat Prod* 75(3):484–488. doi:[10.1021/np200741x](https://doi.org/10.1021/np200741x)
14. Matsumori N, Hiradate Y, Shibata H, Oishi T, Shimma S, Toyoda M, Hayashi F, Yoshida M, Murata M, Morisawa M (2013) A novel sperm-activating and attracting factor from the ascidian *Ascidia sydneiensis*. *Org Lett* 15(2):294–297. doi:[10.1021/ol303172n](https://doi.org/10.1021/ol303172n)
15. Hu Y, Wang K, MacMillan JB (2013) Hunanamycin A, an antibiotic from a marine-derived *Bacillus humanensis*. *Org Lett* 15(2):390–393. doi:[10.1021/ol303376c](https://doi.org/10.1021/ol303376c)
16. Zhang D, Ge H, Xie D, Chen R, Zou JH, Tao X, Dai J (2013) Periconiasins A-C, new cytotoxic cytochalasans with an unprecedented 9/6/5 tricyclic ring system from endophytic fungus *Periconia* sp. *Org Lett* 15(7):1674–1677. doi:[10.1021/ol400458n](https://doi.org/10.1021/ol400458n)
17. Cao M, Zhang Y, He H, Li S, Huang S, Chen D, Tang G, Di Y, Hao X (2012) Daphnmacromines A-J, alkaloids from *Daphniphyllum macropodum*. *J Nat Prod* 75(6):1076–1082. doi:[10.1021/np200960z](https://doi.org/10.1021/np200960z)
18. Wang SJ, Li YX, Bao L, Han JJ, Yang XL, Li HR, Wang YQ, Li SJ, Liu HW (2012) Eryngiolide A, a cytotoxic macrocyclic diterpenoid with an unusual cyclododecane core skeleton produced by the edible mushroom *Pleurotus eryngii*. *Org Lett* 14(14):3672–3675. doi:[10.1021/ol301519m](https://doi.org/10.1021/ol301519m)
19. Liao WY, Shen CN, Lin LH, Yang YL, Han HY, Chen JW, Kuo SC, Wu SH, Liaw CC (2012) Asperjinone, a nor-neolignan, and terrein, a suppressor of ABCG2-expressing breast cancer cells, from thermophilic *Aspergillus terreus*. *J Nat Prod* 75(4):630–635. doi:[10.1021/np200866z](https://doi.org/10.1021/np200866z)
20. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ (2013) Structure revision of asperjinone using computer-assisted structure elucidation methods. *J Nat Prod* 76:113–116. doi:[10.1021/np300218g](https://doi.org/10.1021/np300218g)
21. Okanya PW, Mohr KI, Gerth K, Jansen R, Muller R (2011) Marinoquinolines A-F, pyrroloquinolines from *Ohtaekwangia kribbensis* (Bacteroidetes). *J Nat Prod* 74(4):603–608. doi:[10.1021/np100625a](https://doi.org/10.1021/np100625a)
22. Luo X, Shi YM, Luo RH, Luo SH, Wang RR, Li SH, Zheng YT, Du X, Xiao WL, Pu JX, Sun HD (2012) Schilancitrilactones A-C: three unique nortriterpenoids from *Schisandra lancifolia*. *Org Lett* 14(5):1286–1289. doi:[10.1021/ol300099e](https://doi.org/10.1021/ol300099e)
23. Ponnappalli MG, Annam S, Ravirala S, Sukki S, Ankireddy M, Tuniki VR (2012) Unusual isomeric corniculatolides from mangrove, *Aegiceras corniculatum*. *J Nat Prod* 75(2):275–279. doi:[10.1021/np200789s](https://doi.org/10.1021/np200789s)
24. Guo Z, She Z, Shao C, Wen L, Liu F, Zheng Z, Lin Y (2007) ¹H and ¹³C NMR signal assignments of Paecilin A and B, two new chromone derivatives from mangrove endophytic fungus *Paecilomyces* sp. (tree 1-7). *Magn Reson Chem* 45:777–780. doi:[10.1002/mrc.2035](https://doi.org/10.1002/mrc.2035)
25. Elyashberg ME, Williams AJ, Blinov KA (2012) Contemporary computer-assisted approaches to molecular structure elucidation, vol 1. New developments in NMR. RSC Publishing, Cambridge
26. Zhao BX, Wang Y, Zhang DM, Huang XJ, Bai LL, Yan Y, Chen JM, Lu TB, Wang YT, Zhang QW, Ye WC (2012) Virosaines A and B, two new birdcage-shaped Securinega alkaloids with an unprecedented skeleton from *Flueggea virosa*. *Org Lett* 14(12):3096–3099. doi:[10.1021/ol301184j](https://doi.org/10.1021/ol301184j)
27. Le Goff G, Martin MT, Servy C, Cortial S, Lopes P, Bialecki A, Smadja J, Ouazzani J (2012) Isolation and characterization of alpha, beta-unsaturated gamma-lactono-hydrazides from *Streptomyces* sp. *J Nat Prod* 75(5):915–919. doi:[10.1021/np300026p](https://doi.org/10.1021/np300026p)

28. Reina M, Ruiz-Mesia W, Lopez-Rodriguez M, Ruiz-Mesia L, Gonzalez-Coloma A, Martinez-Diaz R (2012) Indole alkaloids from *Geissospermum reticulatum*. *J Nat Prod* 75(5):928–934. doi:[10.1021/np300067m](https://doi.org/10.1021/np300067m)
29. Wyche TP, Hou Y, Vazquez-Rivera E, Braun D, Bugni TS (2012) Peptidolipins B-F, antibacterial lipopeptides from an ascidian-derived *Nocardia* sp. *J Nat Prod* 75(4):735–740. doi:[10.1021/np300016r](https://doi.org/10.1021/np300016r)
30. Chen Z, Song Y, Chen Y, Huang H, Zhang W, Ju J (2012) Cyclic heptapeptides, cordyheptapeptides C–E, from the marine-derived fungus *acremonium Persicinum* SCSIO 115 and their cytotoxic activities. *J Nat Prod* 75(6):1215–1219
31. Sorres J, Martin MT, Petek S, Levaïque H, Cresteil T, Ramos S, Thoison O, Debitus C, Al-Mourabit A (2012) Pipestelides A–C: cyclodepsipeptides from the Pacific marine sponge *Pipestela candelabra*. *J Nat Prod* 75(4):759–763. doi:[10.1021/np200714m](https://doi.org/10.1021/np200714m)
32. Dai J, Liu Y, Zhou YD, Nagle DG (2007) Cytotoxic metabolites from an Indonesian sponge *Lendenfeldia* sp. *J Nat Prod* 70(11):1824–1826. doi:[10.1021/np070337f](https://doi.org/10.1021/np070337f)
33. Podlesny EE, Kozłowski MC (2012) Structural reassignment of a marine metabolite from a binaphthalenetetrol to a tetrabrominated diphenyl ether. *J Nat Prod* 75(6):1125–1129. doi:[10.1021/np300141t](https://doi.org/10.1021/np300141t)
34. Chen M, Lin S, Li L, Zhu C, Wang X, Wang Y, Jiang B, Wang S, Li Y, Jiang J, Shi J (2012) Enantiomers of an indole alkaloid containing unusual dihydrothiopyran and 1,2,4-thiadiazole rings from the root of *Isatis indigotica*. *Org Lett* 14(22):5668–5671. doi:[10.1021/ol302660t](https://doi.org/10.1021/ol302660t)
35. Lorente A, Pla D, Canedo LM, Albericio F, Alvarez M (2010) Isolation, structural assignment, and total synthesis of barmumycin. *J Org Chem* 75(24):8508–8515. doi:[10.1021/jo101834c](https://doi.org/10.1021/jo101834c)
36. Uehata K, Kimura N, Hasegawa K, Arai S, Nishida M, Hosoe T, Kawai K, Nishida A (2013) Total synthesis of schizocommunin and revision of its structure. *J Nat Prod* 76(11):2034–2039. doi:[10.1021/np400263f](https://doi.org/10.1021/np400263f)
37. Elyashberg M, Williams AJ, Blinov K (2010) Structural revisions of natural products by computer-assisted structure elucidation (CASE) systems. *Nat Prod Rep* 27(9):1296–1328. doi:[10.1039/c002332a](https://doi.org/10.1039/c002332a)
38. Sakano Y, Shibuya M, Yamaguchi Y, Masuma R, Tomoda H, Omura S, Ebizuka Y (2004) Epohelmins A and B, novel lanosterol synthase inhibitors from a fungal strain FKI-0929. *J Antibiot (Tokyo)* 57(9):564–568
39. Snider BB, Gao X (2005) Structure revision and syntheses of epohelmins A and B. *Org Lett* 7(20):4419–4422. doi:[10.1021/ol0516061](https://doi.org/10.1021/ol0516061)
40. Nicolaou KC, Snyder SA (2005) Chasing molecules that were never there: misassigned natural products and the role of chemical synthesis in modern structure elucidation. *Angew Chem Int Ed* 44:1012–1044. doi:[10.1002/anie.200460864](https://doi.org/10.1002/anie.200460864)

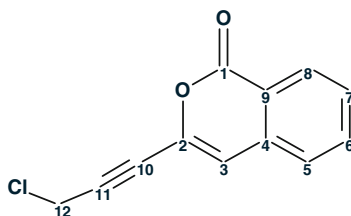
Chapter 5

Problems Solved Using Fuzzy Structure Generation

Abstract A number of more challenging problems (33 items total) are collected in this chapter. These problems are solved using *Fuzzy Structure Generation (FSG)*—a sophisticated approach which solves problems under the condition that an unknown number of correlations of unknown “nonstandard” lengths (of 4, 5, and more bonds) are present in a full set of 2D NMR data. The chapter shows how this approach allows the researcher to solve very complex problems including those that initially seemed unresolvable. The FSG approach significantly enhances the ability of a scientist to perform structure elucidation from fuzzy, complex, and contradictory 2D NMR data. The problems allow the reader to see that Structure Elucidator (the most advanced modern expert system) is not a robot intending to exclude human expertise from the process of structure elucidation, but is a powerful amplifier of human intelligence, an engine for inferring all logical corollaries (structures) from NMR spectroscopic data, and fuses spectrum-structure knowledge with assumptions introduced by the chemist.

5.1 Gymnopalynes A

Novel antibiotics with unusual scaffolds and new modes of actions are urgently needed since antibiotic resistance is constantly increasing. Most of the known antibiotics are derived from fungal or microbial cultures. In the course of the search for novel antibiotics, Thongbai et al. [1] have focused on tropical basidiomycetes from Asia. In an antimicrobial screen, extracts prepared from submerged cultures of this strain showed prominent activity. The research led to the isolation of the *unprecedented* antimicrobial metabolite gymnopalynes A (**5.1**) whose structure was used in this book earlier to illustrate the main features of a molecular connectivity diagram (MCD) (see Sect. 2.1.3).



5.1

Though the molecule is small and relatively simple, elucidation of its structure with the aid of the CASE approach is not straightforward. With this in mind, we used this problem to illustrate some nuances associated with the utilization of the Structure Elucidation software.

Gymnopalynes A was obtained as a colorless oil. The presence of a chlorine atom was indicated by its characteristic isotopic pattern in the HRESIMS that provided the molecular formula of **5.1** as $C_{12}H_7O_2Cl$, implying nine degrees of unsaturation. Absorption bands were observed in the IR spectrum at 1,720 and 1,600 cm^{-1} suggesting the presence of a carbonyl and a benzene rings correspondingly. As the number of skeletal atoms is twice as many as the number of hydrogens contained in the molecular formula, the problem can be considered as being challenging in accordance with Crews' rule. To elucidate the structure of gymnopalynes A, the spectroscopic NMR data presented in Table 5.1 were used.

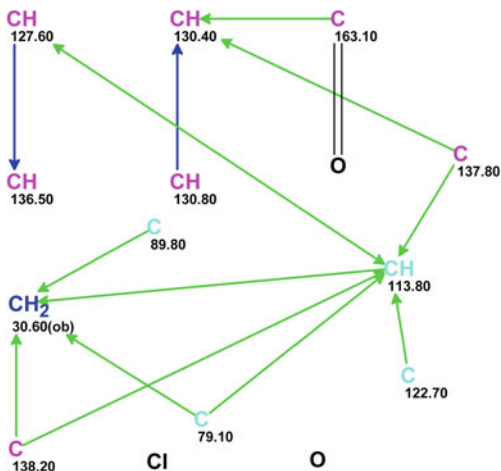
As no sign of the $C\equiv C$ bond presence was mentioned in [1], the MCD (Fig. 5.1) was first created with the option "Allow sp Carbons" deselected.

MCD overview Four carbon atoms C 79.10, C 89.80, CH 113.80, and C 122.70 are colored in light blue indicating that hybridization of these atoms is sp^3 or sp^2 . To decrease the number of conceivable structures during structure generation, two

Table 5.1 Gymnopalynes A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	163.1	161.07	C	–	–	–	–
C2	138.2	134.78	C	–	–	–	–
C3	113.8	112.80	CH	7.02	u	–	C4, C2, C5, C9, C10
C4	137.8	136.06	C	–	–	–	–
C5	127.6	126.00	CH	7.6	u	7.82	C3
C6	136.5	134.9	CH	7.82	u	7.6	–
C7	130.8	129.8	CH	7.62	u	8.24	–
C8	130.4	129.35	CH	8.24	u	7.62	C1, C4
C9	122.7	121.08	C	–	–	–	–
C10	79.1	75.98	C	–	–	–	–
C11	89.8	91.17	C	–	–	–	–
C12	30.6	31.34	CH_2	4.55	u	–	C2, C3, C11, C10

Fig. 5.1 Gymnopalynes A: The MCD was created assuming that triple bonds are not expected in the structure



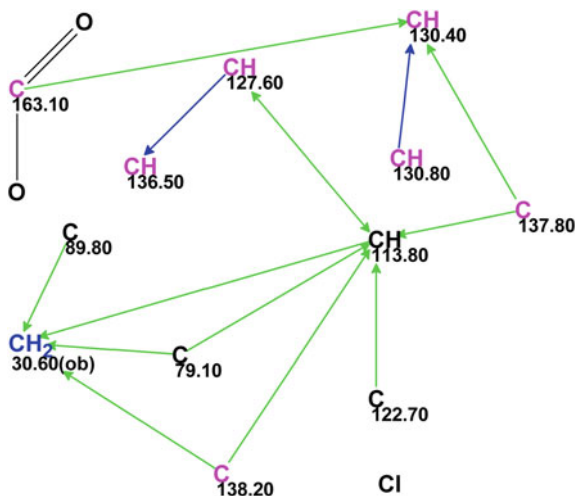
obvious constraints were introduced into MCD by the user: the atom CH₂ (30.60, 4.55) was marked with the label “ob” (i.e., a neighboring heteroatom, most probably a chlorine in this case, is obligatory) and the carbonyl group was drawn by hand at carbon C 163.10 (IR band at 1,720 cm⁻¹). Checking the MCD did not reveal the presence of any contradictions, so strict structure generation accompanied by ¹³C chemical shift prediction (with $d(^{13}\text{C}) = 5$ ppm as a threshold) was initiated. Generation was completed in a second to produce an empty structural file: all generated structures were rejected due to the huge values of average deviations that were calculated.

It was suggested that at least one latent (implicit) nonstandard correlation is present in the HMBC data, and FSG was performed with the options: $m = 1-20$, $a = 16$, “Stop generation when structures generated.” ¹³C chemical shift calculation was not switched on during the generation. Results: $k = 987,316 \rightarrow 8 \rightarrow 4$, $t_g = 10$ min and the structures generated are shown in Fig. 5.2.

<p>1</p>	<p>2</p>	<p>3</p>
<p>$d_A(^{13}\text{C}): 7.684$ $d_N(^{13}\text{C}): 9.974$ $d_I(^{13}\text{C}): 4.809$ $\text{max}_dA(^{13}\text{C}): 22.830$</p>	<p>$d_A(^{13}\text{C}): 11.452$ $d_N(^{13}\text{C}): 7.588$ $d_I(^{13}\text{C}): 4.726$ $\text{max}_dA(^{13}\text{C}): 38.130$</p>	<p>$d_A(^{13}\text{C}): 11.575$ $d_N(^{13}\text{C}): 9.614$ $d_I(^{13}\text{C}): 4.492$ $\text{max}_dA(^{13}\text{C}): 38.260$</p>

Fig. 5.2 Gymnopalynes A: The ranked output structure file obtained by FSG with *sp* carbon atom hybridization forbidden

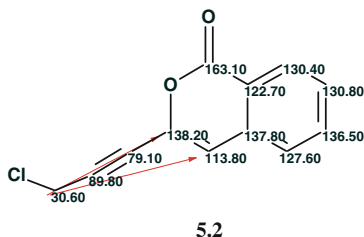
Fig. 5.3 Gymnopalynes A:
Slightly edited MCD



The figure convincingly demonstrates that the solution is wrong (very large average deviations and the structures appear to be extremely exotic). Therefore, it was suggested that the possibility of the presence of triple bonds in the molecule should be investigated.

The MCD was created anew with the option “Allow sp Carbons” selected. A slightly edited MCD is presented in Fig. 5.3.

In the MCD four carbon atoms C 79.10, C 89.80, CH 113.80, and C 122.70 are colored in black now indicating that the hybridization states of these atoms were not set automatically because all types of hybridization (sp^3 , sp^2 and sp) are allowed for them if the presence of triple bonds is permitted. All possible hybridizations of these atoms will be tried during the structure generation. To decrease the number of conceivable structures, an obvious ester group was drawn by hand at carbon C 163.10. Checking the MCD again did not reveal the presence of any contradictions and, as above, strict structure generation combined with ^{13}C chemical shift calculation was initiated which also quickly produced an empty structural file. Therefore, FSG was run with the following options: $m = 1-20$, $a = 16$, “Stop generation when structures generated.” Results: $k = 17,967 \rightarrow 4 \rightarrow 1$, $t_g = 24$ s, 2 from 13 connectivities were extended during generation, 78 from 78 possible connectivity combinations were used during generation. The single structure coincided with the proposed structure of gymnopalynes A and its correctness was confirmed by the small values of the average deviations ($d_A(^{13}\text{C}) = 1.7$ ppm). The assignments of the ^{13}C chemical shifts along with the nonstandard HMBC connectivities (their lengths are of three and four skeletal bonds) are shown in structure 5.2:



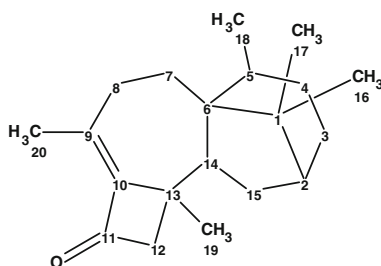
It was interesting to see if the presence of these nonstandard correlations (NSCs) could be suggested from a visual analysis of the HMBC pattern. Figure 5.4 shows part of the gymnopalynes A HMBC experimental spectrum (see SI to [1]).

Figure 5.4 shows that the intensities of two peaks corresponding to NSCs (in squares) are comparable with those belonging to other peaks. This suggests that the peaks indicating correlations between 4.55–138.2 and 4.55–113.8 ppm are unlikely to be of nonstandard length.

FSG led to a single and unambiguous solution to the problem without any suggestions regarding the presence or absence of NSCs in the 2D NMR data. At the same time, the presence of a triple bond in a molecule was detected by the expert system without any additional experiments. Commonly, this would require the application of Raman spectroscopy.

5.2 Harzianone

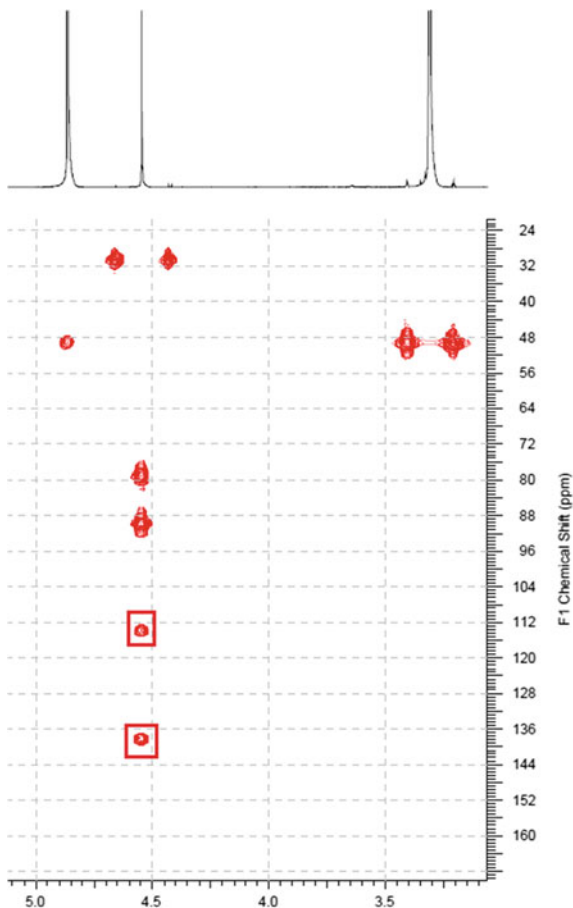
Miao et al. [2] isolated and identified a new harziane diterpene—Harzianone (**5.3**), containing a *unique* tetracyclic scaffold (fused four-, five-, six-, and seven-membered carbon rings).



5.3

The molecular formula was determined to be C₂₀H₃₀O on the basis of HRESIMS (*m/z* 286.2304 [M]⁺, calculated for C₂₀H₃₀O and 286.2297), requiring six degrees of unsaturation. The ¹H NMR spectrum along with the HSQC data displayed four

Fig. 5.4 Gymnopalynes A:
A fragment of the HMBC
experimental spectrum.
The peaks corresponding to the
nonstandard connectivities are
marked by *squares*

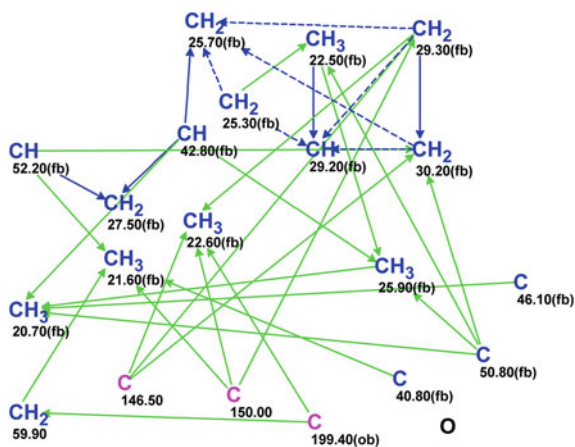


methyl singlets and one methyl doublet. The ^{13}C and DEPT NMR spectra demonstrated the presence of five methyls, six methylenes, three methines, and six nonprotonated carbons. The IR spectrum contained a strong carbonyl band at $1,740\text{ cm}^{-1}$ and a band of medium intensity at $1,660\text{ cm}^{-1}$ that is most probably associated with the stretching vibrations of the carbon double bond. The tabulated 1D NMR and HSQC data and *selected* ^1H - ^1H COSY and ^1H - ^{13}C HMBC correlations depicted by arrows on the target structure in the source article are shown in Table 5.2. The MCD is presented in Fig. 5.5.

MCD overview The MCD contains many COSY and HMBC ambiguous connectivities. For instance, all connectivities associated with CH_3 (20.70) and CH_2 (25.30) are ambiguous. The results from the accidental degeneration of ^1H chemical shifts (H17, H18) and (H4, H7, H8) in the ^1H spectrum are as shown below:

Table 5.2 Harzianone: Spectroscopic data

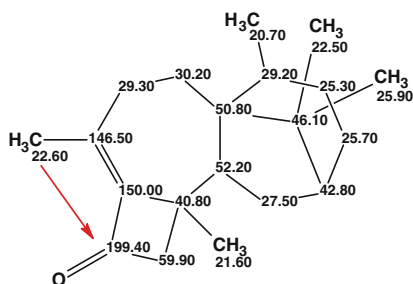
Label	δC	δC_{calc}	CH _n	δH	M(J)	COSY	HMBC
C1	46.1	43.1	C	–	–	–	–
C2	42.8	45.95	CH	1.66	u	1.31, 1.37	–
C3	25.7	28.61	CH ₂	1.94	u	–	–
C3	25.7	28.61	CH ₂	1.31	u	1.26, 1.66	–
C4	25.3	32.48	CH ₂	2.08	u	–	–
C4	25.3	32.48	CH ₂	1.26	u	–	–
C5	29.2	35.88	CH	2.42	u	1.04, 1.26	–
C6	50.8	58.59	C	–	–	–	–
C7	30.2	27.32	CH ₂	1.83	u	1.89	C9, C6, C14
C7	30.2	27.32	CH ₂	1.26	u	1.31, 2.42	–
C8	29.3	30.64	CH ₂	1.89	u	1.83	C9, C10
C8	29.3	30.64	CH ₂	1.26	u	–	–
C9	146.5	146	C	–	–	–	–
C10	150	129.67	C	–	–	–	–
C11	199.4	199.67	C	–	–	–	–
C12	59.9	57.04	CH ₂	2.38	u	–	C11
C12	59.9	57.04	CH ₂	2.53	u	–	–
C13	40.8	39.03	C	–	–	–	–
C14	52.2	55	CH	2.16	u	1.37	–
C15	27.5	33.12	CH ₂	1.37	u	2.16, 1.66	–
C15	27.5	33.12	CH ₂	1.86	u	–	–
C16	25.9	25.05	CH ₃	0.85	s	–	C6, C2, C17
C17	22.5	23.05	CH ₃	1.04	s	2.42	C1, C6, C2, C5, C16, C4
C18	20.7	17.17	CH ₃	1.04	d	–	–
C19	21.6	23.51	CH ₃	1.49	s	–	C10, C12, C14, C13
C20	22.6	21.39	CH ₃	2.09	s	–	C11, C8, C9, C10

Fig. 5.5 Harzianone:
Molecular connectivity
diagram

20.7	CH ₃	H18	1.04
22.5	CH ₃	H17	1.04
25.3	CH ₂	H4	1.26
29.3	CH ₂	H8	1.26
30.2	CH ₂	H7	1.26

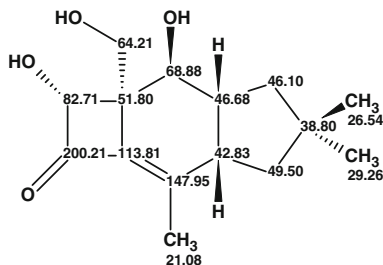
MCD checking for contradictions was completed with a program message which suggested the presence of at least one NSC in the 2D NMR data. In such a case, the first run can be successful if FSG is initiated in the mode **Determine Options Automatically**.

Automatic FSG with ¹³C chemical shift prediction and structural filtering with thresholds of $d < 4$ ppm and $d(\max) < 20$ ppm resulted in the generation of a single structure **5.4**, (deviations values: $d_A = 3.65$, $d_N = 3.58$ and $d_I = 2.96$ ppm), $t_g = 0.083$ s.



5.4

One HMBC connectivity was elongated by one bond during structure generation. It should be noted that deviations calculated for the output structure are relatively large, which is explained by the uniqueness of the target structure and the absence of molecules containing appropriate structural elements in the ACD/DB. The worst accuracy of the HOSE method based prediction was observed for carbon C10 at δC 150.0—the calculated value was 129.7 ppm. The chemical shift prediction protocol allowed us to establish the cause of the program failure: only one structure **5.5** (see below) was found in the ACD/DB that could be used for chemical shift prediction of the C10 atom.

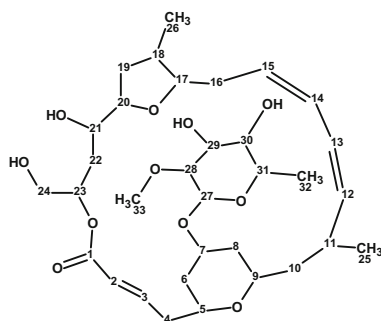


5.5

In structure **5.5**, the chemical shift of the corresponding carbon atom (neighboring with the carbonyl group) is 113.8 ppm, so this value was used for further calculations. However, because the solution to the problem contains only a single structure, the accuracy of chemical shift prediction played a minor role in this case.

5.3 Mandelalide A

Sikorska et al. [3] isolated and investigated the *unusual* polyketide macrolides mandelalides A–D which were isolated from a new species of *Lissoclinum ascidian*. Their planar structures were elucidated on submilligram samples by comprehensive analysis of 1D and 2D NMR data, supported by mass spectrometry. Here we will describe the structure elucidation of Mandelalide A (**5.6**).



5.6

The HRESIMS data for mandelalide A gave a pseudomolecular ion $[M+Na]^+$ at m/z 647.3394, which is consistent with a molecular formula of $C_{33}H_{52}O_{11}$, and implies eight degrees of unsaturation.

For structure elucidation, the authors [3] used 1D NMR spectra in combination with HSQC, HMBC, COSY, TOCSY, and ROESY data acquired at 700 (1H) and 175 (^{13}C) MHz on a 5 mm inverse cryogenic probe. Because the molecular formula contains a large number of hydrogen atoms to produce a rich HMBC spectrum, we input into the program only 1D, HSQC, and HMBC data (Table 5.3).

The MCD (slightly edited) is presented in Fig. 5.6.

MCD overview According to the characteristics of the ^{13}C and 1H chemical shifts (see Table 5.3), carbon atoms C(14.50)–C(43.10) were supplied with the property sp^3/ffb , while carbons C(59.10)–C(94.2) were assigned as sp^3/ob . Carbon C(167.4) was given the hybridization sp^2/ob . Figure 5.6 shows that the unknown molecule has to contain four hydroxyl groups and seven oxygen atoms connected to carbons only.

Checking the MCD gave the message “The minimum number of non-standard connectivities is 2.” Therefore, FSG was initiated with automatic option selection,

Table 5.3 Mandelalide A: NMR spectroscopic data

Label	δC	δC_{calc}	CH_n	δH	M/J	HMBC
C1	167.4	166.58	C	–	–	–
C2	123.1	124.78	CH	6.01	u	C1, C4
C3	147.1	146.47	CH	6.97	u	C1, C2, C4, C5
C4	38.8	38.28	CH ₂	2.36	u	C5, C6, C2, C3
C4	38.8	38.28	CH ₂	2.39	u	–
C5	73.9	69.16	CH	3.36	u	C3, C4
C6	37.6	38.46	CH ₂	1.2	u	C4, C5, C8
C6	37.6	38.46	CH ₂	2.02	u	C8, C7
C7	73.1	67.97	CH	3.82	u	C27, C9, C8
C8	39.7	39.45	CH ₂	1.87	u	–
C8	39.7	39.45	CH ₂	1.22	u	C6, C9, C7, C10
C9	72.5	67.6	CH	3.32	u	C7, C5
C10	43.1	42.64	CH ₂	1.21	u	C11, C25, C12
C10	43.1	42.64	CH ₂	1.51	u	C11, C8, C9, C25, C12
C11	34.2	33.89	CH	2.37	u	C12, C10, C25
C12	141.5	139.08	CH	5.45	u	C14, C11, C10, C25
C13	123.9	131.19	CH	6.28	u	C15, C14, C10, C11
C14	131.3	132.11	CH	6.05	u	C12, C16, C17, C13
C15	126.9	132.1	CH	5.28	u	C13, C17, C16
C16	31.1	38.96	CH ₂	2.28	u	C14, C17, C15, C18
C16	31.1	38.96	CH ₂	1.88	u	C14, C17, C15
C17	81	85.03	CH	3.98	u	C20, C15, C19
C18	37.3	34.79	CH	2.52	u	C16, C19, C17, C26
C19	36.8	37.25	CH ₂	2.01	u	C18, C17
C19	36.8	37.25	CH ₂	1.17	u	C21, C18, C20, C26
C20	83.2	82.17	CH	3.63	u	C22, C21
C21	73	72.58	CH	3.42	u	C22, C20, C23
C22	34.1	35.08	CH ₂	1.76	u	C24, C23
C22	34.1	35.08	CH ₂	1.46	u	C20, C21
C23	72.3	70.17	CH	5.23	u	C1
C24	66.1	64.67	CH ₂	3.81	u	–
C24	66.1	64.67	CH ₂	3.61	u	C22, C23
C25	18.3	17.5	CH ₃	0.85	u	C10, C11, C12
C26	14.5	15.36	CH ₃	1.03	u	C18, C17, C19
C27	94.2	95.8	CH	5.02	u	C29, C28, C31, C7
C28	80.8	85.58	CH	3.4	u	C30, C29, C33
C29	71.7	77.98	CH	3.68	u	–
C30	74.3	76.59	CH	3.34	u	C29, C32, C31
C31	68.1	70.74	CH	3.62	u	C29, C30
C32	17.7	15.91	CH ₃	1.27	u	C30, C31
C33	59.1	60.64	CH ₃	3.45	u	C28

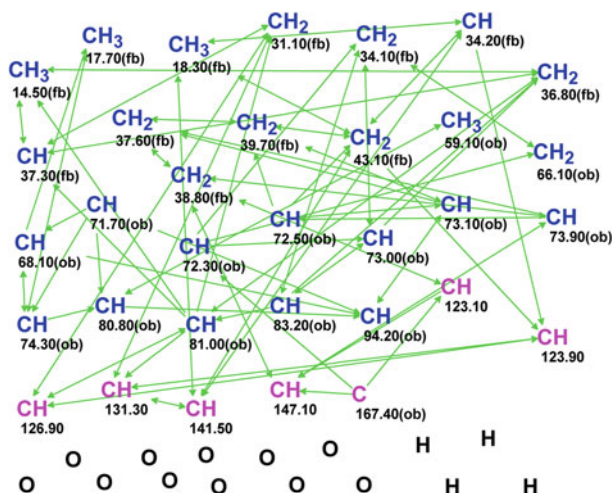
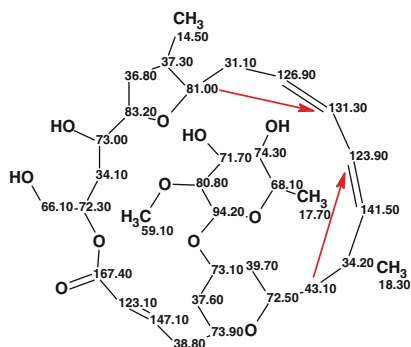


Fig. 5.6 Mandelalide A: HMBC MCD

which gave the following results: $k = 1$, $t_g = 3$ s, 2 from 63 correlations have been extended during generation and only 42 from 1953 ($\sim 2\%$) possible connectivity combinations were used.

The output structure **5.7** coincided with the structure of mandelalide A and the following deviation values were calculated: $d_A = 2.40$, $d_N = 1.78$, and $d_I = 1.56$ ppm. Automatic chemical shift assignment provided the same outcome suggested by authors [3].



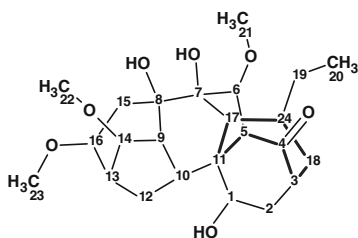
5.7

Therefore the structure of a new natural product—an *unusual* polyketide macrolide—was instantly and unambiguously elucidated using only HMBC

correlations. Two nonstandard connectivities and their lengths were identified by the system automatically. Multiplicities in the ^1H NMR spectrum were not used.

5.4 Puberunine

Aconitum barbatum Pers. var. puberulum Ledeb. (Subgen *Lycotconum*), a herb distributed in the northern part of China, as well as in the Siberia region of Mongolia and Russia, has been used over a long period of time in China as a folk medicine to treat rheumatism and pain. As a result of the study of this herb, Mu et al. [4] isolated six new C18-diterpenoid alkaloids including puberunine (**5.8**) that possesses an *unusual* rearranged 7-membered ring, highlighted by bold lines. This structural feature is *unprecedented* in the field of diterpenoid alkaloids.



5.8

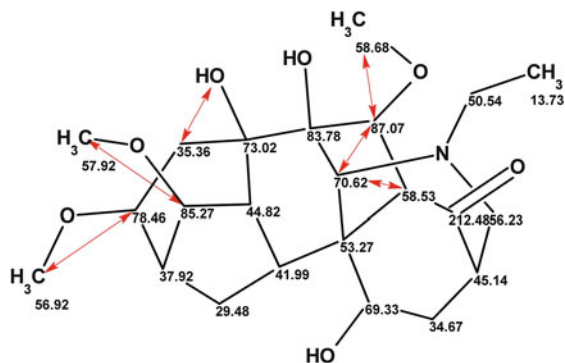
Puberunine, isolated as a colorless amorphous solid, was shown by HRESI-TOF-MS (m/z 438.2496 $[\text{M}+\text{H}]^+$, calculated 438.2486) to have the molecular formula $\text{C}_{23}\text{H}_{35}\text{NO}_7$.

Absorption bands in the IR spectrum indicate the presence of hydroxyl ($3,600\text{--}3,000\text{ cm}^{-1}$) and carbonyl ($1,712\text{ cm}^{-1}$) groups.

Spectroscopic data used for the structure elucidation of the unknown are collected in Table 5.4 and the MCD is shown in Fig. 5.7.

No user edits of the connectivities and atom properties were made. MCD checking revealed the presence of contradictions. FSG was therefore initiated using the mode **Determine Options Automatically**. The results were: $k = 480 \rightarrow 216 \rightarrow 25$, $t_g = 1$ s. It turned out that the best structure of the ranked file was characterized by large deviations, and in addition the length of one COSY correlation was 7 bonds. As five correlations were automatically elongated by the program, then the next run was performed with parameters $m = 6$, $a = 1$. Results: $k = 22 \rightarrow 10 \rightarrow 8$, $t_g = 3$ s, 820 of 8,145,060 combinations of connectivities were used during the Fuzzy Generation. The three top structures of the ranked output structural file are shown in Fig. 5.8.

Figure 5.8 shows that the first ranked structure is structure **5.8** and the ^{13}C chemical shift assignment is displayed on structure **5.9**. Six COSY NSCs detected by the program are marked by arrows.



5.9

Table 5.4 Puberunine: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	COSY	HMBC
C1	69.5	69.86	CH	3.67	2.53, 2.03	C17, C10
C2	39.3	36.01	CH ₂	2.53	3.67, 2.65	–
C2	39.3	36.01	CH ₂	2.03	3.67	C3, C4, C1
C3	48.8	45.69	CH	2.65	2.53, 3.07, 3.25	C2
C4	213.7	213.8	C	–	–	–
C5	60.8	54.83	CH	2.35	3.00, 4.14	C6, C17, C4
C6	92.9	87.5	CH	4.14	2.35, 3.00, 3.50	C8, C4, C7, C5
C7	87.7	88.11	C	–	–	–
C8	79.2	79.84	C	–	–	–
C9	44.2	44.25	CH	2.63	2.10, 3.70	C12, C13, C8
C10	48	43.27	CH	2.1	1.92, 2.27, 2.63	C8
C11	53	52.97	C	–	–	–
C12	28	29.34	CH ₂	2.27	2.1	C11, C14, C16
C12	28	29.34	CH ₂	1.92	2.45, 2.10	C16, C10
C13	39	38.07	CH	2.45	1.92, 3.70	C9
C14	83.6	84.77	CH	3.7	2.45, 2.63, 3.43	–
C15	34.3	38.16	CH ₂	1.75	3.23, 4.27	C8, C7, C16
C15	34.3	38.16	CH ₂	2.7	3.23	C16, C13, C9, C7, C8
C16	83.1	78.54	CH	3.23	1.75, 2.70, 3.37	C14, C12, C15, C13
C17	65.4	71.12	CH	3	2.35, 4.14	C19, C10, C6, C5, C8, C11, C18
C18	51.2	54.67	CH ₂	3.25	2.65	C17, C4
C18	51.2	54.67	CH ₂	3.07	2.65	C4, C17, C3
C19	50.4	49.28	CH ₂	2.96	–	C17, C18
C19	50.4	49.28	CH ₂	3.21	1.19	C17, C20, C18
C20	14.5	13.49	CH ₃	1.19	3.21	C19
C21	59.3	60.43	CH ₃	3.5	4.14	C6
C22	57.9	58.54	CH ₃	3.43	3.7	C14
C23	56.3	56.75	CH ₃	3.37	3.23	C16
O 1	100 ^a	–	OH	4.27	1.75	C15, C8

^a Fictitious ¹⁷O chemical shift

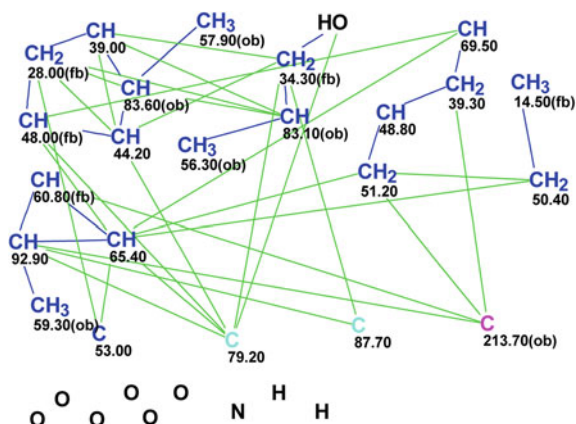


Fig. 5.7 Puberunine: Molecular connectivity diagram

1	2	3
$d_A(^{13}\text{C}): 2.151$ $d_I(^{13}\text{C}): 2.407$ $d_N(^{13}\text{C}): 2.483$ $\text{max}_dA(^{13}\text{C}): 5.970$	$d_A(^{13}\text{C}): 2.667$ $d_I(^{13}\text{C}): 4.195$ $d_N(^{13}\text{C}): 3.580$ $\text{max}_dA(^{13}\text{C}): 8.340$	$d_A(^{13}\text{C}): 3.101$ $d_I(^{13}\text{C}): 3.656$ $d_N(^{13}\text{C}): 3.202$ $\text{max}_dA(^{13}\text{C}): 7.060$

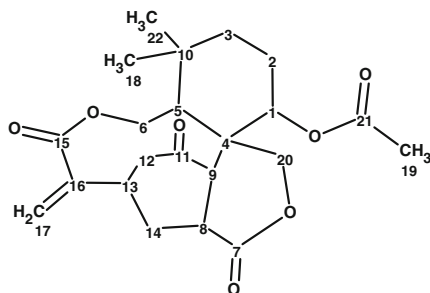
Fig. 5.8 Puberunine: Three top structures of the ranked output file

At the end of the article [4], the author underlines once more that puberunine (1) possesses an unprecedented skeleton containing a rearranged seven-membered ring (consisting of C3, C4, C5, C11, C17, N, and C18), in which the C18–C4 bond typically present in C18- and C19-diterpenoid alkaloids is missing and a new C18–C3 bond exists. This structure was determined by the program automatically and instantly.

5.5 Ternifolide A

Ternifolide A (**5.10**), a new diterpenoid featuring a *unique 10-membered lactone ring* formed between C6 and C15, was isolated from the leaves of *Isodon ternifolius* by Zou et al. [5]. The structure and absolute configuration of **5.10** was confirmed by

X-ray diffraction study. It is the first time that a diterpenoid having such configurations in the reported ent-kaurane diterpenoids was discovered and classified to a new diterpene type, ternifonane.



5.10

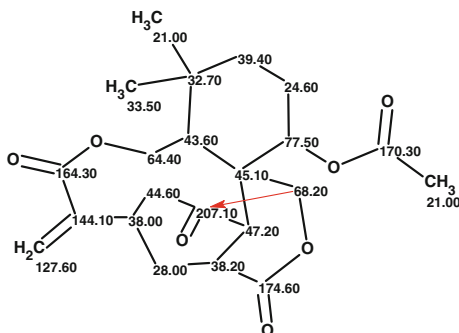
The molecular formula of compound **5.10** was determined as $C_{22}H_{28}O_7$ by HRESIMS ($[M+Na]^+$, 427.1731, calculated 427.1732), corresponding to nine degrees of unsaturation. The IR spectrum (see SI to [5]) shows the absorption bands at 3,450 (water adsorbed in KBr), 3,110, and 3,152 (probably symmetric and anti-symmetric stretching C–H vibrations of $=CH_2$ group), 1,739 and 1,683 (stretching vibrations of C=O groups), and $1,633\text{ cm}^{-1}$ (likely C=C stretching vibrations of double bond C=C). The statement of the authors that the 1,739 and 1,683 bands “indicate the existence of carbonyl groups for lactone and conjugated lactone, respectively” is not convincing without the molecular structure being determined beforehand.

To perform the computer-assisted structure elucidation of ternifolide A, tabulated ^{13}C and 1H NMR spectra were used in combination with the selected HMBC and COSY correlations presented on structure **5.10** (Fig. 5.9) and summarized in Table 5.5.

The MCD created from the data presented in Table 5.5 is shown in Fig. 5.10.

MCD overview Three atoms—C 32.7, C 45.1, and C 170.3—have no correlations, but the hybridizations of all carbon atoms were determined automatically. No user intervention was used to modify the atom properties. Checking the MCD for contradictions showed at least one nonstandard correlation is present in the HMBC data.

FSG was initiated in the mode **Determine Options Automatically**. Results: $k = 2$ Ternifolide A1, $t_g = 0.5$ s, one from 28 correlations have been extended during generation, 28 from 28 possible connectivity combinations were tried during generation. The calculated average deviations are: $d_A = 2.29$, $d_N = 2.37$ and $d_I = 2.92$ ppm. The output structure **5.11** was identical to the structure of ternifolide A with assigned ^{13}C chemical shifts and a single NSC is shown below:

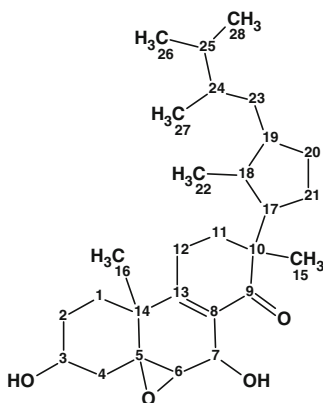


5.11

A new diterpenoid featuring a unique 10-membered lactone ring was therefore instantly and automatically identified using the StrucEluc software.

5.6 Strophasterol A

In a search for bioactive compounds from the mushroom *Stropharia rugosoannulata*, Wu et al. [6] discovered four novel steroids named strophasterols A, B, C, and D having a *very unique* and *unprecedented* carbon skeleton. Here we will describe the computerized structure determination of strophasterol A (5.12).



5.12

Strophasterol A was obtained as white crystals. Its molecular formula was determined to be $C_{28}H_{44}O_4$ by HRESIMS with m/z 467.3100 $[M+Na]^+$ (calculated as $C_{28}H_{44}NaO_4$ 467.3137), thus indicating the presence of seven degrees of

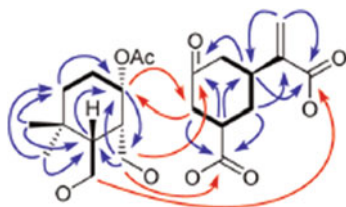
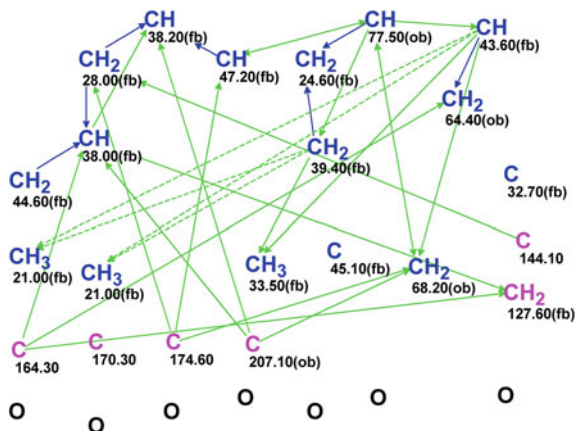


Fig. 5.9 Ternifolide A: Graphical representation of COSY (*bold lines*) and HMBC (*arrows*) correlations [5]

Table 5.5 Ternifolide A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	HMBC
C1	77.5	74.3	CH	5.92	u	1.62	C9, C20
C2	24.6	23.43	CH ₂	1.86	u	–	–
C2	24.6	23.43	CH ₂	1.62	u	5.92, 1.26	–
C3	39.4	38.38	CH ₂	1.35	u	–	–
C3	39.4	38.38	CH ₂	1.26	u	1.62	C1
C4	45.1	42.09	C	–	–	–	–
C5	43.6	45.11	CH	2.75	u	4.4	C1
C6	64.4	61.45	CH ₂	4.52	u	–	–
C6	64.4	61.45	CH ₂	4.4	u	2.75	C15
C7	174.6	171.15	C	–	–	–	–
C8	38.2	38.22	CH	3.44	u	3.58, 1.96	C11, C13
C9	47.2	55.43	CH	3.58	u	3.44	C1, C7
C10	32.7	35.12	C	–	–	–	–
C11	207.1	207.69	C	–	–	–	–
C12	44.6	48.86	CH ₂	2.58	u	–	–
C12	44.6	48.86	CH ₂	2.47	u	3.05	–
C13	38.0	36.82	CH	3.05	u	2.47, 1.96	C15, C11
C14	28.0	29.12	CH ₂	1.96	u	3.05, 3.44	C7, C16
C14	28.0	29.12	CH ₂	2.76	u	–	–
C15	164.3	165.44	C	–	–	–	–
C16	144.1	141.12	C	–	–	–	–
C17	127.6	126.67	CH ₂	6.37	u	–	C13, C15
C17	127.6	126.67	CH ₂	5.57	u	–	–
C18	33.5	33.19	CH ₃	0.74	u	–	C3, C5
C19	21.0	21.21	CH ₃	0.92	u	–	C5, C3
C20	68.2	69.63	CH ₂	4.52	u	–	–
C20	68.2	69.63	CH ₂	4.94	u	–	C11, C7, C1, C5
C21	170.3	169.75	C	–	–	–	–
C22	21.0	29.68	CH ₃	2.02	u	–	–

Fig. 5.10 Ternifolide A:
Molecular connectivity
diagram



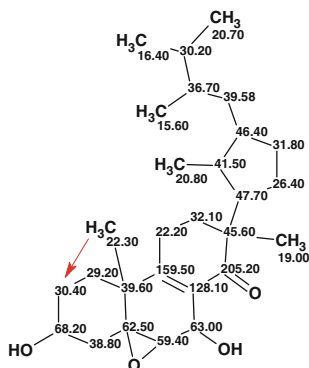
unsaturation in the molecule. The planar structure of **5.12** was elucidated by interpretation of the NMR spectra, including DEPT, COSY, HMBC, and HMQC data. Confirmation of the planar structure and determination of its absolute and relative configuration were performed by X-ray crystallography analysis of its bis (*p*-bromo)benzoate derivative. The DEPT experiment indicated the presence of six methyl, eight methylene, and eight methine groups, as well as six quaternary carbon atoms.

The experimental ^1H , ^{13}C , HMQC, and HMBC data were tabulated in the article [6]. They are presented in Table 5.6 and their graphical mapping in the form of the MCD is shown in Fig. 5.11.

MCD overview Two specific attributes for the MCD are: (a) there are many sp^3 -hybridized carbon atoms which are supplied by the program with the label “*fb*” (oxygen cannot be connected to these atoms), (b) the connectivity net is dense, i.e., the number of HMBC correlations (68) is high. Under such conditions, we can expect that structure generation can be successfully performed without any MCD edits though some of them are obvious (for example, carbons C 63.00 (4.84) and C 68.2 (3.94) definitely have an oxygen atom as a neighbor).

MCD checking revealed the presence of contradictions in the HMBC data. Therefore, FSG was initiated in the mode **Determine Options Automatically** with the results: $k = 152 \rightarrow 23 \rightarrow 6$, $t_g = 1$ s; 1 of the 68 HMBC connectivities was extended during the structure generation process and 5 of the 68 possible connectivity combinations were used during generation. The three top structures of the ranked output file are presented in Fig. 5.12.

It is obvious that structure #1 is identical to structure **5.12** of strophasterol A determined in the original work [6]. The closest competing structures are similar to structure #1, but they are reliably ruled out by the program from the set of candidate structures. Structure **5.13** shows the ^{13}C chemical shift assignment carried out automatically during structure generation.

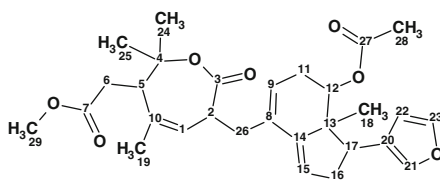


5.13

Thus a molecule with an unprecedented skeleton was automatically and instantly identified without the utilization of X-ray analysis or user intervention.

5.7 Aphanamixoid A

Limonoids, a series of structurally diverse and highly oxygenated tetranortriterpenoids mainly found in the family of Meliaceae, have been attracting continuous attention from biogenetic and synthetic points of view. In recent years, a number of limonoids have been isolated from the plant *Aphanamixis polystachya* by several research groups. No significant bioactivity has been found for those compounds. A new limonoid with a *new carbon skeleton* and potent antifeedent activity, aphanamixoid A (**5.14**), was isolated from the leaves and twigs of *A. polystachya* by Cai et al. [7].



5.14

Aphanamixoid A was obtained as colorless crystals (in acetone). Its molecular formula, $C_{29}H_{36}O_7$, was established from the quasi-molecular ion peak at m/z 519.2361 $[M+Na]^+$ (calculated 519.2358, $C_{29}H_{36}O_7Na$) in the positive HRESIMS, which indicated twelve degrees of unsaturation. The 1D NMR and HSQC data, and selected key HMBC and COSY correlations available from the article [7] in a graphical form only, are presented in Table 5.7. Figure 5.13 displays the MCD.

Table 5.6 Strophasterol A: Spectroscopic NMR data (in CDCl₃)

Label	δC	δC_{calc}	CH _n	δH	M(J)	C HMBC
C1	29.2	30.81	CH ₂	1.7	u	C5, C14, C3, C16
C1	29.2	30.81	CH ₂	1.84	u	–
C2	30.4	30.93	CH ₂	1.99	u	–
C2	30.4	30.93	CH ₂	1.69	u	C16, C1, C4, C3
C3	68.2	68.62	CH	3.94	u	–
C4	38.8	38.08	CH ₂	2.19	u	–
C4	38.8	38.08	CH ₂	1.48	u	C3, C6, C5, C2, C14
C5	62.5	64.98	C	–	–	–
C6	59.4	62.1	CH	3.21	u	C8, C5, C7, C4
C7	63	63.82	CH	4.84	u	C5, C9, C8, C13, C6
C8	128.1	134.38	C	–	–	–
C9	205.2	205.29	C	–	–	–
C10	45.6	47.73	C	–	–	–
C11	32.1	34.96	CH ₂	2	u	–
C11	32.1	34.96	CH ₂	1.65	u	C12, C17, C9, C15, C13, C10
C12	22.2	23.13	CH ₂	2.19	u	–
C12	22.2	23.13	CH ₂	2.39	u	C8, C14, C10, C13, C11
C13	159.5	158.56	C	–	–	–
C14	39.6	41.92	C	–	–	–
C15	19	21.98	CH ₃	0.93	u	C17, C10, C9, C11
C16	22.3	21.06	CH ₃	1.31	u	C1, C14, C5, C13
C17	47.7	52.42	CH	1.88	u	C9, C15, C20, C21, C22, C18, C10
C18	41.5	38.01	CH	1.25	u	C17, C22, C19, C20, C23
C19	46.4	41.96	CH	1.36	u	C26, C22, C28, C25, C23
C20	31.8	30.89	CH ₂	1.73	u	–
C20	31.8	30.89	CH ₂	1.03	u	C21, C19, C18, C17, C23
C21	26.4	27.42	CH ₂	1.37	u	–
C21	26.4	27.42	CH ₂	1.27	u	C17, C20, C18, C19, C10
C22	20.8	17.9	CH ₃	0.97	u	C19, C18, C17
C23	39.58	37.53	CH ₂	0.86	u	C19, C25, C20, C27, C24, C18
C23	39.58	37.53	CH ₂	1.44	u	–
C24	36.7	36.78	CH	1.36	u	–
C25	30.2	32.39	CH	1.6	u	C23, C26, C28, C27
C26	16.4	18.67	CH ₃	0.72	u	C25, C24, C28
C27	15.6	15.84	CH ₃	0.74	u	C25, C24, C23
C28	20.7	20.87	CH ₃	0.83	u	C25, C26, C24

MCD overview The MCD displays four carbon atoms for which hybridization was not assigned by the program (marked as light blue). The methyl group C 25.70 has no connectivity with any other atoms. Three quaternary carbons C 170.9,

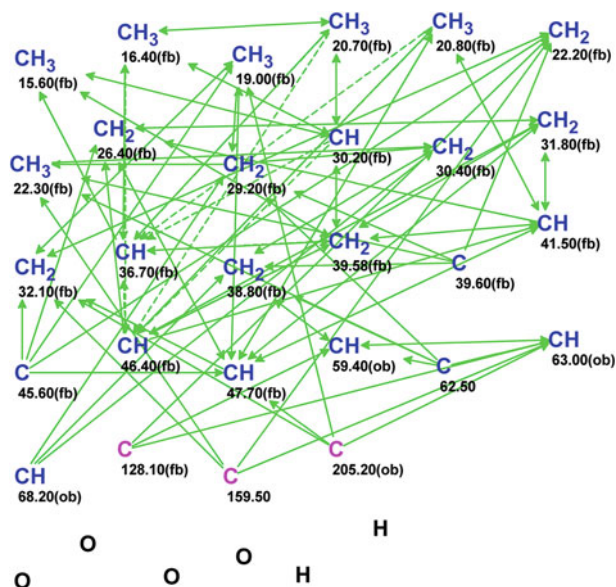


Fig. 5.11 Strophasterol A: The MCD as originally generated

<p>1</p>	<p>2</p>	<p>3</p>
<p>$d_A(^{13}\text{C}): 1.912$ $d_N(^{13}\text{C}): 2.097$ $d_I(^{13}\text{C}): 1.950$ $\text{max_}d_A(^{13}\text{C}): 6.280$</p>	<p>$d_A(^{13}\text{C}): 2.534$ $d_N(^{13}\text{C}): 2.914$ $d_I(^{13}\text{C}): 2.779$ $\text{max_}d_A(^{13}\text{C}): 15.070$</p>	<p>$d_A(^{13}\text{C}): 2.655$ $d_N(^{13}\text{C}): 2.965$ $d_I(^{13}\text{C}): 2.716$ $\text{max_}d_A(^{13}\text{C}): 13.110$</p>

Fig. 5.12 Strophasterol A: The three top structures of the ranked output file

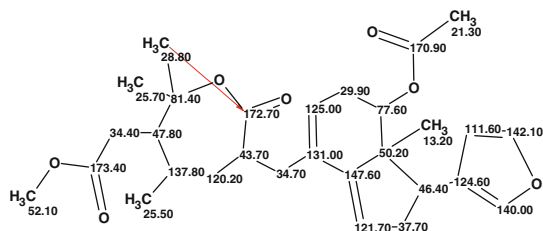
C 172.7, and C 173.4 which probably belong to ester carbonyls are not marked with the “ob” label by the program. Nevertheless, no MCD edits were made to accelerate problem solving. Checking the MCD gave a program message that at least one nonstandard connectivity is present in the 2D NMR data. Therefore, FSG was carried out in the mode **Determine Options Automatically** and completed with the results: $k = 80 \rightarrow 5 \rightarrow 5$, $t_g = 3$ s, 1 from 37 correlations was extended during generation and 37 from 37 possible combinations of connectivities were used

Table 5.7 Aphanamixoid A: The spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	120.2	123.036	CH	5.18	u	3.42	–
C2	43.7	44.48	CH	3.42	u	5.18, 2.30	–
C3	172.7	173.596	C	–	–	–	–
C4	81.4	84.87	C	–	–	–	–
C5	47.8	48.532	CH	2.74	u	2.23	C1, C7
C6	34.4	34.1	CH ₂	2.82	u	–	–
C6	34.4	34.1	CH ₂	2.23	u	2.74	C7, C10, C5, C4
C7	173.4	172.235	C	–	–	–	–
C8	131	135.229	C	–	–	–	–
C9	125	123.554	CH	5.5	u	2.22	–
C10	137.8	136.996	C	–	–	–	–
C11	29.9	30.184	CH ₂	2.47	u	–	–
C11	29.9	30.184	CH ₂	2.22	u	5.07, 5.50	–
C12	77.6	74.885	CH	5.07	u	2.22	C27
C13	50.2	51.415	C	–	–	–	–
C14	147.6	148.269	C	–	–	–	–
C15	121.7	121.815	CH	5.65	u	2.58	C14, C13, C8
C16	37.7	39.85	CH ₂	2.58	u	5.65, 3.15	–
C17	46.4	46.495	CH	3.15	u	2.58	C22, C21, C20
C18	13.2	18.45	CH ₃	0.9	u	–	C13, C12, C14, C17
C19	25.5	23.331	CH ₃	1.78	u	–	C1, C5, C10
C20	124.6	126.186	C	–	–	–	–
C21	140	139.657	CH	7.23	u	–	–
C22	111.6	109.96	CH	6.32	u	7.35	–
C23	142.1	142.48	CH	7.35	u	6.32	–
C24	28.8	24.304	CH ₃	1.36	u	–	C4, C3, C5
C25	25.7	24.304	CH ₃	1.61	u	–	–
C26	34.7	36.542	CH ₂	2.97	u	–	–
C26	34.7	36.542	CH ₂	2.3	u	3.42	C3, C14, C1, C2, C9, C8
C27	170.9	169.943	C	–	–	–	–
C28	21.3	21.186	CH ₃	1.9	u	–	C27
C29	52.1	51.78	CH ₃	3.7	u	–	C7

during the structure generation process. The three top structures of the ranked output file are shown in Fig. 5.14.

The first ranked structure coincides with the structure of aphanamixoid A (5.15, an arrow shows a nonstandard HMBC connectivity) and is characterized by average deviations of $d_A = 1.66$, $d_N = 1.53$, and $d_I = 1.80$ ppm. The remaining structures can be confidently rejected both by their deviation values and by their exotic nature.

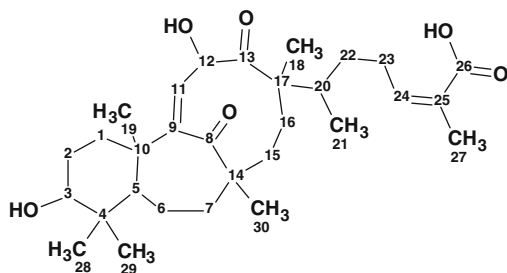


5.15

Thus a natural product which possesses a new skeleton was identified by StrucEluc automatically with no user intervention almost instantly.

5.8 Schiglautone A

Several plants from the genus *Schisandra* of the family *Schisandraceae* are widely used in Traditional Chinese Medicine. In addition to the presence of a large number of lignans, *Schisandra* was also found to be rich in triterpenoids with numerous pharmaceutical effects, which has aroused a lot of interest from pharmacologists. *Schisandra glaucescens* Diels is a vine plant mainly distributed in China. Its stems have been used for the treatment of various diseases in folk medicine. The chemical constituents and pharmacological potential of *S. glaucescens* have never been reported. Meng et al. [8] investigated potentially biologically active substances from plants, the chemical constituents of the stems of *S. glaucescens*. As a result a novel triterpenoid possessing an *unusual* 6/7/9-fused tricyclic ring system was obtained, which was designated as schiglautone A (**5.16**).



5.16

Schiglautone A was obtained as a colorless crystal with a molecular formula of $C_{30}H_{46}O_6$ as deduced from HRESIMS data (m/z 501.3204 $[M+H]^+$, calculated for 501.3216), requiring eight degrees of unsaturation. The 1H NMR data of **5.16** indicated the existence of seven methyls, two olefinic methines, and two

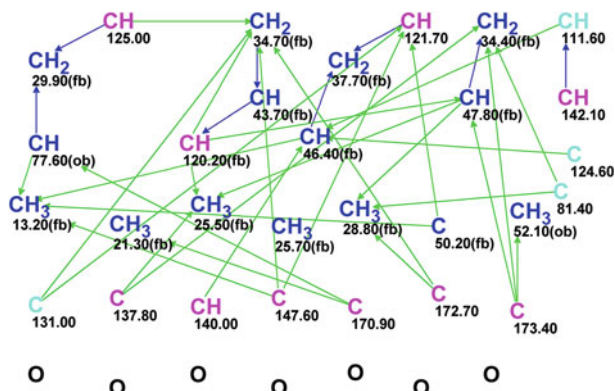


Fig. 5.13 Aplanamixoid A: The MCD

1	2	3
$d_A(^{13}\text{C}): 1.659$ $d_N(^{13}\text{C}): 1.531$ $d_I(^{13}\text{C}): 1.796$ $\text{max}_A d_A(^{13}\text{C}): 5.470$	$d_A(^{13}\text{C}): 2.462$ $d_N(^{13}\text{C}): 2.089$ $d_I(^{13}\text{C}): 2.182$ $\text{max}_A d_A(^{13}\text{C}): 7.270$	$d_A(^{13}\text{C}): 3.609$ $d_N(^{13}\text{C}): 2.487$ $d_I(^{13}\text{C}): 2.288$ $\text{max}_A d_A(^{13}\text{C}): 27.050$

Fig. 5.14 Aplanamixoid A: The three top structures of the ranked output file

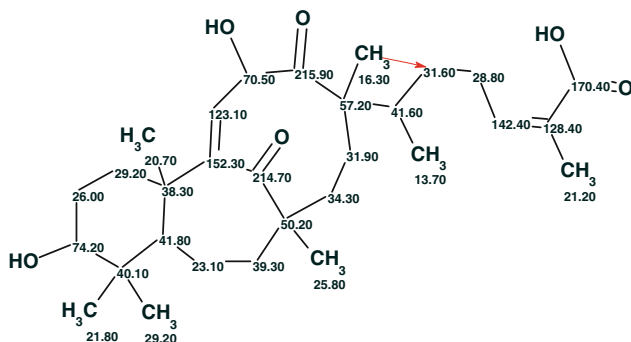
oxymethines. The spectroscopic NMR data (1D NMR, HSQC, HMBC, and COSY) are presented in Table 5.8.

The MCD is displayed in Fig. 5.15.

MCD overview As the degree of unsaturation (8) is not large, the number of HMBC and COSY connectivities is sufficient to allow us to omit the stage of MCD editing. Multiplicities determined in the ^1H NMR spectrum were also not used.

MCD checking revealed the presence of at least one nonstandard connectivity in the 2D NMR data. FSG was initiated in the mode **Determine Options Automatically** which was completed with the following results: $k = 2 \rightarrow 2 \rightarrow 2$, $t_g = 0.015$ s, 1 from 59 correlations has been extended during generation, and 1 from 59 possible connectivity combinations was used during generation. Both generated structures are shown in Fig. 5.16.

The first ranked structure is identical to structure 5.16. The structure of schi-glautone A supplied with ^{13}C chemical shift assignment (5.17) is shown below. The arrow denotes a single nonstandard HMBC connectivity.

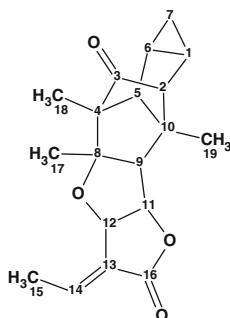


5.17

A structure containing 36 skeletal atoms and a unique 6/7/9-fused skeleton was therefore elucidated almost instantly ($t_g = 0.015$ s!) and unambiguously in the fully automated mode of FSG. This was possible owing to the rich HMBC and COSY data and successful automatic setting of the carbon atom properties. In this case, the initial set of axioms can be characterized as “complete” but contradictory.

5.9 Pallambin A

Liverworts are the most primitive group of terrestrial plants and are widely distributed throughout the world. They produce rich terpenoids and aromatic compounds, many of which exhibit a variety of fascinating structures and interesting biological activities. Chemical investigations on liverworts of the genus *Pallavicinia* afforded many kinds of di- and sesquiterpenoids, mainly including the labdane and clerodane diterpenoids, as well as sesquiterpenoids. Wang et al. [9] isolated pallambin A and pallambin B, two novel 19-nor-7,8-secolabdane diterpenoids with *unprecedented tetracyclodecane skeletons*. We will examine the structure elucidation of pallambin A (**5.18**).



5.18

Table 5.8 Schiglaune A: The spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	29.2	34.08	CH ₂	2.36	u	1.77	C9, C3, C10, C19
C1	29.2	34.08	CH ₂	1.43	u	–	C19, C3, C9
C2	26	27.15	CH ₂	1.95	u	–	–
C2	26	27.15	CH ₂	1.77	u	2.36, 3.57	C1, C3
C3	74.2	78.02	CH	3.57	u	1.77	C5
C4	40.1	39.52	C	–	–	–	–
C5	41.8	54.23	CH	2.43	u	1.6	C7, C19, C9, C3
C6	23.1	22.12	CH ₂	1.38	u	–	–
C6	23.1	22.12	CH ₂	1.6	u	2.43, 1.51	C14
C7	39.3	41.43	CH ₂	1.21	u	–	–
C7	39.3	41.43	CH ₂	1.51	u	1.6	–
C8	214.7	205.96	C	–	–	–	–
C9	152.3	150.75	C	–	–	–	–
C10	38.3	41.18	C	–	–	–	–
C11	123.1	138.27	CH	5.74	u	5.27	C8, C9, C10, C12
C12	70.5	73.79	CH	5.27	u	5.74	C17, C13, C11, C9
C13	215.9	217.2	C	–	–	–	–
C14	50.2	47.35	C	–	–	–	–
C15	34.3	36.57	CH ₂	2.54	u	2.21	C17, C30, C16, C8
C15	34.3	36.57	CH ₂	1.232	u	–	C8, C17
C16	31.9	31.83	CH ₂	2.21	u	2.54	C15, C14, C18, C17
C16	31.9	31.83	CH ₂	1.64	u	–	C17, C13
C17	57.2	51.73	C	–	–	–	–
C18	16.3	18.34	CH ₃	1.32	u	–	C13, C22, C17, C20
C19	20.7	25	CH ₃	1.2	u	–	C5, C9, C10
C20	41.6	33.53	CH	1.7	u	1.02, 1.27	C17, C21, C13
C21	13.7	15.86	CH ₃	1.02	u	1.7	C22, C20, C17
C22	31.6	31.06	CH ₂	1.89	u	–	C17, C21, C23, C24
C22	31.6	31.06	CH ₂	1.27	u	1.70, 2.75	C21, C23, C24
C23	28.8	29.87	CH ₂	2.75	u	1.27, 5.98	C22, C25, C20, C24
C23	28.8	29.87	CH ₂	2.9	u	–	C20, C22, C25, C24
C24	142.4	144.08	CH	5.98	u	2.75	C26, C27, C22
C25	128.4	127.57	C	–	–	–	–
C26	170.4	172.34	C	–	–	–	–
C27	21.2	20.75	CH ₃	2.05	u	–	C26, C24, C25
C28	21.8	27.14	CH ₃	0.77	u	–	C29, C4, C3
C29	29.2	16.15	CH ₃	1.16	u	–	C28, C4, C3
C30	25.8	22.3	CH ₃	1.5	u	–	C7, C8, C14, C15

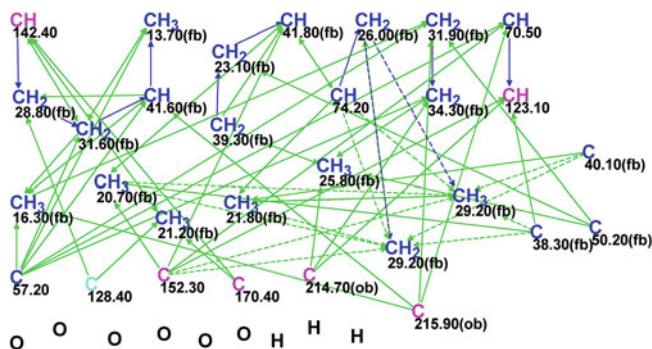


Fig. 5.15 Schiglautone A: The MCD. No MCD edits were made

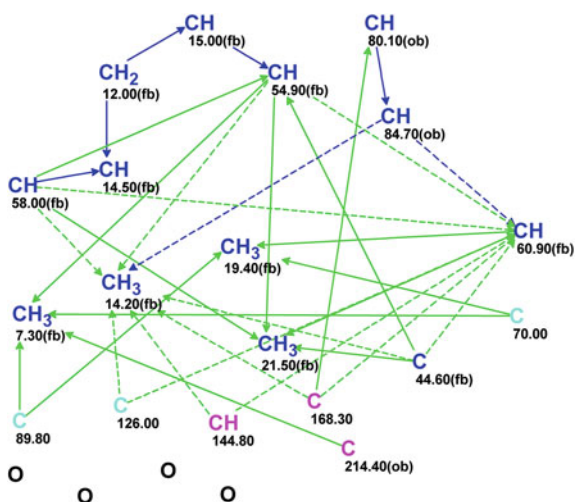
1	2
$d_A(^{13}\text{C}): 3.430$ $d_N(^{13}\text{C}): 3.244$ $d_I(^{13}\text{C}): 2.903$ $\text{max}_d_A(^{13}\text{C}): 15.170$	$d_A(^{13}\text{C}): 4.644$ $d_N(^{13}\text{C}): 3.675$ $d_I(^{13}\text{C}): 4.069$ $\text{max}_d_A(^{13}\text{C}): 24.420$

Fig. 5.16 Schiglautone A: The output structures as ranked by ^{13}C average deviations

Pallambin A was isolated as a colorless crystal. Its molecular formula was determined to be $\text{C}_{19}\text{H}_{22}\text{O}_4$ by HRESIMS requiring nine degrees of unsaturation (positive HRESIMS m/z 337.1412 $[\text{M}+\text{Na}]^+$, calculated for $\text{C}_{19}\text{H}_{22}\text{O}_4\text{Na}$, 337.1410). According to the characteristic spectral features, the ^{13}C NMR spectrum allowed the authors [9] to suggest the presence of one double bond (C 126.0 and C 144.8), an ester carbonyl (C 168.3), and a ketone carbonyl (C 214.4). If the assignments are correct then the compound must be a six-ring structure in order to achieve its degree of unsaturation. The ^{13}C , ^1H , and 2D NMR data which were used in [9] for the structure elucidations of pallambin A are collected in Table 5.9. The MCD is presented in Fig. 5.17.

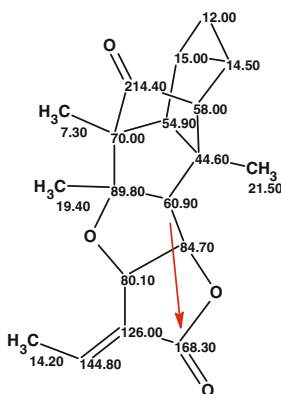
Table 5.9 Pallambin A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	14.5	18.38	CH	1.43	u	2.49, 0.56	–
C2	58	58.61	CH	2.49	u	1.43	–
C3	214.4	217.62	C	–	–	–	–
C4	70	59.76	C	–	–	–	–
C5	54.9	55.81	CH	2.48	u	0.87	C10, C2, C6
C6	15	15.14	CH	0.87	u	2.48, 0.56	–
C7	12	8.11	CH_2	1.4	u	–	–
C7	12	8.11	CH_2	0.56	u	0.87, 1.43	–
C8	89.8	94.13	C	–	–	–	–
C9	60.9	55.13	CH	2.27	d(7.1)	4.95	C5, C13, C14, C16, C10, C2
C10	44.6	53.76	C	–	–	–	–
C11	84.7	82.82	CH	4.95	u	2.27, 4.75	–
C12	80.1	77.22	CH	4.75	d(3.5)	4.95	C16
C13	126	129.55	C	–	–	–	–
C14	144.8	141.28	CH	6.69	q(7.3)	–	–
C15	14.2	13.92	CH_3	2.27	d(7.3)	–	–
C16	168.3	170.1	C	–	–	–	–
C17	19.4	19.2	CH_3	1.11	S	–	C9, C4, C8
C18	7.3	12	CH_3	1.18	S	–	C5, C4, C3, C8
C19	21.5	21.21	CH_3	1.34	S	–	C10, C9, C2, C5

Fig. 5.17 Pallambin A: Molecular connectivity diagram

MCD overview As the proton chemical shifts of the hydrogens attached to C9 and C15 are the same (2.27 ppm), the MCD contains ambiguous connectivities. Three carbon atoms (70.00, 89.80, and 126.00) are colored in light blue (i.e., their hybridization types are not defined). No user interventions were made except for inputting the numbers of hydrogen atoms attached to the neighboring carbon atoms (see proton signal multiplicities in Table 5.9, column M(J)). MCD checking revealed the presence of at least one nonstandard correlation in the 2D NMR data.

FSG accompanied by ^{13}C chemical shift prediction was initiated in the mode **Determine Options Automatically** which was completed with the following results: $k = 564 \rightarrow 1$, $t_g = 1.5$ s, 1 from 25 connectivities has been extended, 25 from 25 possible connectivity combinations were used, $d_A = 3.22$, $d_N = 3.43$, $d_I = 2.96$ ppm. It turned out that the single structure obtained was identical to pallambin A. An arrow denotes a nonstandard HMBC connectivity in structure **5.19** which displays the ^{13}C chemical shift assignment.



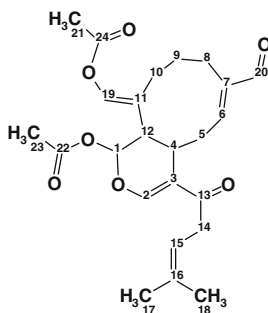
5.19

Average deviations calculated for structure **5.19** are relatively large, which can be explained by an unusual skeleton of the molecule. Moreover, the results of the HOSE-based chemical shift calculations were displayed along with the following program message: “ACD/CNMR cannot guarantee the precise spectrum calculation for this structure. There is not enough information about the same or similar structures.” In article [9], the correctness of structure **5.19** was confirmed by NOESY analysis as well as by X-ray crystallography.

5.10 Cristaxenicin A

More than one billion people are affected by neglected tropical diseases, of which leishmaniasis is thought to be the most difficult to control. Toxicity and the high cost for the treatments of antileishmanial drugs pose serious problem for controlling

the disease. Thus, new drugs should be urgently developed. More than 90 marine natural products that may have efficacy in treating the disease have been reported so far, but none of them have reached clinical trials. In the course of these efforts to discover potential drug leads from marine invertebrates, Ishigami et al. [10] have tested 1,565 extracts of marine organisms. They found promising activity in the lipophilic extract of the deep-sea gorgonian *Acanthoprimnoa cristata* collected in southern Japan. Bioassay-guided isolation yielded a highly antileishmanial new xenicane diterpenoid, Cristaxenicin A (**5.20**).



5.20

Cristaxenicin A had a molecular formula of $C_{24}H_{30}O_7$ as determined by HRFABMS at m/z 431.2089 $[M+H]^+$ [calculated for $C_{24}H_{30}O_7$ m/z 431.2070 (Δ -1.9 mmu)]. For the structure elucidation 1H , ^{13}C , HSQC, HMBC, and COSY NMR spectra were used (see Table 5.10).

The MCD created from the data presented in Table 5.10 is shown in Fig. 5.18.

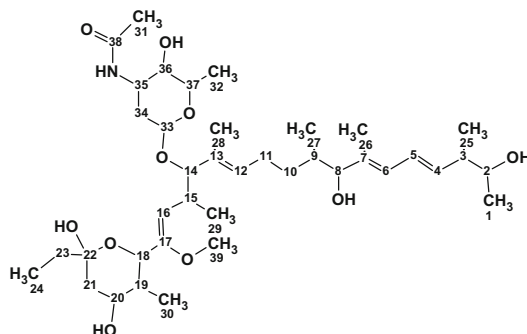
MCD overview The diagram contains three light blue carbon atoms CH 94.5, C 120.2, and C 120.6, all being involved in the net of connectivities, and two “free” methyl groups (δC 19.00 and δ 19.10). No MCD edits were made. As the presence of contradictions in the 2D NMR data was detected by the program, FSG was run with options which were defined automatically and the results gave $k = 2 \rightarrow 2 \rightarrow 1$, $t_g = 0.13$ s.

The single output structure with its assigned ^{13}C chemical shifts and calculated average deviations is shown in Fig. 5.19.

Comparison of the resultant structure with structure **5.20** leads to the conclusion that the correct solution was found instantaneously and fully automatically.

5.11 Juniperolide A

Raju et al. [11] described the isolation and structure elucidation of a new linear polyketide, Juniperolide A (**5.21**).



5.21

Table 5.10 Cristaxenicin A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	94.5	90.63	CH	5.86	u	2.48, 3.04	C22
C2	154.6	153.24	CH	7.68	u	–	C13, C3, C4, C1
C3	120.2	118.92	C	–	–	–	–
C4	35.9	31.55	CH	3.04	u	2.86, 5.86, 2.48	–
C5	30	33.25	CH ₂	2.65	u	–	–
C5	30	33.25	CH ₂	2.86	u	3.04, 6.65	–
C6	153.4	152.84	CH	6.65	u	2.86	C20, C8
C7	144.9	143.97	C	–	–	–	–
C8	21.2	24.73	CH ₂	2.32	u	–	–
C8	21.2	24.73	CH ₂	2.39	u	1.83	C20, C6, C7
C9	22.5	24.05	CH ₂	1.83	u	2.54, 2.39	C7
C10	28.8	27.23	CH ₂	2.54	u	1.83	C11, C12, C8
C10	28.8	27.23	CH ₂	2	u	–	–
C11	120.6	119.81	C	–	–	–	–
C12	49	43.22	CH	2.48	u	3.04, 5.86	C19
C13	198.2	198.36	C	–	–	–	–
C14	37.6	36.46	CH ₂	3.33	u	–	–
C14	37.6	36.46	CH ₂	3.4	u	5.3	C16, C13
C15	116.9	116.42	CH	5.3	u	3.4	–
C16	134.9	135.62	C	–	–	–	–
C17	16.7	18.3	CH ₃	1.68	u	–	C18, C15, C16
C18	24.5	26.06	CH ₃	1.75	u	–	C16, C17, C15
C19	136	132.65	CH	6.98	u	–	C24, C12, C11, C10
C20	195.8	195.74	CH	9.27	u	–	C8, C7, C6
C21	19.1	20.97	CH ₃	2.05	u	–	–
C22	168.9	169.26	C	–	–	–	–
C23	19	20.89	CH ₃	2.1	u	–	–
C24	167.8	167.87	C	–	–	–	–

Fig. 5.18 Cristaxenicin A:
Molecular connectivity
diagram

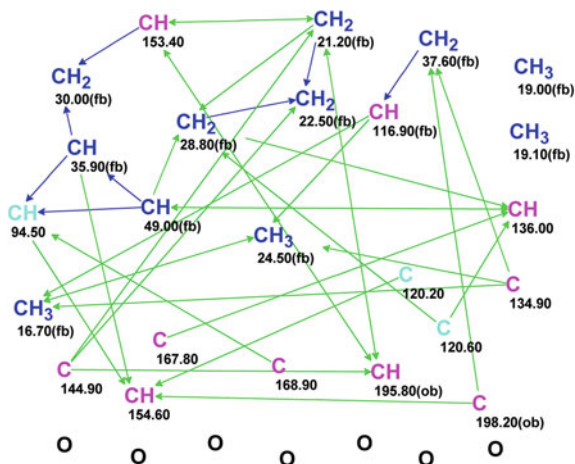
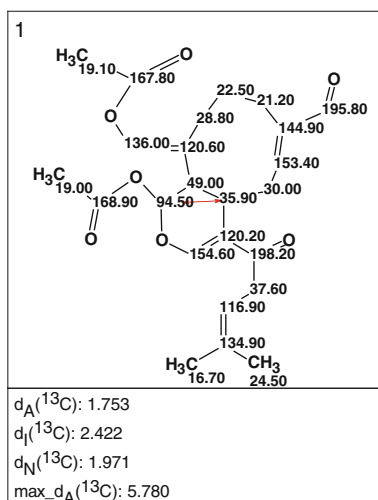


Fig. 5.19 Cristaxenicin A:
The resultant structure. The
red arrow shows a
nonstandard COSY
connectivity



It was produced by the terrestrial actinomycete (Lv1-48) isolated from the rhizosphere of the plant *Juniperus excelsa*. HPLC-DAD-MS analysis of a small-scale liquid fermentation revealed the production of Juniperolide A (m/z 709). HRESIMS (+) analysis of **5.21** returned a molecular formula $\text{C}_{39}\text{H}_{67}\text{NO}_{10}$ requiring seven double bond equivalents. The UV absorbance at 240 nm was suggestive of a diene functionality in the structure.

To elucidate the structure of Juniperolide A the ^{13}C , ^1H , HSQC, HMBC, and COSY data presented in Table 5.11 were used. The NMR data (obtained in methanol- d_4) revealed resonances for an amide or ester carbonyl (δC 173.5), and four double bonds (δC 118.9–154.7), requiring **5.21** to incorporate two rings.

The MCD created from the spectroscopic data is presented in Fig. 5.20.

Table 5.11 Juniperolide A: Spectroscopic NMR data (methanol-d4)

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	20.2	20.45	CH ₃	1.11	u	3.65	C3, C2
C2	72.1	71.28	CH	3.65	u	1.11, 2.25	C4, C1, C25, C3
C3	45.3	43.16	CH	2.25	u	1.03, 3.65	C1, C25, C5, C2, C4
C4	136.9	138.35	CH	5.64	u	6.33	C6, C25, C5, C3, C2
C5	127.5	127.03	CH	6.33	u	5.64, 5.99	C6, C3, C7
C6	127.7	123.97	CH	5.99	u	6.33	C26, C8, C5, C4
C7	138.2	139.39	C	–	–	–	–
C8	83	82.37	CH	3.67	u	1.65	C26, C6, C9
C9	36.7	36.35	CH	1.65	u	3.67, 1.06, 0.95	–
C10	34	33.07	CH ₂	1.06	u	1.65, 2.10	C27
C10	34	33.07	CH ₂	1.34	u	–	–
C11	25.9	25.06	CH ₂	2.1	u	1.06, 5.29	C13, C12
C11	25.9	25.06	CH ₂	1.95	u	–	–
C12	129.5	129.56	CH	5.29	u	2.1	C11, C10, C28, C14
C13	136.4	134.07	C	–	–	–	–
C14	92.7	85.38	CH	3.54	u	2.85	C28, C12, C29, C16, C15
C15	33.9	33.59	CH	2.85	u	3.54, 0.99, 4.65	C14, C29, C16, C17
C16	118.9	110.44	CH	4.65	u	2.85	C15, C17, C29, C14, C18
C17	154.7	153.61	C	–	–	–	–
C18	73.2	76.61	CH	4.2	u	1.7	C20, C19, C17, C22, C16, C30
C19	37.9	39.12	CH	1.7	u	4.20, 2.08, 3.92	–
C20	70.7	70.99	CH	3.92	u	1.70, 2.08	–
C21	38.9	39.83	CH ₂	1.9	u	–	–
C21	38.9	39.83	CH ₂	<u>2.08</u>	u	1.70, 3.92	C20, C35, C33, C36
C22	99.9	99.01	C	–	–	–	–
C23	26.3	31.14	CH ₂	1.6	u	0.93	C22, C24
C24	8	7.61	CH ₃	0.93	u	1.6	C22
C25	16.4	16.11	CH ₃	1.03	u	2.25	C2, C4, C3
C26	12.5	11.88	CH ₃	1.72	u	–	C7, C8, C6
C27	15.6	15.93	CH ₃	0.95	u	1.65	C9, C8, C10
C28	12.2	11.46	CH ₃	1.57	u	–	C12, C13, C14
C29	18.3	14.31	CH ₃	0.99	u	2.85	C14, C16, C15
C30	14.4	11.62	CH ₃	0.85	u	–	C18, C19, C20
C31	22.5	22.9	CH ₃	1.97	u	–	C38
C32	18.4	18.11	CH ₃	1.21	u	3.19	C36, C37
C33	101.1	97.83	CH	4.5	u	1.44	C14, C34

(continued)

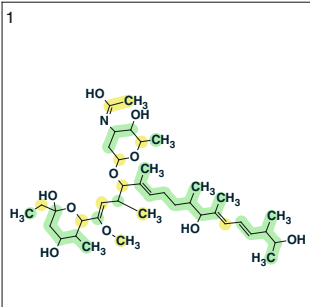
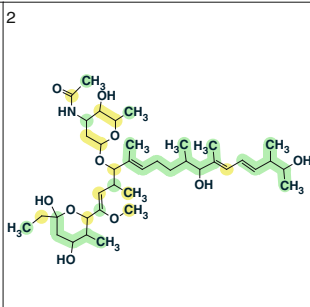
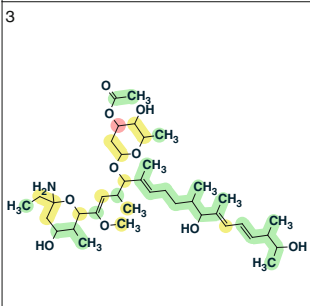
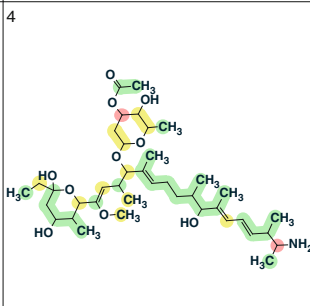
	
$d_A(^{13}\text{C}): 2.085$ $d_N(^{13}\text{C}): 2.216$ $d_I(^{13}\text{C}): 2.160$ $\text{max_}d_A(^{13}\text{C}): 8.460$	$d_A(^{13}\text{C}): 2.285$ $d_N(^{13}\text{C}): 2.031$ $d_I(^{13}\text{C}): 2.121$ $\text{max_}d_A(^{13}\text{C}): 9.470$
	
$d_A(^{13}\text{C}): 2.861$ $d_N(^{13}\text{C}): 2.528$ $d_I(^{13}\text{C}): 3.079$ $\text{max_}d_A(^{13}\text{C}): 17.980$	$d_A(^{13}\text{C}): 3.170$ $d_N(^{13}\text{C}): 2.969$ $d_I(^{13}\text{C}): 3.276$ $\text{max_}d_A(^{13}\text{C}): 20.590$

Fig. 5.21 Juniperolide A: The first four structures of the ranked output file. The color attributes of the ^{13}C chemical shift prediction accuracy are: ± 3 ppm—green, between 3 and 15 ppm—yellow, greater than 15 ppm—red

which mark the accuracy of the ^{13}C chemical shift prediction for each carbon atom in the candidate structure with a specific color (see Sect. 2.2.2).

Figure 5.21 shows that structures #1 and #2 are characterized by close average and maximum deviation values and have very similar coloring, but structure #2, coinciding with structure 5.21, is preferred by neural net and increment-based prediction, while structure #1 is ranked as the best by the HOSE code-based algorithm. This is accounted for by the fact that both structures #1 and #2 are tautomers. Figure 5.22 presents the tautomer-dependent parts of these structures supplied with the experimental and calculated ^{13}C chemical shifts.

Figure 5.22 shows that the predicted chemical shifts for carbons C 173.5 and C 22.5 favor structure Juniperolide A suggested by the authors [11]. Structures #3–#4 (see Fig. 5.21) contain an ester instead of an amide group and they are rejected by

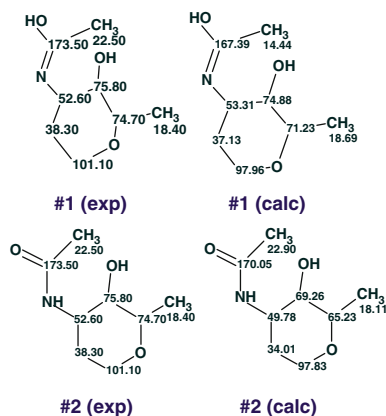
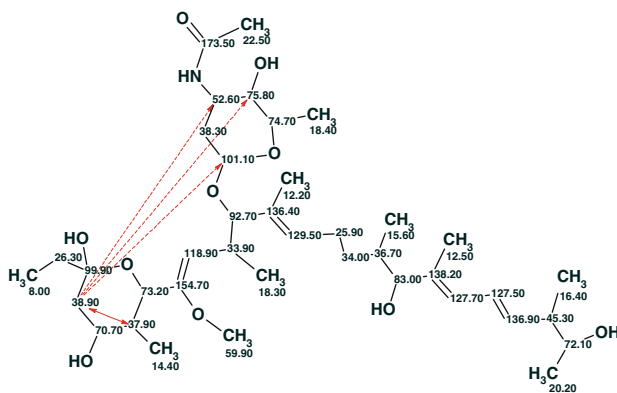


Fig. 5.22 Juniperolide A: The tautomer-dependent parts of structures #1 and #2 with the experimental (*left*) and predicted (*right*) ^{13}C chemical shifts

the large chemical shift discrepancies at carbon atoms C35 (in #3) and C35 and C2 (in #4) with all of the mentioned atoms colored in red.

The structure of Juniperolide A supplied with the ^{13}C chemical shift assignment is presented as structure 5.22.

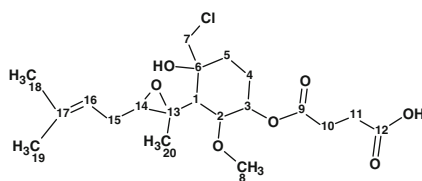


5.22

The solid double-sided arrow denotes a COSY nonstandard connectivity, while the ambiguous connectivities are marked by dotted lines. The carbon atoms C 99.90 and C 101.1 are sp^3 -hybridized and included into ketal groups but not into a $\text{C}=\text{C}$ double bond as would be suggested according to the chemical shift values.

5.12 Ligerin

In the course of searching for new antiproliferative compounds from marine-derived fungi, Vansteelandt and co-workers [12] investigated a strain of *Penicillium* sp. belonging to the section *Canescentia* of the subgenus *Penicillium*. This strain was selected after screening for cytotoxicity against cancer cell lines versus non-tumor cell lines. The authors isolated a new chlorinated sesquiterpenoid compound related to fumagillin, Ligerin (**5.23**) and elucidated its structure by interpreting the spectroscopic data including IR, UV, and HRESIMS, together with analyses of the 1D and 2D NMR spectra.



5.23

Ligerin was obtained as a colorless oil after three purification steps using bio-assay-guided fractionation. The analysis of its isotopic pattern by LRESIMS revealed the presence of a chlorine atom in the molecule with a 3:1 ratio for peak intensities of, respectively, m/z 441, i.e., $[M+Na]^+$, and m/z 443, i.e., $[M+2+Na]^+$. The HRESIMS spectrum showed an $[M+Na]^+$ ion at m/z 441.16507 and its cluster ion $[2M+Na]^+$ at m/z 859.34186, suggesting $C_{20}H_{31}ClO_7$ as the elemental composition of compound **5.23**, indicating five degrees of unsaturation.

The NMR data for **5.23** obtained from ^{13}C , 1H , HSQC, HMBC, and COSY spectra are listed in Table 5.12.

The MCD is presented in Fig. 5.23.

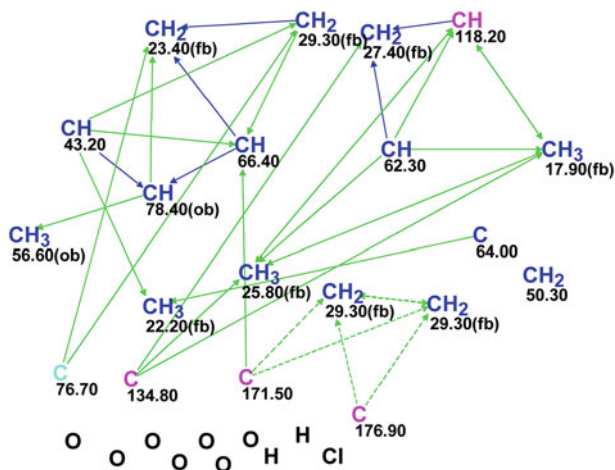
MCD overview The MCD contains many ambiguous connectivities (marked by dotted lines). This is a consequence of the presence of three methylene groups whose resonances in the ^{13}C NMR spectrum are overlapped at 29.3 ppm, while for two of them (C10 and C11, see Table 5.12) the 1H signals from the attached protons are also overlapped. Hybridization of the carbon atom at δ 76.7 was not specified by the program. No edits of the MCD were made. MCD checking for contradictions detected the presence of at least one nonstandard connectivity in the COSY and HMBC data.

FSG in the mode **Determine Options Automatically** combined with ^{13}C chemical shift prediction was completed with the following results: $k = 27,327 \rightarrow 65 \rightarrow 65$, $t_g = 38$, 2 from 28 connectivities have been extended during generation, 36 from 378 possible connectivity combinations were used. The three top structures of the ranked output file are presented in Fig. 5.24.

Figure 5.24 shows that all three structures are similar and are made up by permuting the elements included into a set of structural blocks. Therefore, it is not

Table 5.12 Ligerin: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	COSY	C HMBC
C1	43.2	48	CH	2.39	u	3.3	–
C2	78.4	80.25	CH	3.3	u	5.52, 2.39	–
C3	66.4	65.72	CH	5.52	u	1.84, 3.30	C1, C5, C9, C2
C4	23.4	24.56	CH ₂	1.84	u	5.52, 1.40	C3, C2, C6, C5
C5	29.3	30.08	CH ₂	1.4	u	1.84	C1, C6
C5	29.3	30.08	CH ₂	1.96	u	–	C4
C6	76.7	73.01	C	–	–	–	–
C7	50.3	55.41	CH ₂	3.85	u	–	–
C7	50.3	55.41	CH ₂	3.52	u	–	–
C8	56.6	56.4	CH ₃	3.28	u	–	C2
C9	171.5	173.03	C	–	–	–	–
C10	29.3	29.42	CH ₂	2.73	u	–	C9, C12, C11
C11	29.3	31.43	CH ₂	2.73	u	–	–
C12	176.9	174.36	C	–	–	–	–
C13	64	62.08	C	–	–	–	–
C14	62.3	60.84	CH	2.98	u	2.19	C15
C15	27.4	27.58	CH ₂	2.19	u	5.21, 2.98	C16, C17, C14
C15	27.4	27.58	CH ₂	2.47	u	–	–
C16	118.2	118.98	CH	5.21	u	2.19	C18, C19, C14
C17	134.8	134.9	C	–	–	–	–
C18	25.8	25.7	CH ₃	1.76	u	–	C14, C19, C16, C17
C19	17.9	18	CH ₃	1.68	u	–	C16, C18, C14, C17
C20	22.2	16.97	CH ₃	1.51	u	–	C1, C13

**Fig. 5.23** Ligerin: Molecular connectivity diagram

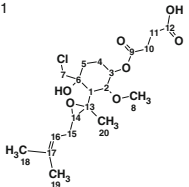
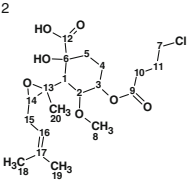
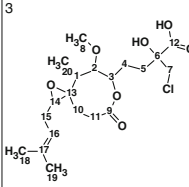
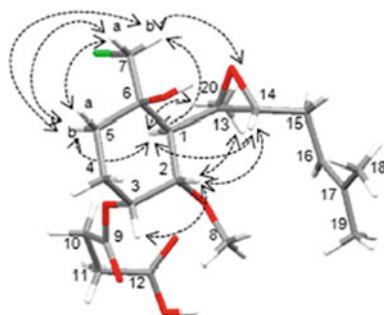
1	2	3
		
$d_A(^{13}\text{C}): 1.723$ $d_N(^{13}\text{C}): 1.763$ $d_I(^{13}\text{C}): 1.918$ $\text{max}_A(^{13}\text{C}): 5.230$	$d_A(^{13}\text{C}): 1.790$ $d_N(^{13}\text{C}): 2.084$ $d_I(^{13}\text{C}): 2.216$ $\text{max}_A(^{13}\text{C}): 7.890$	$d_A(^{13}\text{C}): 2.366$ $d_N(^{13}\text{C}): 1.989$ $d_I(^{13}\text{C}): 2.019$ $\text{max}_A(^{13}\text{C}): 10.740$

Fig. 5.24 Ligerin: Three top structures of the ranked output file

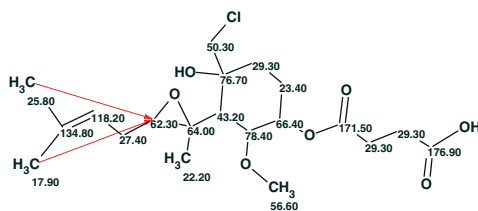
Fig. 5.25 Ligerin: Selected NOESY correlations



surprising that the average deviations characterizing the structures are very close. Nevertheless, both the average and maximum deviations distinguished the correct structure as the best. The correctness of the choice can be checked using the NOESY data. Figure 5.25 represents the NOESY correlations detected by the authors [12].

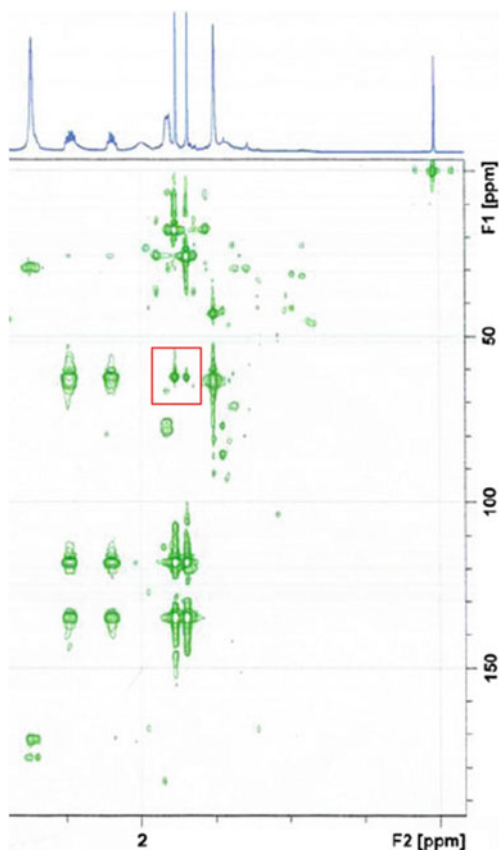
A comparison of structures #2 and #3 with Fig. 5.25 leads to the conclusion that NOESY correlations associated with hydrogen atom H7 to H1 and H20 would hardly be observed for structures #2 and #3 and this endorses the choice of the right structure.

The correct structure 5.24 with the associated ^{13}C chemical shift assignment and two HMBC nonstandard connectivities (arrows) is presented below.



5.24

Fig. 5.26 Ligerin: A part of the HMBC spectrum. The peaks corresponding to ${}^5J_{\text{CH}}$ couplings are contained within the *red square*



Note that both NSCs correspond to ${}^5J_{\text{CH}}$ couplings and they were both automatically detected and elongated by the appropriate number of chemical bonds (two in this case). The HMBC pattern available in the supporting information associated with article [12] (see Fig. 5.26) shows that the intensities of the peaks corresponding to the NSCs are rather significant and this gives no grounds to a priori assign them to coupling constants of the type ${}^5J_{\text{CH}}$.

The Ligerin structure was determined quickly and automatically. It should be noted that this is just the case when the preference of the structure #1 could be confirmed also by utilizing an additional NMR technique, for instance a 1,1-ADEQUATE experiment.

5.13 Phosphodiyn A

Polyacetylenic natural products characterized by carbon–carbon triple bonds or alkynyl functional groups are widely distributed in natural objects. A number of polyacetylenes have been isolated from marine sponges. Polyacetylenes vary by a

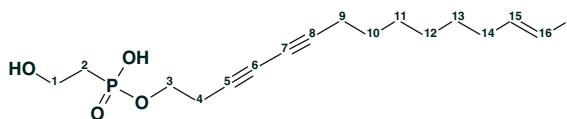
Table 5.13 Phosphoiodyd A: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	C HMBC
C1	35.5	54.98	CH_2	3.14	u	1.89	C2
C2	24.5	30.36	CH_2	1.89	u	3.14	C1
C3	61.9	63.03	CH_2	3.94	u	2.6	C4, C5
C4	21.2	21.77	CH_2	2.6	u	3.94	C5, C6, C3, C7
C5	73.1	74.61	C	–	–	–	–
C6	66.1	66.27	C	–	–	–	–
C7	64.8	64.78	C	–	–	–	–
C8	77.1	77.29	C	–	–	–	–
C9	18.2	19.1	CH_2	2.25	u	1.5	C6, C10, C8
C10	27.9	28.1	CH_2	1.5	u	2.25, 1.39	C11, C9, C8
C11	28	27.94	CH_2	1.39	u	1.5	C10, C12
C12	28.2	27.92	CH_2	1.32	u	1.43	C13, C11
C13	28	29.1	CH_2	1.43	u	2.07, 1.32	C15, C14
C14	35.5	36	CH_2	2.07	u	1.43, 6.52	C16, C15, C13
C15	147	146.3	CH	6.52	u	6.10, 2.07	C13, C14, C16
C16	73.6	74.5	CH	6.1	u	6.52	C15, C14

number of structural features, including the number of triple bonds, chain lengths, and the substituted functional groups. They exhibit diverse bioactivities such as cytotoxic, antiviral, antifouling, RNA-cleaving, and enzyme-inhibitory activities.

Kim et al. [13] isolated and identified two *unprecedented* phosphorus-containing iodinated polyacetylenes, Phosphoiodyd A and B, from a Korean marine sponge *Placospongia* sp. Here we will describe the structural elucidation of Phosphoiodyd A, as well as its structure revision on the basis of another axiom postulated by the same group of authors later [14].

The new compound was obtained as an amorphous solid. Its molecular formula was deduced as $\text{C}_{16}\text{H}_{24}\text{O}_4\text{PI}$ based on the pseudomolecular ion peak at m/z 438.0459 $[\text{M}+\text{H}]^+$ in the HRFABMS, the ^{13}C NMR data, and the suggestion that the admitted elemental composition is CHOPI. Structure **5.25** was derived for Phosphoiodyd A using the ^1H , ^{13}C , HSQC, COSY, and HMBC spectroscopic data (Table 5.13).

**5.25**

The ^1H NMR spectrum displayed 2 methines and 10 methylenes, two of which, C1 and C3, the authors [13] assigned as oxygenated methylenes. Progressing with the analysis it is worthy to note the discrepancy between the experimental (35.5)

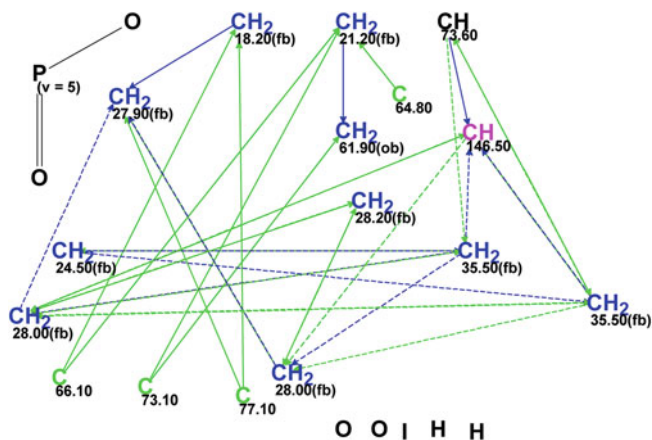


Fig. 5.27 Phosphoiodyne A: Molecular connectivity diagram. The four *sp*-hybridized carbon atoms are marked in a green color

and predicted (55.5) values of the chemical shifts assigned to carbon atom C1. The ^{13}C NMR spectrum in combination with the HSQC and HMBC spectra indicates four fully substituted carbons at δC 73.1, 66.1, 64.8, and 77.1, which allowed authors to assign to a diyne unit.

The MCD modified in accordance with some of the authors' assumptions regarding atom properties is presented in Fig. 5.27.

MCD overview When the MCD was created, the check box **Allow *sp* Carbons** in the window **Create MCD Options** was selected because the presence of triple bonds between carbon atoms was postulated on the basis of the sample origin and chemical knowledge. As a result, the carbon atoms C 64.8–C 77.10 were marked by the program with the hybridization attribute “not defined” (all mentioned atoms were colored in black). Following the authors' assumptions, the atoms C 64.8, C 66.1, C 73.1, and C 77.1 were assigned by the user as *sp*-hybridized (**Edit Properties of Atom #**) and the program colored them in a characteristic green color. The properties of CH (73.60, 6.10) were left without any change, as its hybridization might be both sp^2 and sp^3 . The methylene at 63.9 received the label “*ob*”, while all remaining methylene groups (including that at C 35.5) were labeled by the program with “*fb*” on the basis of the Atom Property Correlation Table (APCT). Finally, the phosphorus-containing fragment was drawn onto the MCD by hand as was suggested by authors.

An attempt to perform structure generation immediately gave zero structures. It was suggested that it was worth performing the next attempt with the APCT switched off considering an unprecedented structure for the unknown. The next run was performed in the FSG mode of **Determine Options Automatically** with the result: $k = 104 \rightarrow 104 \rightarrow 52$, $t_g = 0.2$ s. Connectivity lengths were elongated by one chemical bond ($m = 1$). Structural file ranking led to selection of the best structure with the average deviation range of 10–12 ppm, which can be considered to indicate

1	2	3
$d_A(^{13}\text{C}): 11.036$ $d_N(^{13}\text{C}): 12.612$ $d_I(^{13}\text{C}): 10.841$ $\text{max}_dA(^{13}\text{C}): 49.550$	$d_A(^{13}\text{C}): 11.076$ $d_N(^{13}\text{C}): 13.321$ $d_I(^{13}\text{C}): 11.749$ $\text{max}_dA(^{13}\text{C}): 45.420$	$d_A(^{13}\text{C}): 12.773$ $d_N(^{13}\text{C}): 14.113$ $d_I(^{13}\text{C}): 12.577$ $\text{max}_dA(^{13}\text{C}): 54.850$

Fig. 5.28 Phosphodiodyn A: Top structures of the ranked output file at $m = 2$

increasing the m value from 1 to 2. The third run gave the following result: $k = 48 \rightarrow 48 \rightarrow 12$, $t_g = 1.2$ s, and Fig. 5.28 shows the top structures of the ranked output file.

Figure 5.28 shows that both the average and maximum deviations are extremely high, which can be considered as a hint to the need to revise the initial set of axioms. As shown in Fig. 5.27, two carbons with overlapped signals in the ^{13}C spectrum at 35.5 ppm are labeled as carbons for which neighboring with heteroatoms is forbidden. Meanwhile, one of these atoms, C1, is a neighbor of a hydroxyl group in structure 5.25 according to the authors [13] suggestion. Therefore, to provide for the generation of structure 5.25 and its retention during spectral filtering it is necessary to remove the label “*fb*” at atoms C 35.5 (select “not defined” in the dialog window **Edit Properties of Atom #**) and switch off APCT (choose “not used” in the dialog window **CSB Generator Options**) and **Filter**. Results of FSG: $k = 101 \rightarrow 101 \rightarrow 20$, $t_g = 1$ s, 2 from 26 connectivities have been extended during generation, 94 from 325 possible connectivity combinations have been used during generation. The three top structures of the ranked output file are presented in Fig. 5.29.

1	2	3
$d_A(^{13}\text{C}): 2.067$ $d_N(^{13}\text{C}): 3.313$ $d_I(^{13}\text{C}): 3.410$ $\text{max}_dA(^{13}\text{C}): 19.480$	$d_A(^{13}\text{C}): 2.991$ $d_N(^{13}\text{C}): 4.283$ $d_I(^{13}\text{C}): 4.160$ $\text{max}_dA(^{13}\text{C}): 27.910$	$d_A(^{13}\text{C}): 6.782$ $d_N(^{13}\text{C}): 7.233$ $d_I(^{13}\text{C}): 7.380$ $\text{max}_dA(^{13}\text{C}): 43.990$

Fig. 5.29 Phosphodiodyn A: Top structures of the ranked output file with FSG ($m = 2$) under the condition that an oxygen atom may be a neighbor of C 35.5

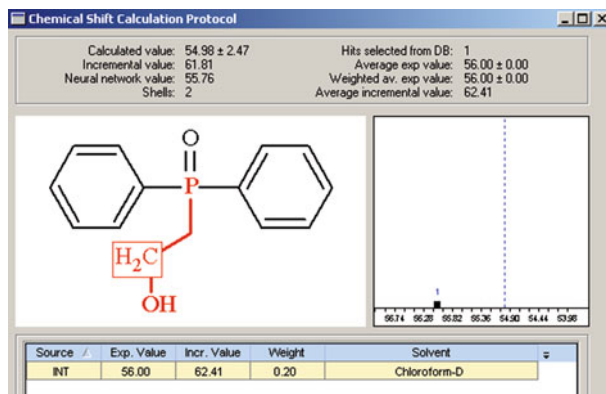


Fig. 5.30 Phosphiodyn A: Chemical shift calculation protocol generated for the carbon $\text{CH}_2\text{-OH}$ of structure #1 presented in Fig. 5.29

We see that the first ranked structure does coincide with the authors' [13] structural suggestion and the average deviations are acceptable if we take into account that the structure is unprecedented, but the maximum deviation (19.5 ppm) is too high. The ^{13}C chemical shift prediction protocol for $-\text{CH}_2\text{-OH}$ is presented in Fig. 5.30.

Figure 5.30 shows that only one hit was found in the database for the ^{13}C chemical shift calculation and the predicted value is 55 ppm, that is, the difference between the experimental and calculated values is ~ 20 ppm. On the other hand, as can be seen from Table 5.13, the predictions for the other carbon atoms are excellent. In addition, the following two reference structures were found in the ACD/CNMR database (Fig. 5.31):

Thus, we can conclude that the chemical shift 35.5 ppm cannot be assigned to the methylene group $\text{CH}_2\text{-OH}$ and therefore either the NMR experimental data (Table 5.13) are erroneous or erroneous is the structure 5.25.

The results obtained with Structure Elucidator were sent to the authors [13]. As a result of compound reinvestigation, the authors revised the structures both of phosphiodyn A and B (structures are very similar [14]). It turned out that the

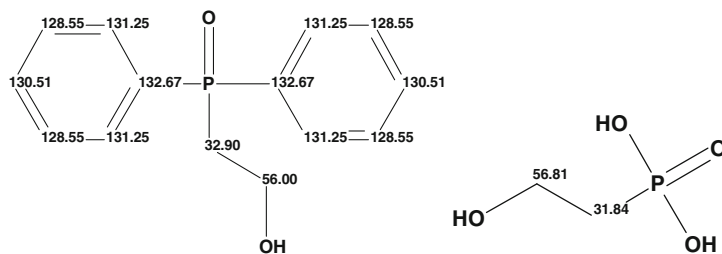


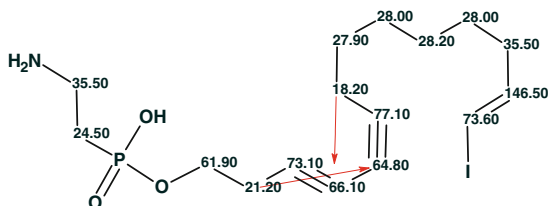
Fig. 5.31 Phosphiodyn A: Reference structures found in the ACD/CNMR database

1	2	3
$d_A(^{13}\text{C}): 0.723$ $d_N(^{13}\text{C}): 1.905$ $d_I(^{13}\text{C}): 1.813$ $\text{max}_A(^{13}\text{C}): 4.550$	$d_A(^{13}\text{C}): 2.201$ $d_N(^{13}\text{C}): 3.432$ $d_I(^{13}\text{C}): 3.053$ $\text{max}_A(^{13}\text{C}): 14.420$	$d_A(^{13}\text{C}): 6.147$ $d_N(^{13}\text{C}): 6.589$ $d_I(^{13}\text{C}): 6.711$ $\text{max}_A(^{13}\text{C}): 17.300$

Fig. 5.32 Phosphodiodyn A: Top structures of the ranked output file generated from the molecular formula $\text{C}_{16}\text{H}_{25}\text{O}_3\text{NPI}$

reason for deducing the wrong structure was a mistaken assumption about the possible chemical composition of the unknown. The “axiom” regarding the chemical composition of the unknowns was revised by the authors: the possibility of the presence of a nitrogen atom was allowed. With the new axiom (chemical composition CHNOPI) and the pseudomolecular ion peak at m/z $[\text{M}+\text{H}]^+$ 438.0692, the molecular formula $\text{C}_{16}\text{H}_{25}\text{O}_3\text{NPI}$ was generated and reliably selected as the most probable one with $\Delta m = 0.0003$. The revised molecular formula was input into the program and a new MCD was created. In contrast to the previous MCD (Fig. 5.27), no labels specifying the possibility of CH_2 35.5 carbons to have neighboring heteroatoms were assigned. This is because the program took into account the fact that a nitrogen atom is present in the molecular formula (i.e., the presence of the fragment $\text{CH}_2\text{-N}$ is allowed). FSG with parameters used during the last program run ($m = 2$, $a = 1$) gave the following result: $k = 101 \rightarrow 20$, $t_g = 1$ s. The three top structures of the ranked output file are shown in Fig. 5.32.

Figure 5.32 shows that the best structure is characterized by small average and maximum deviations and it can be reliably selected as the real structure of phosphodiodyn A (5.26) which coincides with the revised structure [14].



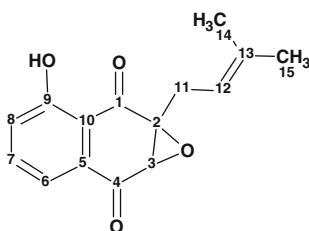
5.26

Here red arrows denote two $^4J_{\text{CH}}$ HMBC correlations. It is worth noting that only two structures (#1 and #2 in Fig. 5.32) were generated in 0.2 s when FSG was repeated with the APCT and **Filter** switched on.

The example demonstrates how a wrong assumption (axiom) regarding the conceivable chemical composition leads to inference of an erroneous structure. At the same time, the example underlines once more that chemical shift prediction for a structural hypothesis allows one to reveal the fact that the structure is erroneous and hints at the need to revise all experimental data starting from the chemical composition and molecular formula.

5.14 (2S,3R)-2,3-Epoxy-2,3-Dihydro-8-Hydroxylapachol

Whitson and co-workers [15] isolated from *Barleria alluaudii* and identified the naphthoquinone epoxide **5.27**, a compound not previously isolated from natural sources.



5.27

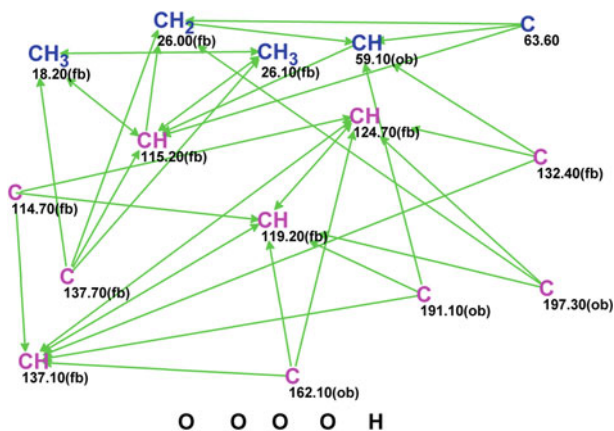
^1H , ^{13}C , HSQC, and HMBC NMR spectra (Table 5.14) of this natural product were used as input data to the Structure Elucidator.

The molecular formula for **5.27**, $\text{C}_{15}\text{H}_{14}\text{O}_4$, was derived from the HRESIMS ion at m/z 259.0966 ($[\text{M}+\text{H}]^+$; $\Delta + 0.39$ ppm). Initial interpretation of the NMR data indicated that **5.27** contains seven quaternary carbons, five methines, one methylene, and two methyls. The MCD for which atom properties were slightly edited is presented in Fig. 5.33.

MCD overview All carbon atoms for which $\delta\text{C} > 100$ ppm were assigned as sp^2 -hybridized and labels showing the possibility of connection with a heteroatom were set according to their characteristic ^{13}C and ^1H chemical shifts. The methine C 59.6 (δH 3.79) was assigned the label “ob.” MCD checking was completed with the program message “The minimum number of non-standard connectivities is 1.” As the molecule under investigation is small, an attempt was undertaken to perform FSG with automatically set options and ^{13}C NMR chemical shift prediction during the structure generation. Results: $k = 1,266 \rightarrow 2 \rightarrow 2$, $t_g = 6$ h 26 min, 7 from 31

Table 5.14 (2S,3R)-2,3-Epoxy-2,3-dihydro-8-hydroxylapachol: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	C HMBC
C1	197.3	198.06	C	–	–	–
C2	63.6	60.75	C	–	–	–
C3	59.1	59.6	CH	3.79	u	C2, C4, C11, C5
C4	191.1	192.56	C	–	–	–
C5	132.4	133.15	C	–	–	–
C6	119.2	119.1	CH	7.44	u	C1, C4, C9, C10, C8, C7
C7	137.1	137.06	CH	7.55	u	C10, C6, C4, C9, C8, C5
C8	124.7	124.32	CH	7.18	u	C9, C7, C6, C10, C1, C5
C9	162.1	160.89	C	–	–	–
C10	114.7	117.38	C	–	–	–
C11	26	28.38	CH ₂	3.04	u	C12, C13, C3, C1, C2
C11	26	28.38	CH ₂	2.58	u	C1, C12, C13, C2, C3
C12	115.2	115.7	CH	5.01	u	C14, C3, C2, C15, C13
C13	137.7	137.13	C	–	–	–
C14	18.2	17.99	CH ₃	1.63	u	C12, C15, C13
C15	26.1	25.8	CH ₃	1.68	u	C13, C12, C14

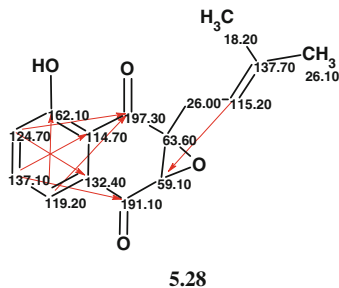
**Fig. 5.33** (2S,3R)-2,3-Epoxy-2,3-dihydro-8-hydroxylapachol: Slightly edited MCD

connectivities have been extended during generation, and all 2,629,575 possible connectivity combinations were used. The ranked output structures are presented in Fig. 5.34.

1	2
$d_A(^{13}\text{C}): 0.979$ $d_N(^{13}\text{C}): 1.579$ $d_I(^{13}\text{C}): 2.529$ $\text{max}_dA(^{13}\text{C}): 2.850$	$d_A(^{13}\text{C}): 3.933$ $d_N(^{13}\text{C}): 4.669$ $d_I(^{13}\text{C}): 3.508$ $\text{max}_dA(^{13}\text{C}): 12.040$

Fig. 5.34 (2S,3R)-2,3-Epoxy-2,3-dihydro-8-hydroxylapachol: Ranked output structures

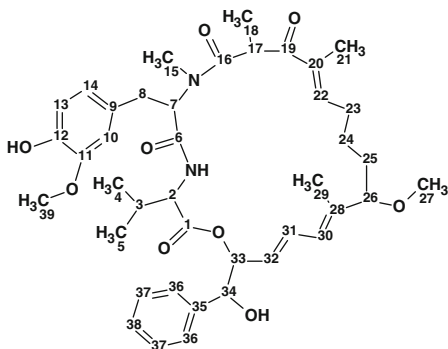
We see that the best structure coincides with structure **5.27**, and the ^{13}C chemical shift assignment along with 7 nonstandard HMBC connectivities (arrows) are shown on structure **5.28**.



So, an unambiguous structure elucidation was carried out automatically in the presence of seven nonstandard connectivities but it took a long time (~ 6.5 h) for structure generation. The molecule is small, consists of well-known fragments (no unusual elements contained within the skeleton) but its solution was rather time-consuming because it was necessary to undertake ca. 2.6 million attempts to generate structures only during the last step when the number of lengthened connectivities (parameter m) was equal to 7. No structure was obtained as a result of the FSG and structural filtering at $m = 1-6$, therefore in reality many millions of attempts to generate plausible structures were made. Taking into account that time-consuming structure generation can be run overnight (this approach is recommended), the associated computational cost is not too high.

5.15 Aetheramide

Aetheramides are structurally distinctive cyclic peptides isolated from a novel myxobacterial genus proposed to be termed “Aetherobacter”. These natural products are characterized by both their unique structures and their novel modes of action. Plaza et al. [16] isolated two new unusual depsipeptides, termed aetheramides A and B. Their structures were elucidated using extensive NMR (^1H , ^{13}C , HSQC, COSY, HMBC, HOHAHA) and ESIMS analysis, chemical derivatizations, and by the combined analysis of homonuclear (H–H) and heteronuclear $^{2,3}J_{\text{CH}}$ -couplings and quantum mechanical calculations. Here we will discuss the computer-assisted structure elucidation of aetheramide B (**5.29**).



5.29

The HRESIMS of aetheramide B displayed an $[\text{M}+\text{H}]^+$ peak at m/z 719.3920 consistent with the molecular formula $\text{C}_{41}\text{H}_{54}\text{N}_2\text{O}_9$ (with a calculated mass of 719.3908). The degree of unsaturation is equal to 16. The IR spectrum displayed absorption bands at 3,367 and 1,673 cm^{-1} , suggesting the presence of OH or/and NH (3,367 cm^{-1}) and carbonyl (most probably amide) functionalities. The NMR spectroscopic data used for CASE analysis are presented in Table 5.15.

The MCD (Fig. 5.35) is modified slightly in accordance with observed characteristic spectral features.

MCD overview The degree of unsaturation is fairly large (16) but the molecular structure produces a large number of detected HMBC correlations. Four carbonyl bonds are evident and two O–CH₃ bonds were drawn by hand to reduce the time required for structure generation. Carbon atoms with ^{13}C chemical shifts in the range 54–85 ppm were supplied with the property sp^3/ob , which was supported by the corresponding ^1H chemical shifts (see Table 5.15). Since the neighborhood of a nitrogen atom can vary the chemical shifts of sp^2 carbons over a wide interval, the properties of the carbon atoms at 112.5–146.9 were marked as sp^2/nd (“nd” means *not defined* and must be determined automatically during the structure generation). The methyl group CH₃ (δC 31.10; δH 2.96) is most probably chemically bonded to

Table 5.15 Aetheramide B: NMR spectroscopic data

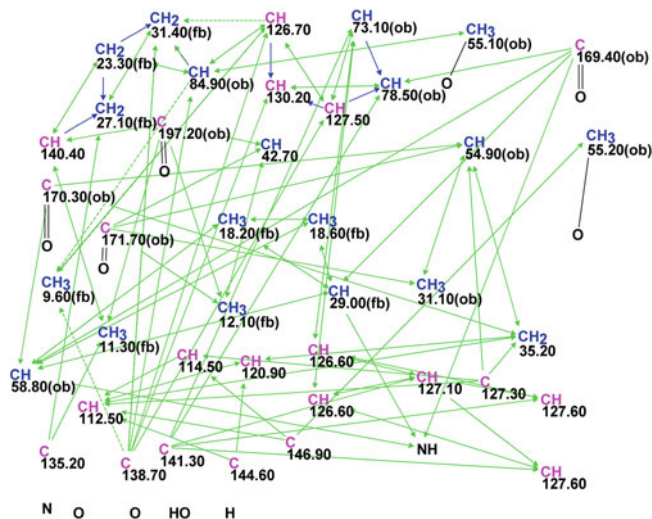
Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C1	169.4	171.09	C	–	–	–	–
C2	58.8	57.91	CH	3.92	u	–	C6, C5, C1, C4, C3
C3	29	31.05	CH	1.91	u	–	C4, C2, C5, C1
C4	18.6	17.58	CH ₃	0.56	d(6.8)	–	C5, C2, C3
C5	18.2	18.65	CH ₃	0.78	d(6.8)	–	C4, C3, C2
C6	170.3	169.9	C	–	–	–	–
C7	54.9	56.29	CH	5.5	s	–	C8, C16, C15, C9, C6
C8	35.2	35.62	CH ₂	3.07	u	–	–
C8	35.2	35.62	CH ₂	2.77	u	–	C10, C6, C7, C9, C14
C9	127.3	130.26	C	–	–	–	–
C10	112.5	111.82	CH	6.77	u	–	C14, C13, C8, C9, C12, C11
C11	146.9	146.34	C	–	–	–	–
C12	144.6	145.24	C	–	–	–	–
C13	114.5	114.85	CH	6.6	d(8)	–	C9, C11
C14	120.9	122.05	CH	6.59	d(8)	–	C12, C8, C10
C15	31.1	31.49	CH ₃	2.96	s	–	C16, C7
C16	171.7	171.19	C	–	–	–	–
C17	42.7	45.86	CH	4.24	u	–	C18, C19, C16
C18	12.1	13.31	CH ₃	0.58	d(6.7)	–	C19, C17, C16
C19	197.2	197.54	C	–	–	–	–
C20	135.2	137.3	C	–	–	–	–
C21	11.3	10.09	CH ₃	1.65	s	–	C22, C20, C19
C22	140.4	144.7	CH	6.45	u	2.17	C24, C21, C19, C23
C23	27.1	29.03	CH ₂	2.17	u	6.45, 1.11	–
C23	27.1	29.03	CH ₂	1.99	u	–	C20, C25, C22, C24
C24	23.3	24.25	CH ₂	1.11	u	1.42, 2.17	C22, C23, C25, C26
C25	31.4	32.11	CH ₂	1.42	u	3.48, 1.11	C28, C30, C26
C25	31.4	32.11	CH ₂	1.55	u	–	C26, C28, C23, C24
C26	84.9	80.54	CH	3.48	u	1.42	C25, C28, C24, C30, C27
C27	55.1	55.52	CH ₃	3.03	u	–	C26
C28	138.7	134.66	C	–	–	–	–
C29	9.6	19.45	CH ₃	1.55	s	–	–
C30	126.7	129.1	CH	5.8	d(11.2)	6.44	C26, C29, C28, C32, C31
C31	130.2	133.42	CH	6.44	u	5.80, 5.60	C33, C28, C30, C32
C32	127.5	130.12	CH	5.6	u	5.38, 6.44	C28, C34, C33, C30
C33	78.5	79.31	CH	5.38	u	5.60, 4.77	C1, C34, C31, C35, C32
C34	73.1	74.65	CH	4.77	u	5.38	C32, C35, C36, C33
C35	141.3	138.66	C	–	–	–	–
C36	126.6	126.32	CH	7.39	d(7.5)	–	C34, C37, C38
C37	127.6	127.93	CH	7.32	t(7.5)	–	C38, C36, C35

(continued)

Table 5.15 (continued)

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C38	127.1	128.88	CH	7.23	t(7.5)	–	C36, C35
C39	55.2	55.99	CH ₃	3.71	u	–	C11
N1	100 ^a	–	NH	8.13	u	–	C3, C2, C1
O1	50 ^a	–	OH	8.67	u	–	–

^a Chemical shifts of nitrogen and oxygen atoms (marked by a) are conventionally assigned to formally distinguish between different heteroatoms of the same type

**Fig. 5.35** Aetheramide B: Slightly edited MCD

a nitrogen atom, but there are two nitrogen atoms (N and NH) in the molecule and it is not known which of them should be chosen as the associated atom, hence the methyl group was left without any chemical bond.

MCD checking revealed the presence of a minimum of one nonstandard connectivity. Taking this fact into account, FSG was initiated using the mode **Determine Options Automatically** accompanied with ¹³C chemical shift prediction and structure filtering. The result gave: $k = 53 \rightarrow 6 \rightarrow 2$, $t_g = 15 \text{ min } 10 \text{ s}$, 3 of the 75 correlations were extended during the generation process and 12,785 from 67,525 (~20%) possible connectivity combinations were used. The structures elucidated are shown in Fig. 5.36.

A comparison with structure 5.29 shows that structure #1 ranked as first by all three methods is aetheramide B. The differences between deviations related to structures #1 and #2 are small because both structures are very similar: they differ only in the positions of the OH and CH₃ groups on the benzene ring. Difficulties associated with the presence of three NSCs in the HMBC data were overcome by

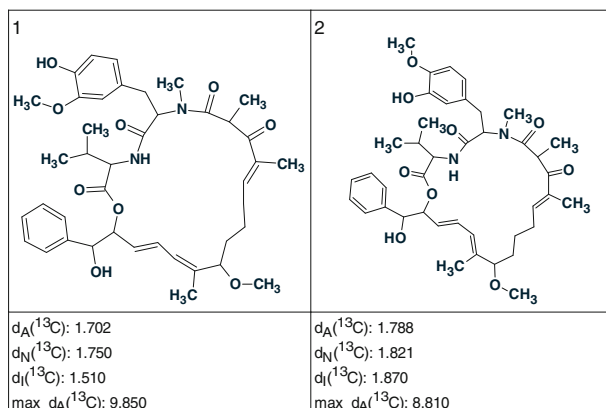
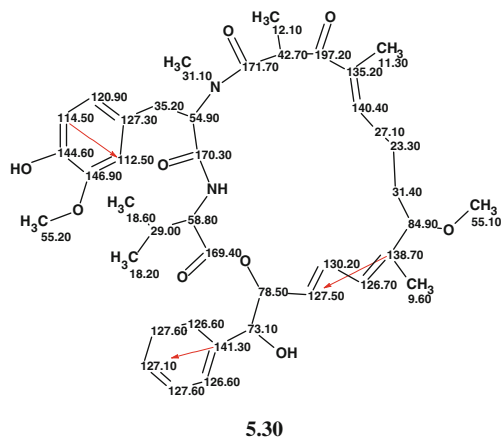


Fig. 5.36 Aetheramide B: Ranked output structure file

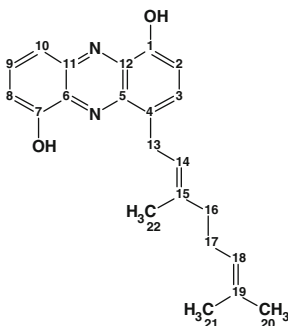
the program in automatic mode. The structure of aetheramide B (**5.30**) supplied with the chemical shift assignments and with drawn NSCs included is shown below:



Comparison of the experimental and calculated ^{13}C chemical shifts (listed in Table 5.15) leads to the conclusion that chemical shift prediction performed very well.

5.16 Geranylphenazinediol

Ohlendorf and co-workers [17] describe the isolation and structure elucidation of Geranylphenazinediol (**5.31**), a phenazinediol substituted with an isoprenoid side chain.



5.31

This new phenazine natural product was produced by the *Streptomyces* sp. strain LB173, which was isolated from a marine sediment sample. The structure was established by analysis of NMR and MS data. To identify the structure of a new compound, the authors [17] used the known fact that the phenazine core structure has a characteristic UV spectrum, and substitution with hydroxyl groups leads to a bathochromic shift. With UV maxima at 204, 273, 354, 374, and 458 nm, the UV spectrum of **5.31** was almost identical to that of 1,6-phenazinediol. The molecular formula $C_{22}H_{24}N_2O_2$ was established by HRESIMS, which was congruent with the NMR data. The ^{13}C NMR spectrum corresponded to a total of 22 individual carbons. A common set of NMR data (1H , ^{13}C , DEPT, 1H - ^{13}C HSQC, 1H - 1H COSY, and 1H - ^{13}C HMBC) was employed for structure elucidation.

We used StrucEluc to elucidate the structure of this new natural product as if “ab initio” under the condition that only NMR spectroscopic data (^{13}C , 1H , HSQC, and HMBC) are available and assuming that the similar structure of 1,6-phenazinediol with similar spectra is unknown.

The spectroscopic data input into StrucEluc are presented in Table 5.16.

The MCD used for structure generation is shown in Fig. 5.37.

MCD Overview The chemical shifts associated with two hydroxyl groups (9.01 and 9.20 ppm) were determined in the 1H NMR spectrums. An IR spectrum of compound **5.31** was not provided in article [17], but the assignment could be confirmed by using the IR spectrum as stretching vibrations for OH and NH groups are observed at different frequencies in a CCl_4 solution [18]. The most downfield ^{13}C chemical shifts at 151.7 and 153.8 are typical for hydroxylated aromatic carbons, so it was assumed that OH groups were connected with the corresponding carbons using drawn chemical bonds. These user constraints (“axioms”) significantly reduce the dimension of the problem, as the program will not try to connect nitrogen atoms with these carbons. The number of hydrogen atoms $n(H)$ attached to the neighboring carbon atoms were added to the atom properties in accordance with the 1H multiplicities shown in Table 5.16. It is known that the chemical shifts of sp^2 -hybridized carbons connected to a nitrogen atom can resonate over a chemical shift value range. The atom properties displayed on the MCD allow nitrogen atoms

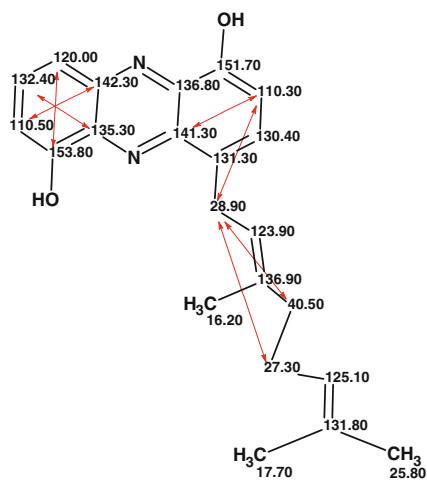
Table 5.16 Geranylphenazinediol: NMR spectroscopic data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C1	151.7	151.89	C	–	–	–	–
C2	110.3	113.55	CH	7.17	d(7.7)	7.61	C12, C5, C1, C4
C3	130.4	130.12	CH	7.61	d(7.7)	4.05, 7.17	C2, C5, C1, C13
C4	131.3	129.56	C	–	–	–	–
C5	141.3	142.64	C	–	–	–	–
C6	135.3	134.15	C	–	–	–	–
C7	153.8	153.42	C	–	–	–	–
C8	110.5	109.49	CH	7.25	d(7.70)	7.73, 7.82	C10, C11, C7, C6
C9	132.4	131.39	CH	7.82	dd (8.7;7.7)	7.25, 7.73	C7, C8, C6, C11
C10	120	119.38	CH	7.73	d(8.90)	7.25, 7.82	C7, C6, C9, C8
C11	142.3	142.43	C	–	–	–	–
C12	136.8	134.15	C	–	–	–	–
C13	28.9	29.91	CH ₂	4.05	d(7.40)	1.80, 2.10, 5.54, 7.61	C3, C16, C4, C5, C15, C17, C14, C2
C14	123.9	122.48	CH	5.54	u	1.80, 2.10, 4.05	C13, C16, C22
C15	136.9	135.89	C	–	–	–	–
C16	40.5	39.49	CH ₂	2.1	u	2.13, 5.54, 4.05	C14, C22, C17, C15, C18
C17	27.3	26.5	CH ₂	2.13	u	1.55, 1.59, 2.10, 5.10	C15, C19, C18, C16
C18	125.1	124.4	CH	5.1	u	1.55, 1.59, 2.13	C21, C20, C17
C19	131.8	131.6	C	–	–	–	–
C20	25.8	25.65	CH ₃	1.59	s	2.13, 5.10	C18, C19, C21
C21	17.7	17.75	CH ₃	1.55	s	2.13, 5.10	C19, C20, C18
C22	16.2	15.98	CH ₃	1.8	s	4.05, 5.54	C14, C16, C15
O1	100 ^a	–	OH	9.01	u	–	–
O2	110 ^a	–	OH	9.2	u	–	–

^a Fictitious ¹⁷O chemical shift values

to be connected with any of the quaternary carbons over the chemical shift range of 135.3–142.3 ppm.

MCD checking resulted in the detection of contradictions in the HMBC data and the program reported that the minimum number of nonstandard connectivities is 1. A posterior checking of structure **5.31** supplied with the assigned ¹³C chemical shifts allowed us to establish seven HMBC nonstandard connectivities, one of which is 4 bonds in length (⁵J_{CH}), as shown on structure **5.32**.



5.32

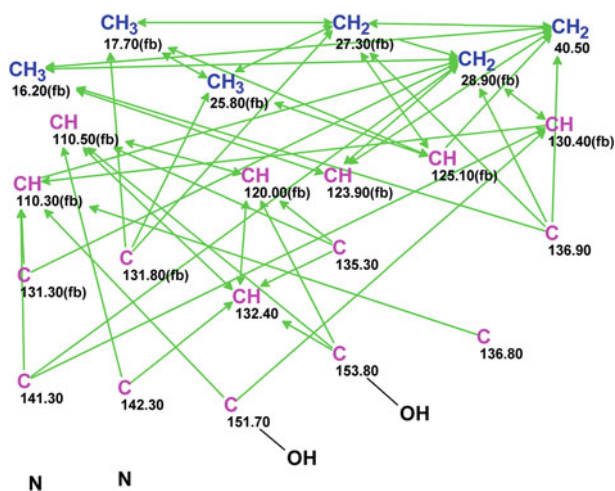


Fig. 5.37 Geranylphenazinediol: The MCD of Geranylphenazinediol

To overcome the contradictions in the HMBC data FSG was initiated with the following options (Fig. 5.38):

Figure 5.38 shows that FSG was run in the safest mode ($m = 1-20$, $a = 16$) when the lengthening correlations are replaced by their deletion. FSG was stopped by the program when calculations with $m_g = 6$ were completed (i.e., attempts to generate structures with $m = 1-5$ yielded no structures) with the result: $k = 2,503 \rightarrow 6 \rightarrow 6$, $t_g = 1 \text{ h } 32 \text{ min}$; 2,142,400 from 4,496,388 possible combinations of connectivities

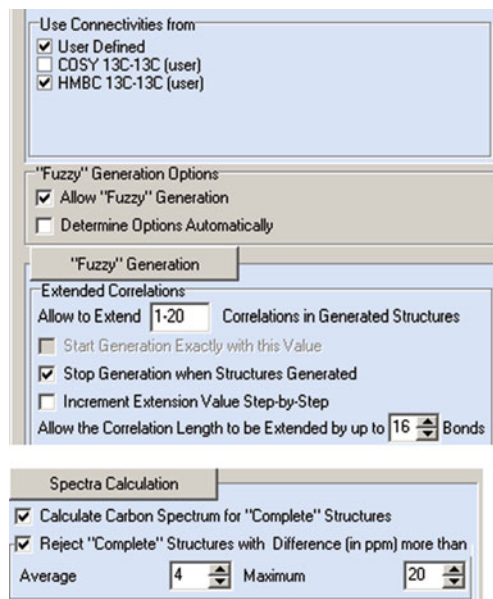


Fig. 5.38 Geranylphenazinediol: Options of FSG applied to Geranylphenazinediol

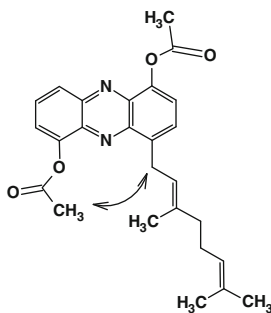
were used for attempts to generate structures at $m_g = 6$ and this is the reason for the long generation time.

The beginning of the ranked output file is shown in Fig. 5.39. The figure shows that the best structure with minimal deviations calculated by all three methods coincides with structure **5.31**.

It was interesting to find out why the correct solution was found at $m_g = 6$, while the number of NSCs is 7. The answer gives a comparison of shift assignments which were automatically performed for the generated structure with the assignments suggested by the authors [17] (Fig. 5.40).

Automated assignment placed the chemical shift of 110.50 ppm in the C10 position because the experimental $^4J_{\text{HC}}$ connectivity of 110.5–142.3 (see structure **5.32**) was not deleted by the program using a setting of $m_g = 6$. Thus we obtained the *correct* structure but with a pair of exchanged ^{13}C chemical shifts.

The authors [17] report that "...the structure of **5.31** could be unequivocally proven with one exception: from the spectroscopic data it was *impossible* to discriminate between a 1,7-phenazinediol and a 1,10-phenazinediol core structure. The problem could be solved by NOESY correlations from the hydroxyl groups to methylene groups of the isoprenoid chains which were connected to the other aromatic ring. However, comparable NOESY correlations were not observed for **5.31**. In order to determine the constitution, **5.31** was acetylated using pyridine-acetic anhydride, which yielded compound **5.31a**.



5.31a

As expected, two additional signals appeared in the ^1H NMR spectrum accounting for one acetylmethyl group each. The respective NOESY spectrum showed no correlation of the two methyl signals to each other, but a correlation from one of the signals to CH_2 (28.9) of the isoprenoid side chain. Thus, we concluded the core structure to be 1,7-phenazinediol, a substitution pattern that is congruent with that of other naturally occurring phenazines.”

Figure 5.39 shows that the competing structure mentioned above which could not be distinguished from the correct one without additional experiments is ranked as *second*. Therefore, even the “approximate solution” found at $m_g = 6$ allowed the program to make the correct choice between the two most probable structures.

1	2	3
$d_A(^{13}\text{C}): 1.731$ $d_N(^{13}\text{C}): 2.020$ $d_I(^{13}\text{C}): 2.527$ $\text{max_}d_A(^{13}\text{C}): 10.510$	$d_A(^{13}\text{C}): 1.884$ $d_N(^{13}\text{C}): 2.078$ $d_I(^{13}\text{C}): 2.605$ $\text{max_}d_A(^{13}\text{C}): 10.240$	$d_A(^{13}\text{C}): 3.337$ $d_N(^{13}\text{C}): 2.833$ $d_I(^{13}\text{C}): 3.739$ $\text{max_}d_A(^{13}\text{C}): 12.150$
4	5	6
$d_A(^{13}\text{C}): 5.283$ $d_N(^{13}\text{C}): 5.396$ $d_I(^{13}\text{C}): 3.549$ $\text{max_}d_A(^{13}\text{C}): 19.700$	$d_A(^{13}\text{C}): 5.462$ $d_N(^{13}\text{C}): 5.978$ $d_I(^{13}\text{C}): 3.907$ $\text{max_}d_A(^{13}\text{C}): 27.720$	$d_A(^{13}\text{C}): 5.810$ $d_N(^{13}\text{C}): 5.869$ $d_I(^{13}\text{C}): 3.990$ $\text{max_}d_A(^{13}\text{C}): 16.010$

Fig. 5.39 Geranylphenazinediol: The top structures of the ranked output file at $m_g = 6$ for Geranylphenazinediol

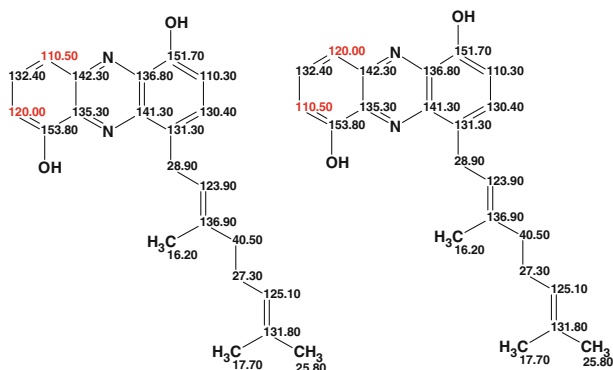


Fig. 5.40 Geranylphenazinediol: Comparison between automatic (*left*, $m_g = 6$) and manual (*right*, $m_g = 7$) carbon chemical shift assignments. The latter corresponds to the assignment suggested in [17]

For the sake of completeness, we repeated FSG with $m = m_g + 1 = 7$, i.e., the number of connectivities, which will be deleted during the structure generation, was set equal to 7. Result: $k = 26,625 \rightarrow 45 \rightarrow 23$, $t_g = 10$ h 44 min. During the process of structure generation, 12,669,412 from 22,481,940 possible connectivity combinations were used as the basis of attempts to generate structures and this significantly increased the generation time. Structure generation was performed overnight.

For comparison with the solution found at $m_g = 6$, the top of the last output file (for $m_g = 7$) is shown in Fig. 5.41.

The comparison of the two solutions—“approximate” and “exact”—shows that all deviations calculated for the best (and *correct*) structure of the “exact” solution are significantly smaller than in the previous case. This is in agreement with the new (and *correct*) automatic chemical shift assignment. The competing structures #2 and #3 are characterized by deviations that are also smaller than in the “approximate” solution due to permutation of the 120 and 110.5 ppm shifts.

5.17 Strynuxline A

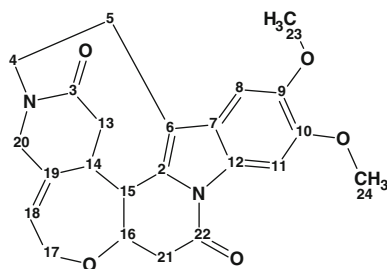
Strychnos nux-vomica is a moderate-sized tree of the family *Loganiaceae* found in southern Asian countries. The dried ripe seeds of this tree have been applied as a traditional folk medicine in China for the treatment of tumors, rheumatic arthritis, swelling pain, trauma, bone fractures, etc. Strychnine and brucine are major components, which are also mainly responsible for most of the pharmacological properties such as antitumor, cytoprotective, and antitussive activities. Studies aimed at the discovery of trace alkaloids in *S. nux-vomica* led Fu and co-workers

1	2	3
$d_A(^{13}\text{C})$: 0.924 $d_N(^{13}\text{C})$: 1.352 $d_I(^{13}\text{C})$: 2.038 $\text{max}_dA(^{13}\text{C})$: 3.250	$d_A(^{13}\text{C})$: 2.511 $d_N(^{13}\text{C})$: 2.300 $d_I(^{13}\text{C})$: 3.273 $\text{max}_dA(^{13}\text{C})$: 12.150	$d_A(^{13}\text{C})$: 2.842 $d_N(^{13}\text{C})$: 2.652 $d_I(^{13}\text{C})$: 3.581 $\text{max}_dA(^{13}\text{C})$: 12.610
4	5	6
$d_A(^{13}\text{C})$: 4.808 $d_N(^{13}\text{C})$: 3.884 $d_I(^{13}\text{C})$: 3.832 $\text{max}_dA(^{13}\text{C})$: 16.730	$d_A(^{13}\text{C})$: 5.215 $d_N(^{13}\text{C})$: 3.772 $d_I(^{13}\text{C})$: 3.810 $\text{max}_dA(^{13}\text{C})$: 30.720	$d_A(^{13}\text{C})$: 5.604 $d_N(^{13}\text{C})$: 4.555 $d_I(^{13}\text{C})$: 3.816 $\text{max}_dA(^{13}\text{C})$: 20.060

Fig. 5.41 Geranylphenazinediol: The top structures of the ranked output file at $m_g = 7$

[19] to the isolation of strynuxlines A and B, two novel alkaloids possessing an *unprecedented skeleton* with a 6/5/9/6/7/6 hexacyclic ring system. This was the first report of the cleavage of the C3–C7 bond in strychnan-type alkaloids. In addition to strynuxlines A and B, 16 known alkaloids were isolated and identified in this work.

Here we will describe how strynuxlines A (**5.33**) could be identified with the assistance of a CASE-based approach.



5.33

Strynuxline A (**1**) was obtained as a white, amorphous solid. Its molecular formula was determined to be $\text{C}_{23}\text{H}_{24}\text{N}_2\text{O}_5$, by HRESIMS (m/z 408.1673 $[\text{M}]^+$; calculated 408.1685), with an index of hydrogen deficiency of 13. The ^{13}C , ^1H , and

Table 5.17 Strynuxline A: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	C HMBC
C2	130.1	133.47	C	–	–	–	–
C3	174.9	171.59	C	–	–	–	–
C5	49.3	47.49	CH_2	3.17	u	2.8	C7, C3, C21
C5	49.3	47.49	CH_2	4.57	u	–	–
C6	25.7	22.64	CH_2	3.14	u	–	–
C6	25.7	22.64	CH_2	2.8	u	3.17	C7, C8
C7	117.7	113.15	C	–	–	–	–
C8	124.7	123.54	C	–	–	–	–
C9	100.4	100.28	CH	6.78	s	–	C7, C11, C8
C10	146.9	146.2	C	–	–	–	–
C11	148.1	146.88	C	–	–	–	–
C12	99.7	98.38	CH	7.88	s	–	C8, C10
C13	129.5	131.32	C	–	–	–	–
C14	39.5	35.25	CH_2	2.86	dd(15.1, 6.3)	–	–
C14	39.5	35.25	CH_2	2.48	d(15.1)	3.51	–
C15	41.5	35.44	CH	3.51	u	3.29, 2.48	C19, C2, C3
C16	48.7	49.95	CH	3.29	u	3.51, 4.13	–
C17	83.7	76.79	CH	4.13	u	2.77, 3.29	C7, C18, C15, C24, C16
C18	68	65.19	CH_2	4.10	d(13.8)	6.01	C20
C18	68	65.19	CH_2	4.26	dd(13.8, 7.0)	–	–
C19	123.3	126.68	CH	6.01	u	4.1	–
C20	145.1	133.39	C	–	–	–	–
C21	52.9	53.75	CH_2	4.19	u	–	C15, C19, C3
C21	52.9	53.75	CH_2	3.86	u	–	–
C23	42.3	42.3	CH_2	2.77	u	4.13	C24
C23	42.3	42.3	CH_2	3.15	u	–	–
C24	168	166.35	C	–	–	–	–
C25	56.3	56.34	CH_3	3.86	s	–	–
C26	56.3	56.47	CH_3	3.86	s	–	C11, C10

HSQC spectroscopic data, as well as key HMBC and COSY correlations presented in the article [19] graphically are given in Table 5.17.

Figure 5.42 shows the initial MCD created from the NMR spectroscopic data.

MCD overview The properties of only four carbon atoms C56.3–C83.7 are fully assigned by the program, while six carbons (marked by a light blue color) have an ambiguous state of hybridization (sp^3 or sp^2) and the possibility of a heteroatom neighborhood is also undefined. Taking into account the ^1H chemical shifts of hydrogens (7.88 and 6.78 ppm) attached to atoms C 99.7 and C 100.4 correspondingly, as well as the characteristic ^{13}C chemical shifts of the other four “light blue” quaternary carbons, a hybridization state of sp^2 can be assigned to all of them. Two absorption bands were observed at 1,670 and 1,640 cm^{-1} in the IR spectrum.

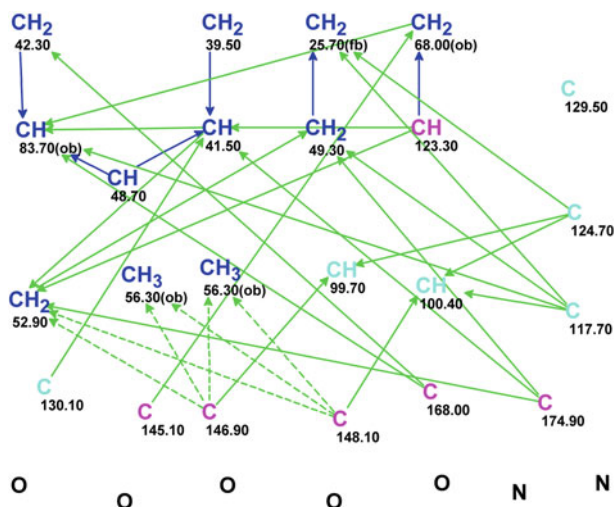


Fig. 5.42 Strynuxline A: The MCD

Therefore, it is possible to suggest that the carbons at C 168.0 and C 174.9 are connected to oxygen atoms with double bonds. Two methyl groups at C 56.3 can be safely connected to oxygen atoms. In the presence of two nitrogen atoms in the molecule, it would be risky to assign the additional labels “*ob*” or “*fb*” to carbon atoms displayed on the MCD. The numbers of hydrogen atoms present in the first sphere around eight carbon atoms were set on the MCD in agreement with the ^1H signal multiplicities shown in Table 5.17. The MCD edited in accordance with the considerations above is shown in Fig. 5.43.

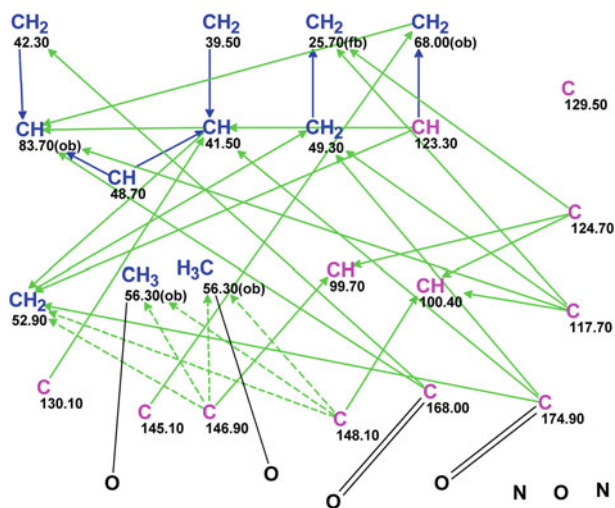


Fig. 5.43 Strynuxline A: The edited MCD

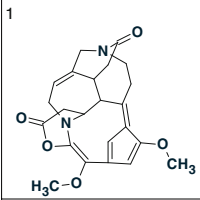
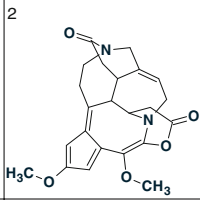
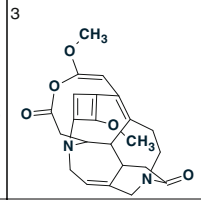
		
$d_A(^{13}\text{C}): 6.408$ $d_N(^{13}\text{C}): 6.073$ $d_I(^{13}\text{C}): 5.934$ $\text{max_}d_A(^{13}\text{C}): 26.700$	$d_A(^{13}\text{C}): 6.587$ $d_N(^{13}\text{C}): 6.395$ $d_I(^{13}\text{C}): 6.283$ $\text{max_}d_A(^{13}\text{C}): 28.450$	$d_A(^{13}\text{C}): 7.550$ $d_N(^{13}\text{C}): 7.355$ $d_I(^{13}\text{C}): 6.357$ $\text{max_}d_A(^{13}\text{C}): 24.000$

Fig. 5.44 Strynuxline A: The three top structures of the ranked output file produced by strict structure generation

MCD checking produced a program message indicating that no contradictions were detected in the 2D NMR data. Strict structure generation was initiated which gave the following result: $k = 132 \rightarrow 94 \rightarrow 77$, $t_g = 0.27$ s. The three top structures of the ranked output file are presented in Fig. 5.44.

The large values of the average deviations, as well as the “exotic” structure skeletons, suggest that the solution is wrong and the next most rationale step is an attempt at FSG. FSG with the options $m = 1$, $a = 16$ was completed with the results: $k = 5,142 \rightarrow 3,411 \rightarrow 1,883$, $t_g = 25$ s, 30 out of 30 possible connectivity combinations were used during the generation process. The first three structures of the ranked structural file are shown in Fig. 5.45.

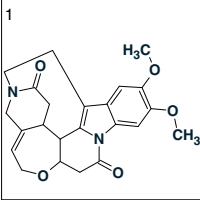
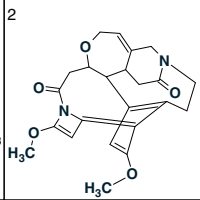
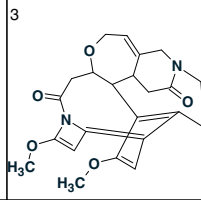
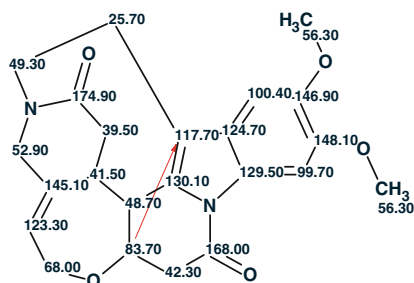
		
$d_A(^{13}\text{C}): 2.673$ $d_N(^{13}\text{C}): 2.868$ $d_I(^{13}\text{C}): 2.439$ $\text{max_}d_A(^{13}\text{C}): 11.710$	$d_A(^{13}\text{C}): 3.987$ $d_N(^{13}\text{C}): 4.618$ $d_I(^{13}\text{C}): 4.805$ $\text{max_}d_A(^{13}\text{C}): 16.970$	$d_A(^{13}\text{C}): 4.693$ $d_N(^{13}\text{C}): 5.409$ $d_I(^{13}\text{C}): 5.323$ $\text{max_}d_A(^{13}\text{C}): 16.790$

Fig. 5.45 Strynuxline A: The three top structures of the ranked output file produced by FSG

The first ranked structure coincides with structure **5.33** derived by the authors [19]. The ^{13}C chemical shift assignment is shown on structure **5.34** (an arrow shows the detected nonstandard HMBC correlation):

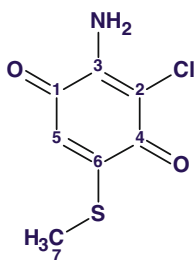


5.34

Despite a very unusual skeleton, the structure of strynuxline A and its chemical shift assignment was unambiguously determined by the program using the FSG mode.

5.18 Hyaladione

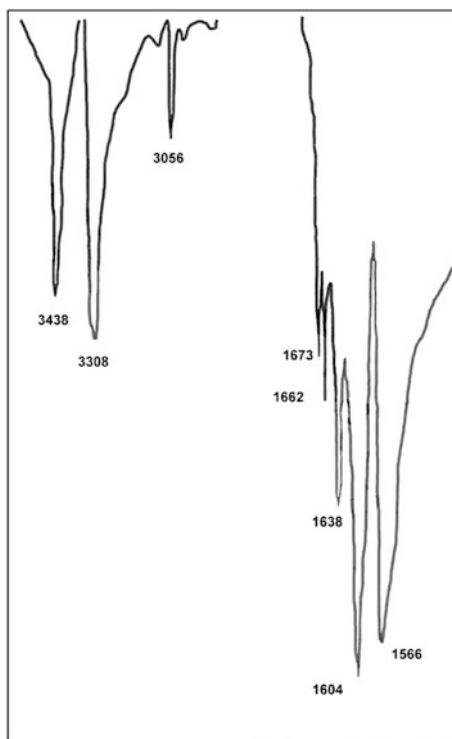
Within the framework of the screening of a myxobacterial strain collection for biologically active compounds, Okanya et al. [20] isolated hyaladione (**5.35**)—the first biologically active secondary metabolite from *Hyalangium minutum*, strain NOCB-2T. The structure of **5.35** was elucidated by extensive HRESIMS, NMR, and X-ray crystallographic analysis.



5.35

The molecular formula $\text{C}_7\text{H}_6\text{ClNO}_2\text{S}$ requiring five degrees of unsaturation was established by ultrahigh-resolution ESI-TOF-MS in the positive-ion mode from a molecular ion peak (HRESIMS m/z 203.9882 $[\text{M}+\text{H}]^+$ (calculated for $\text{C}_7\text{H}_6\text{ClNO}_2\text{S}$ $[\text{M}+\text{H}]^+$, 203.9880). Supporting the empirical formula, all seven carbon atoms were

Fig. 5.46 Hyaladione: A part of the IR spectrum containing characteristic absorption bands



present in the ^{13}C NMR spectrum of **5.35** in acetone- d_6 . The IR spectrum (Fig. 5.46) shows two strong bands at ν_{max} 1,604 and 1,566 cm^{-1} which were assigned by the authors [20] to α,β -unsaturated carbonyl groups.

Additionally, the IR spectrum revealed two sharp bands at ν_{max} 3,438 and 3,308 cm^{-1} . The bands are typical for an NH_2 group. An IR peak at 3,056 is definitely associated with stretching vibrations of a $=\text{C}-\text{H}$ bond.

The ^1H NMR spectrum of **5.35** in acetone- d_6 presented only three singlet signals: of a methyl group (δH 2.41, CH_3 -7), a methine proton (δH 6.26, H5), and a broad singlet of an amine (δH 6.81). COSY and HMBC spectra of the analyzed compound were also acquired and used for structure elucidation (Table 5.18).

The authors [20] concluded that all correlations from the COSY, HMBC, and ROESY spectra (see Fig. 5.47) were insufficient for a complete structure elucidation.

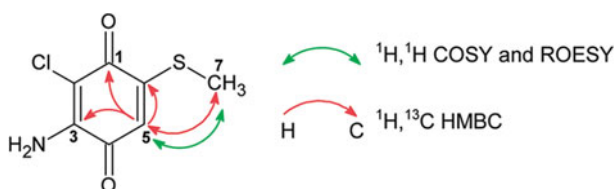
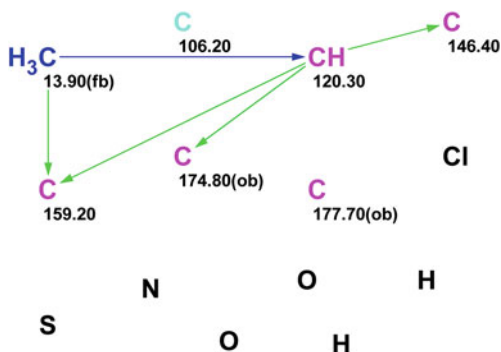
The compound was crystallized from acetone to obtain pink, needle-shaped crystals for an unambiguous structure elucidation by X-ray analysis. It was interesting to verify if the 2D NMR data indeed are insufficient for a complete elucidation of structure **5.35**. With this in mind, the spectroscopic data presented in Table 5.18 were input into StrucEluc and the MCD was created (Fig. 5.48).

MCD overview The hybridization of the atom C 106.2 (colored light blue) is assigned as “ sp^3 or sp^2 ” by the program. The methyl group of the substructure $\text{S}-\text{CH}_3$ is automatically supplied with a label “ fb ”, so one could suggest that

Table 5.18 Hyaladione: The NMR spectroscopic data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	C HMBC
C1	174.8	178.2	C	–	–	–	–
C2	106.2	98.85	C	–	–	–	–
C3	146.4	149.75	C	–	–	–	–
C4	177.7	177.26	C	–	–	–	–
C5	120.3	117.99	CH	6.26	s	2.41	C1, C3, C6
C6	159.2	156.24	C	–	–	–	–
C7	13.9	15.1	CH ₃	2.41	s	6.26	C6, C5
N	100 ^a	–	NH ₂	6.81	s	–	–

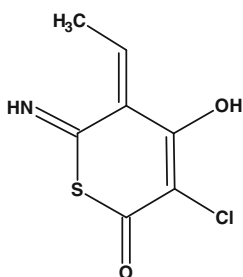
^a Fictitious ¹⁵N NMR chemical shift

**Fig. 5.47** Hyaladione: The COSY and HMBC correlations presented graphically in the work [20]**Fig. 5.48** Hyaladione: The MCD

structure **5.35** will never be generated as the CH₃ group is connected to the S atom. However, the algorithms of StrucEluc take into account that the sulfur atom S influences chemical shifts of neighboring carbons in a similar way to a carbon atom. So the S atom is considered as a non-heteroatom within the program. The number of connectivities shown on the MCD is small, but at the same time the diversity of the heteroatoms (four different chemical types of heteroatoms are present in the molecular formula) is rather high for such a small molecule. Although the IR spectrum pattern allows one to connect the N atom to two hydrogen atoms by hand we will avoid this approach and perform ab initio modeling as much as possible.

Under these conditions we can certainly expect that the number of possible generated structures will be large.

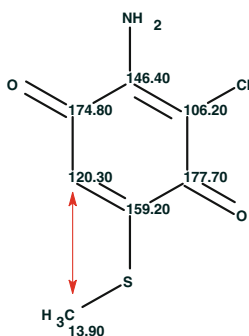
Checking the MCD for contradictions provided the message that the MCD had passed all tests so Strict Structure Generation was initiated. Results: $k = 454 \rightarrow 269 \rightarrow 177$, $t_g = 1$ s. It turned out that structure **5.36** was selected by the spectrum prediction as the best one with average deviation values of 10–11 ppm.



5.36

The result obtained can be interpreted as a hint to the presence of NSCs in the 2D NMR data (see COSY and HMBC correlations from CH_3 to $=\text{C}-\text{H}$ in Fig. 5.47). Therefore, FSG was initiated with the options $m = 1$, $a = 16$. The generation was completed with the results: $k = 37,646 \rightarrow 36,884 \rightarrow 17,807$, $t_g = 3$ min. The four top structures of the ranked file are shown in Fig. 5.49.

Figure 5.49 shows that the best ranked structure is identical to structure **5.35**, and structure **5.37** shows the chemical shift assignment as well as the single nonstandard correlation:



5.37

Consequently, the application of the CASE approach solved the problem even using sparse 2D NMR data in spite of the pessimistic prognosis of the authors [20] who claimed that the problem was unsolvable even for experienced spectroscopists.

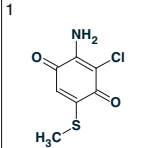
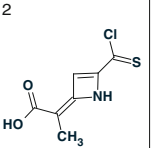
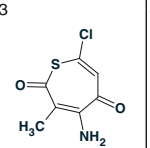
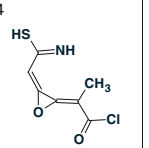
1	2	3	4
			
$d_A(^{13}\text{C}): 3.001$ $d_N(^{13}\text{C}): 4.417$ $d_I(^{13}\text{C}): 1.921$ $\text{max_}d_A(^{13}\text{C}): 7.350$	$d_A(^{13}\text{C}): 4.193$ $d_N(^{13}\text{C}): 4.131$ $d_I(^{13}\text{C}): 4.820$ $\text{max_}d_A(^{13}\text{C}): 10.120$	$d_A(^{13}\text{C}): 4.349$ $d_N(^{13}\text{C}): 4.681$ $d_I(^{13}\text{C}): 4.806$ $\text{max_}d_A(^{13}\text{C}): 9.200$	$d_A(^{13}\text{C}): 4.869$ $d_N(^{13}\text{C}): 7.332$ $d_I(^{13}\text{C}): 4.973$ $\text{max_}d_A(^{13}\text{C}): 10.930$

Fig. 5.49 Hyaladione: The four top structures of the ranked file. Structure #2 can be readily eliminated as the presence of a COOH group contradicts the IR spectrum (Fig. 5.46)

This example emphasizes once more that the estimation of the probability of solving a structural problem from NMR data is valid only if a CASE expert system is applied to this goal (see [21] and Sect. 5.33).

It is interesting to examine the generated structures that are similar to structure 5.35 and estimate the reliability of the best structure selection. A Similarity search of the best structure in the output file was performed (**Structure/Similarity Search In/Generated Molecules**), and structures characterized with a Similarity Coefficient falling in the interval 1.0–0.9 were ranked in *descending* order of Similarity Coefficient (see Fig. 5.50).

Figure 5.50 shows that all structures similar to Hyaladione can be confidently rejected by their average deviations without applying X-ray crystallography. The molecular formula of Hyaladione $\text{C}_7\text{H}_6\text{ClNO}_2\text{S}$ contains 12 skeleton atoms. Therefore, the total number of theoretically possible isomers corresponding to the



















1	2	3	4	5	6
					
$d_A(^{13}\text{C}): 15.28$ SimCoeff: 1.00	$d_A(^{13}\text{C}): 9.799$ SimCoeff: 1.00	$d_A(^{13}\text{C}): 14.31$ SimCoeff: 1.00	$d_A(^{13}\text{C}): 3.001$ SimCoeff: 1.00	$d_A(^{13}\text{C}): 6.294$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 6.981$ SimCoeff: 0.99
7	8	9	10	11	12
					
$d_A(^{13}\text{C}): 14.21$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 14.58$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 13.95$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 14.29$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 10.61$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 5.760$ SimCoeff: 0.99
13	14	15	16	17	18
					
$d_A(^{13}\text{C}): 10.13$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 5.604$ SimCoeff: 0.99	$d_A(^{13}\text{C}): 22.52$ SimCoeff: 0.92	$d_A(^{13}\text{C}): 15.38$ SimCoeff: 0.92	$d_A(^{13}\text{C}): 13.83$ SimCoeff: 0.90	$d_A(^{13}\text{C}): 22.08$ SimCoeff: 0.90

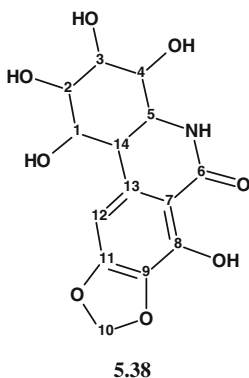
Fig. 5.50 Hyaladione: The structures characterized with Similarity Coefficient falling in the interval 1.0–0.9. The four first structures have the same value of Similarity Coefficient equal to 1. Structure # 4 is Hyaladione

molecular formula can be calculated using Structure Elucidator. The calculation (it took ~ 7 min) gave the number N of isomers equal to 465,326,452.

If n , the number of candidate structures found as a result of problem solving, is equal to 17,807, then in accordance with the Shannon formula (see Sect. 1.1.1), the amount of extracted structural information, I , is equal to $I = \log_2 N - \log_2 n = 28.8 - 14.1 = 14.7$ bits, so almost half of all structural information was extracted by application of the logical and combinatorial algorithms of the Structure Elucidator. At the same time 96.2 % of all possible isomers were rejected by the program. Therefore, the remaining half of the structural information was extracted using the methodology of ^{13}C chemical shift calculation and structural file ranking, which allowed us to select a single and correct structure. Unfortunately, an estimation of the number of isomers, N , is possible only for small molecular formulae as the number of possible isomers and the processing time grows exponentially as the number of skeletal atoms increases.

5.19 Epipancratistatin

The Amaryllidaceae plant family genera and particularly the *Narcissus* genus have a long and quite notable place in the history of traditional and Western medicine. While investigating *Narcissus* varieties from horticultural sources, Pettit et al. [22] isolated and elucidated the structure of a new natural product, 3-epipancratistatin, **5.38**, with molecular formula $\text{C}_{14}\text{H}_{15}\text{NO}_8$.



IR absorption bands observed at 3,230 and $1,670\text{ cm}^{-1}$ suggested the presence of OH and C=O groups, respectively. The 1D NMR, HSQC, HMBC, and COSY spectra (Table 5.19) presented in the article (a total of 46 COSY and HMBC correlations) were input into StrucEluc, and the MCD was created (Fig. 5.51).

MCD overview The hybridization states of the carbon atoms C(97.7), C(101.8), and C(107.4) were assigned as “not sp ” (sp^2 or sp^3) by the program because of the

Table 5.19 3-Epipancratistatin: The spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M (J)	COSY	C HMBC
C1	67.3	73.79	CH	4.27	u	2.94, 3.82, 5.11	C5, C2, C3
C2	72.6	74.1	CH	3.82	u	3.55, 4.27, 4.94	C14, C1
C3	70.4	75.44	CH	3.55	u	3.54, 3.82, 4.62	C4
C4	71.9	73.02	CH	3.54	u	3.55, 5.03, 3.37	C3, C5
C5	54	56.27	CH	3.37	u	2.94, 3.54	C3, C6, C4, C13, C14
C6	169.4	165.19	C	–	–	–	–
C7	107.4	110.79	C	–	–	–	–
C8	145.4	149.26	C	–	–	–	–
C9	131.7	131.12	C	–	–	–	–
C10	101.8	102.63	CH ₂	6.04	u	–	C9, C11
C11	152.1	151.36	C	–	–	–	–
C12	97.7	99.94	CH	6.49	u	–	C8, C7, C11, C6, C14, C9
C13	135.8	139.78	C	–	–	–	–
C14	38.3	39.7	CH	2.94	u	3.37, 4.27	C13, C5, C7, C4, C1, C8, C11
N1	150 ^a	–	NH	7.49	u	–	C4, C5, C14, C13, C8
O1	100 ^a	–	OH	5.11	u	4.27	C14, C1, C2
O2	110 ^a	–	OH	4.94	u	3.82	C3, C1, C2
O3	120 ^a	–	OH	4.62	u	3.55	C4, C2, C3
O4	130 ^a	–	OH	5.03	u	3.54	C5, C3, C4
O5	140 ^a	–	OH	13.02	u	–	C9, C6, C7, C8

^a Fictitious ¹⁵N and ¹⁷O chemical shifts

presence of each of C(*sp*³), a fragment O–C–O, and C(*sp*²) atoms is conceivable. Carbon atoms with ¹³C chemical shifts in the range between 67.3 and 72.6 ppm (all ¹H chemical shifts of attached hydrogens are higher than 3.5 ppm) were marked with the label “*ob*” by the user (atom neighbors are heteroatoms). As the presence of a carbonyl group was evident from the IR spectrum, the carbon C(169.4) was connected to an oxygen atom by a double bond on the MCD. It is interesting to note that after the application of the command “Clean” to the MCD six carbon atoms and four hydroxyl groups formed a readily recognizable part of the molecule encompassing the COSY connectivities. There is also a peculiarity specific for the carbon at C 131.7 in this particular elucidation problem: if a ¹³C signal occurs around 130 ppm in an NMR spectrum then it is usually expected that an oxygen atom

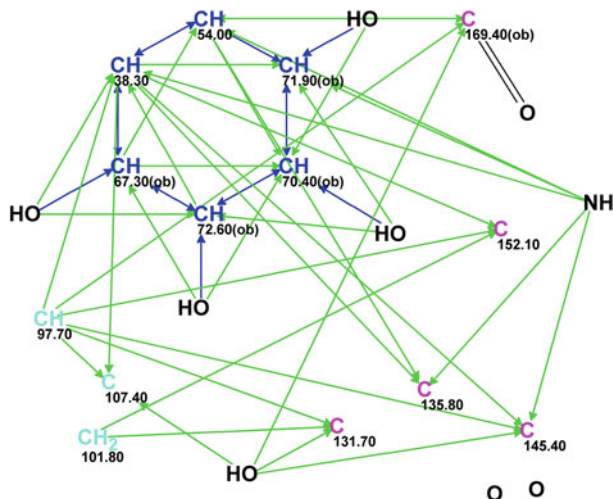
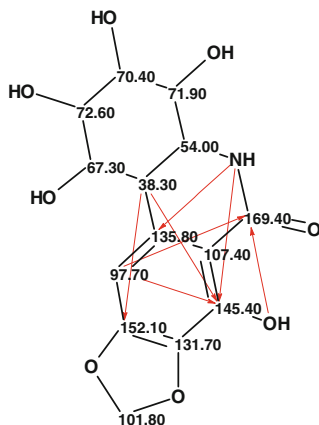


Fig. 5.51 3- Epipancratistatin: The MCD. The atom properties and double bond drawn from C 169.4 to the oxygen were set by the user

cannot be its neighbor [18]. However, in the molecule under consideration, as we will see, such an assumption would be a pitfall: a correct solution would never be obtained if the sp^2 -hybridized C 131.7 atom was marked with “*fb*” label by the user.

Checking the MCD for contradictions resulted in the following program message: “The minimum number of nonstandard connectivities is 2.” In the first program run, we purposely will ignore the existence of the visually observed spin system (six-membered cycle) and will admit the possibility of the presence of NSCs both in the COSY and HMBC data. Therefore, FSG was initiated with the following parameters $m = 2-20$, $a = 16$, “Stop when structure generated.” To reduce the output file, an option “Calculate ^{13}C chemical shifts during generation” was switched on and, consequently, all generated structures with deviation values of $d > 4$ ppm were rejected by the program. The results were: $k = 13,626 \rightarrow 1$, $t_g = 26$ min 50 s, 7 from 46 connectivities have been extended during generation, 46,800,160 from 53,524,680 possible connectivity combinations have been checked during generation. This means that the real number of NSCs in the 2D NMR data is at least 7, and almost 47 million attempts at structure generation were performed with $m = 7$. All attempts at structure generation with the number of NSCs $m < 7$ yielded empty structure files. A single resultant structure turned out to be identical to structure **5.38** ($d_A = 3.21$, $d_I = 2.1$, $d_N = 1.99$ ppm) and its assignment coincided with that suggested by the authors [22]. All seven HMBC NSCs are shown on structure **5.39** by arrows:



5.39

We see that all nonstandard connectivities detected by the program are of three C–C bond lengths and all of them encompass two central six-membered rings, which markedly hampers manual structure elucidation. In the process of structure elucidation, the authors [22] suggested the presence of an aromatic ring, an amide unit and a O–CH₂–O group from the 1D NMR and HMQC data. When StrucEluc was used, these suggestions were not necessary and a single correct structure was found almost automatically on the basis of several evident MCD edits.

For the sake of completeness we repeated the problem solving with only one additional “axiom”: the existence of nonstandard connectivities is allowed only in the HMBC data (the closed system of COSY connectivities shown on the MCD was taken into account). For this goal, the COSY check box was deselected in the **Options of Structure Generation** as shown below:



The results gave: $k = 12,329 \rightarrow 1$, $t_g = 6 \text{ min } 42 \text{ s}$, 7 from 23 correlations have been extended during generation, 225,709 from 245,157 possible connectivity combinations were used during structure generation. We conclude that both the scale of the computations and the processor time were markedly reduced when the necessity of checking the COSY connectivities was excluded.

As noted above, the ¹³C signal occurring at 131.7 ppm could be assigned on the MCD to an *sp*²-hybridized carbon atom having no oxygen in the first sphere of environment, which corresponds to the correlation tables [18]. It would be interesting to see how the program would respond if the label “*fb*” was ascribed to the carbon C 131.7. FSG was repeated under this condition ($m = 2-20$, $a = 16$, the filter was switched off to see all probable structures), and the following results were

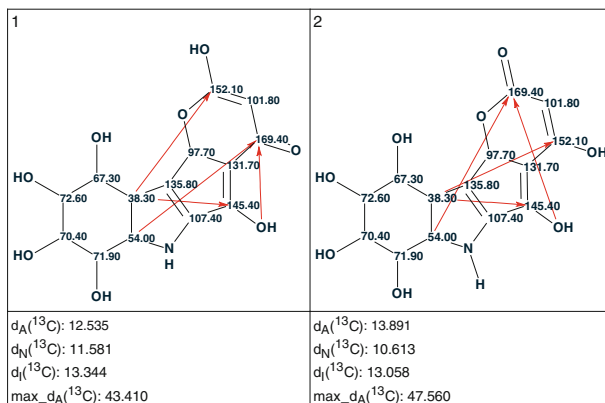


Fig. 5.52 3- Epipancrastatin: The output structural file obtained when carbon C 131.7 was supplied with the label “*fb*” (neighborhood with a heteroatom is forbidden). Nonstandard connectivities are marked with arrows. Note the non-realistic lengths of the NSCs

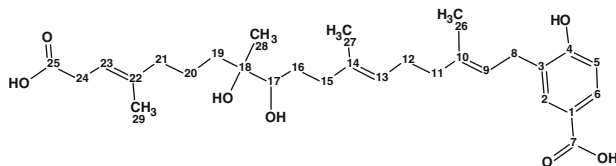
obtained: $k = 2 \rightarrow 2$, $t_g = 4$ s, 4 from 46 connectivities have been extended during generation, 110,825 from 163,185 possible connectivity combinations were used during generation. The structures obtained are presented in Fig. 5.52.

Though both structures look nice from the point of view of chemical aesthetics, they should be rejected due to the huge average and maximum deviations and the unrealistically long NSCs. The Structure Elucidator therefore allowed for the detection and rejection of an incorrect hypothesis simply based on the properties of the carbon C 131.7.

5.20 Erythrolic Acid

Meroterpenoids are natural products of mixed biosynthetic origin, containing a terpene element in combination with a carbon skeleton derived from other biosynthetic pathways, such as the shikimate or polyketide pathways. There are relatively few meroterpenoids from bacterial sources. Hu et al. [23] have found that bacteria of the genus *Erythrobacter* are prolific producers of meroterpenoids.

The authors [23] isolated and identified five meroterpenoids erythrolic acids A–E containing a hydroxybenzoic acid moiety and, in the case of **5.40**, (erythrolic acid A) a two-carbon homologated terpene side chain.



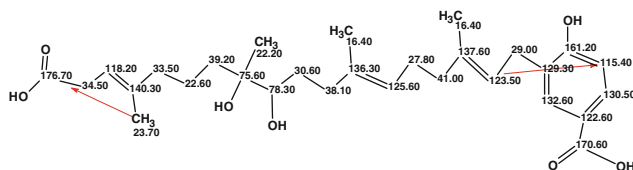
5.40

The molecular formula of Erythrolic acid **A** was determined to be $C_{29}H_{42}O_7$ (calculated for $C_{29}H_{41}O_7$, 501.2852), based on a HRESIMS $[M-H]^-$ of 501.2849, indicating nine degrees of unsaturation. Tabulated 1D NMR and HSQC data as well as selected key HMBC correlations which were graphically displayed in the article [23] are presented in Table 5.20.

The slightly edited MCD is shown in Fig. 5.53.

The following edits were made in the initial MCD. For light blue (sp^2 or sp^3) carbon atoms hybridization states were assigned as follows: 75.6 sp^3 , C 122.6 sp^2 , and C 129.3 sp^2 . The number of hydrogen atoms attached to the carbon atoms in the first sphere of each environment were set for some of the carbon atoms according to the signal multiplicities and coupling constants observed in the 1H NMR spectrum (see column M(J), Table 5.20).

Checking the MCD revealed the presence of at least two nonstandard connectivities in the HMBC data and so FSG was initiated with the options $m = 2-20$, $a = 16$, "Stop Generation when Structures Generated." ^{13}C NMR chemical shifts were calculated during structure generation, and structure filtering was used as described in Sect. 2.2.2. Results: $k = 2 \rightarrow 1$, $t_g = 6$ s, $d_A = 1.08$, $d_N = 1.30$, $d_I = 1.43$ ppm, 2 from 51 correlations have been extended during generation and 194 from 1,275 possible connectivity combinations were used during the generation process. The resulting single structure **5.41** was identical to the structure of erythrolic acid **A** and is shown below. The automatically assigned ^{13}C chemical shift assignments and nonstandard connectivities (marked by red arrows) are displayed on the structure.



5.41

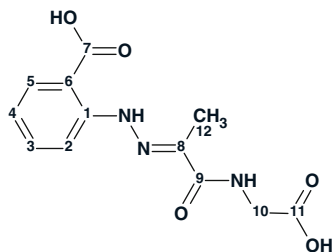
It can be seen that the number of NSCs detected as a result of MCD checking turned out to be equal to the real number of nonstandard connectivities and their lengths correspond to a $^5J_{CH}$ coupling constant, i.e., to remove contradictions in the HMBC data, the mentioned connectivities should be elongated by two chemical bonds. Logical analysis of the HMBC data allowed detection of these correlations, while the procedure of FSG determined their lengths and resulted in a single and correct structure almost instantly. The small average deviation values calculated for structure **5.41** support its correctness.

Table 5.20 Erythrolic acid A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH_n	δH	M(J)	C HMBC
C1	122.6	122.5	C	–	–	–
C2	132.6	132.5	CH	7.76	d(2.0)	C8, C4, C6, C7
C3	129.3	129.2	C	–	–	–
C4	161.2	161.1	C	–	–	–
C5	115.4	115.3	CH	6.77	d(8.4)	C4, C3, C1
C6	130.5	130.42	CH	7.69	d(8.4)	C7, C2, C4
C7	170.6	171.03	C	–	–	–
C8	29	28.9	CH ₂	3.31	u	C10, C4, C9, C2, C3
C9	123.5	123.3	CH	5.34	u	C11, C8, C5
C10	137.6	137.4	C	–	–	–
C11	41	40.9	CH ₂	2.07	t(7.3)	C12, C26, C9, C10, C13
C12	27.8	27.6	CH ₂	2.15	q(7.3)	C11, C14, C13, C10
C13	125.6	124.89	CH	5.21	t(7.3)	C15, C12, C27, C11
C14	136.3	134.02	C	–	–	–
C15	38.1	37.14	CH ₂	2.23	u	C14, C13
C15	38.1	37.14	CH ₂	1.98	u	C17, C16, C14, C13
C16	30.6	28.96	CH ₂	1.72	u	–
C16	30.6	28.96	CH ₂	1.34	u	C17
C17	78.3	78.16	CH	3.26	u	C19, C15, C18
C18	75.6	74.26	C	–	–	–
C19	39.2	38.77	CH ₂	1.51	u	C18
C19	39.2	38.77	CH ₂	1.45	u	–
C20	22.6	22.12	CH ₂	1.44	u	–
C20	22.6	22.12	CH ₂	1.53	u	–
C21	33.5	40.49	CH ₂	2.04	t(7.0)	C22, C29, C23, C19
C22	140.3	139.99	C	–	–	–
C23	118.2	116.12	CH	5.35	u	C21, C29, C25, C24
C24	34.5	33.16	CH ₂	3.01	d(7.20)	C25, C23, C22
C25	176.7	178.45	C	–	–	–
C26	16.4	16.2	CH ₃	1.71	s	C11, C9, C10
C27	16.4	15.4	CH ₃	1.61	s	C13, C15, C14
C28	22.2	22.65	CH ₃	1.07	s	C19, C18, C17
C29	23.7	16.32	CH ₃	1.73	s	C23, C22, C21, C25

5.21 Farilhydrazone

Isaria farinosa is an entomopathogenic fungus that has been used as a biocontrol agent and from which various bioactive metabolites have been reported. Ma et al. [24] have isolated and identified a series of new *N*-hydroxypyridones, particularly Farilhydrazone A (5.42).



5.42

Farylhydrazone A gave a molecular ion $[M+Na]^+$ peak at m/z 302.0742 ($\Delta -0.5$ mmu) by HRESIMS, corresponding to a molecular formula of $C_{12}H_{13}N_3O_5$ (eight degrees of unsaturation). Though the molecular formula is of a modest size, the ratio of the total number of skeletal atoms to the number of hydrogens is 1.5, which along with DBE = 9 and the presence of a chemical bond between heteroatoms suggests that the problem will be not simple to solve. In addition, there are two different kinds of heteroatoms—three nitrogens and five oxygens, all of them related to one category of skeletal atoms.

Some distinctive characteristic frequencies are observed in the IR spectrum which allows us to suggest the presence of OH/NH groups ($3,392\text{ cm}^{-1}$), carbonyl groups ($1730, 1678, 1646\text{ cm}^{-1}$), and a benzene ring ($1,591$ and $1,505\text{ cm}^{-1}$). The NMR spectroscopic data used for the structure elucidation of Farylhydrazone A are collected in Table 5.21 and the corresponding MCD is presented in Fig. 5.54.

MCD overview Two carbon atoms, C 113.2 and CH 114.2, are marked as sp^2 - or sp^3 -hybridized by the program. There are no carbon atoms for which the presence of neighboring heteroatoms could be automatically set. This urges the user to use the structural information extracted from the NMR and IR spectra at the first stage of human data analysis. Particularly, it is possible to take into account that the

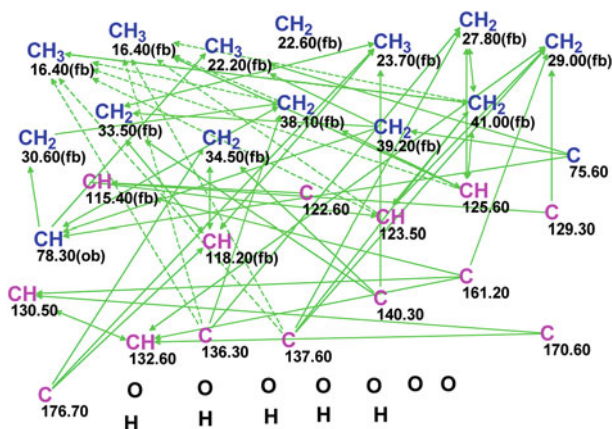


Fig. 5.53 Erythrollic acid A: The slightly edited MCD. Hybridization states were postulated for the following carbon atoms: C 75.6 sp^3 , C 122.6 sp^2 , C 129.3 sp^2

Table 5.21 Farylhydrazone A: The spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	C HMBC
C1	146.6	146.66	C	–	–	–	–
C2	114.2	114.66	CH	7.98	d(8.0)	7.53	C4, C1, C6
C3	134.5	134.56	CH	7.53	t(8.0)	7.98, 6.92	C1, C5
C4	119.7	117.67	CH	6.92	t(8.0)	7.53, 7.91	C6, C2
C5	131.7	131.65	CH	7.91	d(8.0)	6.92	C1, C7, C3
C6	113.2	114.03	C	–	–	–	–
C7	170.6	169.4	C	–	–	–	–
C8	138.3	136.87	C	–	–	–	–
C9	164.9	164.13	C	–	–	–	–
C10	41.5	41.14	CH ₂	3.86	d(6.0)	8.51	C9, C11
C11	171.9	171.48	C	–	–	–	–
C12	10.2	10.13	CH ₃	2.04	s	–	C8, C9
N1	100 ^a	–	NH	8.51	u	3.86	C9, C10

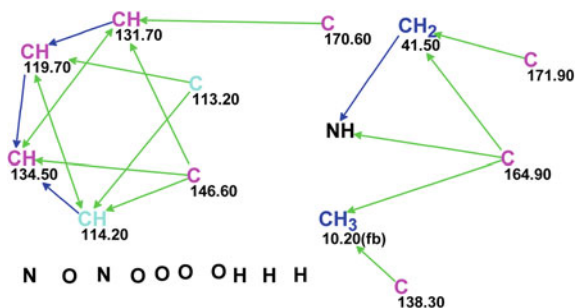
^a Fictitious ¹⁵N chemical shift

HMBC connectivities readily reveal a sequence of carbon atoms that make up a benzene ring skeleton as shown in Fig. 5.54.

Nevertheless, we will try to solve the problem in an “ab initio” fashion to demonstrate the capability of the program to solve a challenging problem without using hints obtained on the basis of data analysis by a human expert.

MCD checking with common options set, where chemical bonds between heteroatoms are *forbidden*, detected the presence of contradictions in the data set and one nonstandard correlation was revealed. FSG accompanied by ¹³C chemical shift calculation was initiated with the options usual for such a case: $m = 1-20$, $a = 16$, “Stop Generation when Structures Generated.” However, structure generation was interrupted by the user when 440,000 structures were generated during ~1 h, while no structure passed filtering. Such program behavior can be considered as a hint to change some of the initial suggestions used during the structure generation process. In this case, the presence of chemical bonds between heteroatoms was allowed and the options for structure generation were changed—the checkbox “Allow Bonds

Fig. 5.54 Farylhydrazone A:
The MCD

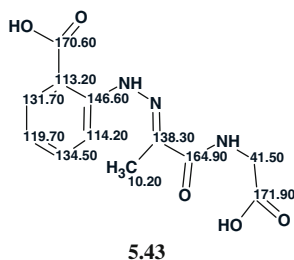


1	2	3	4
$d_A(^{13}\text{C}): 0.645$ $d_N(^{13}\text{C}): 1.476$ $d_I(^{13}\text{C}): 1.142$ $\text{max}_dA(^{13}\text{C}): 2.030$	$d_A(^{13}\text{C}): 2.822$ $d_N(^{13}\text{C}): 2.805$ $d_I(^{13}\text{C}): 2.813$ $\text{max}_dA(^{13}\text{C}): 14.420$	$d_A(^{13}\text{C}): 3.992$ $d_N(^{13}\text{C}): 3.841$ $d_I(^{13}\text{C}): 3.345$ $\text{max}_dA(^{13}\text{C}): 14.930$	$d_A(^{13}\text{C}): 3.535$ $d_N(^{13}\text{C}): 3.155$ $d_I(^{13}\text{C}): 3.358$ $\text{max}_dA(^{13}\text{C}): 9.240$

Fig. 5.55 Farylhydrazone A: The top four structures of the ranked output file

between Heteroatoms of the Same Atom Type” was selected. FSG was run again in the mode “Detect Options Automatically”. Results: $k = 27,948 \rightarrow 9 \rightarrow 9$, $t_g = 11$ s and no connectivity was extended, i.e., in reality strict structure generation was performed because the program found no contradictions in the COSY and HMBC data. The top four structures of the ranked output file are shown in Fig. 5.55.

Figure 5.55 shows that the first ranked structure is identical to the structure of farylhydrazone A and the small values of the average deviations readily confirm the confidence of the structure elucidation (see structure 5.43).



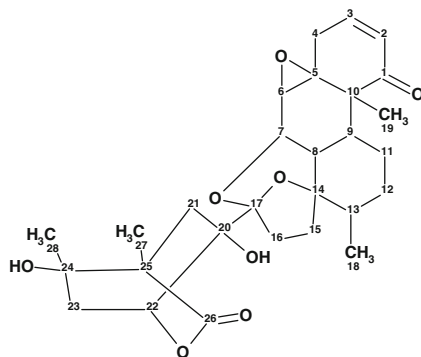
Thus, failure of the FSG process (no conceivable structure after a long generation process) hints at the need to check for the possibility of a chemical bond existing between heteroatoms. This allowed us to confidently identify the correct structure in several seconds and succeed without utilizing any additional assumptions based on the preliminary analysis of the NMR and IR data. Nevertheless, if the additional information was used the result obviously would be obtained much quicker. For instance, when three evident C=O bonds were drawn on the MCD manually (see structure 5.43) and aromatic carbons (see Fig. 5.54) were connected with connectivities of one-bond length, the failure of FSG under the condition that chemical bonds between heteroatoms are forbidden was revealed immediately: the generation process was completed with the result $k = 114,745 \rightarrow 0$, $t_g = 26$ s. The authors [24] note that the proposed structure 5.42 could not be fully corroborated by the HMBC data. Therefore, it was finally confirmed by X-ray crystallographic analysis. Employing Structure Elucidator would allow researchers to avoid the

application of this analytical procedure. The problem of analyzing farylhydrazone A exemplifies the structure elucidation of a small but difficult to elucidate organic molecule.

5.22 Physangulidine A

The genus *Physalis* (Solanaceae) is represented by almost 90 species distributed throughout the tropical and subtropical regions of the world where it has been widely used in folk medicine by developing countries. As a result of its medicinal value, there has been significant interest in evaluating the phytochemical and pharmaceutical properties of *Physalis angulata*. Previously, physanolide, withangulatin, physangulin, and physalin, among other constituents isolated from *P. angulata*, were found to show significant biological activity.

Jin et al. [25] reported the discovery of three antiproliferative withanolides with an *unusual carbon framework*, namely, physangulidines A, B, and C, isolated from *P. angulata* L. Since the date of publication [25], about 650 withanolides have been isolated from different plant sources, physangulidine A (**5.44**) is the first withanolide having a disconnection between C13 and C17. The spectroscopic data used for the structure elucidation of physangulidine A were employed for the explanation of the CASE methodology in this section.



5.44

Physangulidine A was isolated as white crystalline needles whose molecular formula was determined to be $C_{28}H_{36}O_8$ by HRMS (m/z 501.2493 $[M+H]^+$, 523.2311 $[M+Na]^+$). Simple analysis of the 1H , ^{13}C , and HSQC spectra revealed four methyls, seven methylenes, eight methines, and nine quaternary carbons. The IR spectrum (see SI to [25]) suggests the presence of hydroxyl ($3,477\text{ cm}^{-1}$), ester ($1751, 1730\text{ cm}^{-1}$), and ketone groups ($1,713\text{ cm}^{-1}$).

1D and 2D NMR data (COSY, ROESY, and HMBC) were used for the structure elucidation (see Table 5.22).

Table 5.22 Physangulidine A: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	C HMBC
C1	201.8	203.71	C	–	–	–	–
C2	129.3	128.55	CH	6.08	u	6.86	C4
C3	143.7	142.02	CH	6.86	u	6.08, 1.93	C5, C4, C1
C4	32.06	32.61	CH_2	1.93	u	6.86	C10, C1
C4	32.06	32.61	CH_2	2.95	u	–	C2, C3, C6, C5, C19
C5	62.59	64.71	C	–	–	–	–
C6	63	60.89	CH	3.14	u	4.32	C9, C4, C7
C7	67.07	71.29	CH	4.32	u	3.14, 1.45	C9, C4, C6
C8	33.97	40.08	CH	1.45	u	–	–
C9	34.11	33.91	CH	1.82	u	1.45	C1, C11
C10	47.45	54.14	C	–	–	–	–
C11	20.48	27.17	CH_2	1.35	u	–	C12
C11	20.48	27.17	CH_2	2.1	u	–	–
C12	28.53	29.92	CH_2	1.48	u	–	–
C12	28.53	29.92	CH_2	1.73	u	–	–
C13	37.34	40.89	CH	1.76	u	0.95	–
C14	90.02	82.85	C	–	–	–	–
C15	32.95	28.58	CH_2	1.85	u	–	–
C15	32.95	28.58	CH_2	1.71	u	2.17	C17, C14
C16	31.46	29.13	CH_2	2.17	u	1.71	C14, C15, C17
C17	109.7	108.59	C	–	–	–	–
C18	14.47	15.86	CH_3	0.95	d(7.0)	1.76	C13, C14, C12
C19	14.22	15.11	CH_3	1.21	S	–	C11, C9, C5, C10, C1
C20	75.06	75.44	C	–	–	–	–
C21	36.92	37.79	CH_2	1.45	u	4.32, 1.82	C24, C17, C14, C7
C21	36.92	37.79	CH_2	2.4	u	–	C26, C25, C27, C17, C20
C22	78.74	82.57	CH	4.62	u	2.41	C21, C24, C20, C26
C23	40.49	37.14	CH_2	2.09	u	–	–
C23	40.49	37.14	CH_2	2.41	u	4.62	C22, C20, C24, C28
C24	69.49	71.54	C	–	–	–	–
C25	48.46	46.24	C	–	–	–	–
C26	177	174.41	C	–	–	–	–
C27	14.04	16.49	CH_3	1.14	S	–	C25, C24, C26, C21
C28	27.04	27.54	CH_3	1.17	S	–	C23, C24, C25

The initial MCD is presented in Fig. 5.56.

MCD overview The diagram contains ambiguous connectivities which are accounted for by the overlapping ^1H signals at δ 1.45 produced by protons attached to atoms C8 and C21. There are four carbons (C 69.49–C 109.7) for which the state

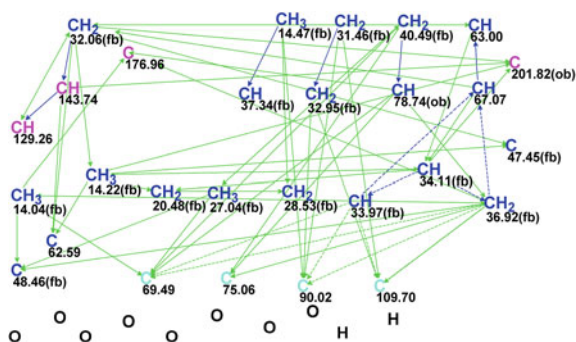


Fig. 5.56 Physangulidine A: Initial MCD

of hybridization is not defined and there is a series of supposedly oxygenated carbon atoms (C 62.59–C 75.06 and 176.96) which did not receive the label “ob” automatically. Taking into account the mentioned observations, as well as the size of the molecule (36 skeletal atoms, RDBE = 11), one can expect that the structure generation process from the initial MCD will be time-consuming. Therefore, the MCD was reasonably edited, and the modified MCD is shown in Fig. 5.57.

As the presence of ester (176.96) and ketone (201.82) groups is highly probable both from the IR and ^{13}C NMR spectra these groups were drawn onto the MCD by hand. Carbon atoms C 90.02 and 109.70, in principle, can belong either to an acetal group or to a double-bonded carbon; therefore the properties of these atoms were left as they are. The distinct multiplicities of several signals in the ^1H NMR spectrum (see Table 5.22, column M(J)) were used for setting the numbers of hydrogen atoms attached to the neighboring carbons.

As MCD checking revealed the minimum number of nonstandard connectivities to be one, FSG accompanied by ^{13}C chemical shift prediction was initiated with the parameters $m = 1\text{--}20$, $a = 16$, “Stop Generation when Structures Generated.” Results: $k = 26,073 \rightarrow 151 \rightarrow 75$, $t_g = 11$ min 30 s, 5 from 52 correlations have

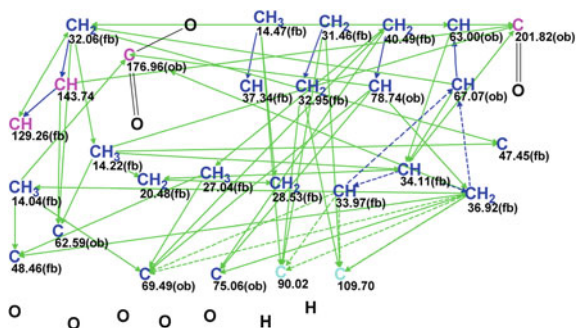
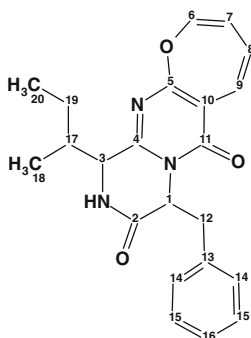


Fig. 5.57 Physangulidine A: Edited MCD

5.23 Protuboxepin A

Fungi have proven to be valuable resources for the discovery of novel secondary metabolites. Because the marine environment provides unique ecosystems and living conditions, marine fungi have been recognized as a potential source of diverse novel secondary metabolites. Lee et al. [26] have investigated the chemical constituents of the extracts obtained from cultures of the marine-derived fungus *Aspergillus* sp. SF-5044. This study led to the isolation of new natural products, particularly, an oxepin-containing diketopiperazine-type metabolite named protuboxepin A (**5.46**).



5.46

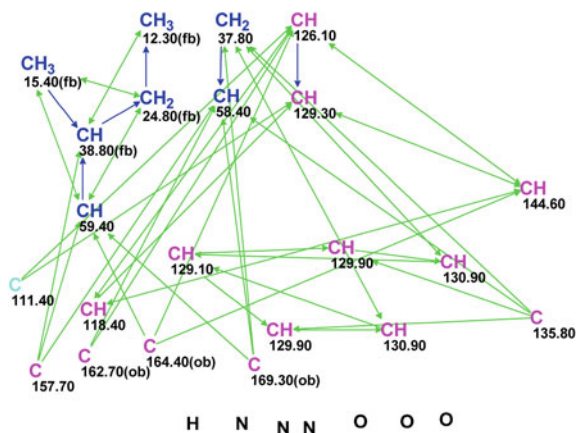
Protuboxepin A was assigned the molecular formula $C_{22}H_{23}N_3O_3$ on the basis of HRESIMS data (m/z 378.1824 $[M+H]^+$), which was fully supported by the 1H and ^{13}C NMR data. Spectroscopic data (^{13}C , 1H NMR, COSY, HMBC) used for the protuboxepin A structure elucidation are collected in Table 5.23 (COSY data were not tabulated in the article [26], so they were determined from the spectral pattern included in the Supporting Information). A slightly modified MCD is presented in Fig. 5.59.

MCD overview Only three atom property refinements were made to the MCD: a label “ob” (a sign of the presence of at least one heteroatom in the first sphere of the given atom environment) was assigned to the carbon atoms C 162.70, C 164.40, and C 169.30. The properties of the light blue carbon atom C 111.4 were left without any edits: along with $C(sp^2)$, the presence of fragments O–C–O and N–C–O is also conceivable. The numbers of hydrogen atoms attached to the carbons neighboring with some CH_n atoms were set in accordance with the 1H multiplicities shown in Table 5.23 (column M/J). The assigned hydrogen numbers can be seen when the cursor is placed on the corresponding carbon atom. The molecule contains one exchangeable proton, and if the IR spectrum of the sample diluted in CCl_4 was available, the choice between OH and NH functionalities would be possible. Unfortunately, no IR spectroscopic data were presented in the article [26], therefore an attempt to resolve the alternative was made using the so-called methodology of “generalized portrait” (see Sect. 1.3.2.6 and monograph [21]).

For this goal, a fragment search against the ACD Fragment Library using the ^{13}C NMR spectrum of the compound was carried out and resulted in the selection of

Table 5.23 Protuboxepin A: Spectroscopic NMR data

Label	δC	δC_{calc}	CH _n	δH	M(J)	COSY	C HMBC
C1	58.4	57.23	CH	5.34	u	3.35	C11, C2, C12, C13, C4
C2	169.3	169.43	C	–	–	–	–
C3	59.4	51.42	CH	2.75	u	2.13	C4, C2, C5, C17, C19, C18
C4	157.7	148.49	C	–	–	–	–
C5	164.4	158.09	C	–	–	–	–
C6	144.6	143.59	CH	6.08	d(5.5)	–	C5, C7, C8, C9
C7	118.4	120.12	CH	5.72	u	–	C6, C9
C8	129.3	124.55	CH	6.21	u	6.73	C10, C6, C7
C9	126.1	130.77	CH	6.73	d(11.4)	6.21	C11, C7, C6, C5, C10
C10	111.4	110.63	C	–	–	–	–
C11	162.7	160.7	C	–	–	–	–
C12	37.8	36.86	CH ₂	3.35	u	5.34	C14, C13, C2, C1
C13	135.8	134.77	C	–	–	–	–
C14	130.9	128.71	CH	6.93	d(7.0)	–	C12, C15
C15	129.9	128.9	CH	7.25	t(7.7)	–	C13, C16
C16	129.1	128.48	CH	7.32	t(7.3)	–	C14
C17	38.8	39.32	CH	2.13	u	1.14, 2.75, 0.84	C3, C20, C18, C19, C4
C18	15.4	14.19	CH ₃	0.84	d(7.0)	2.13	C17, C19, C3
C19	24.8	25.36	CH ₂	1.14	u	0.75, 2.13	C3, C17, C18, C20
C20	12.3	11.22	CH ₃	0.75	t(7.3)	1.14	C19, C17

Fig. 5.59 Protuboxepin A:
The slightly modified MCD

1,029 fragments. Then the following command was activated: **Structure Elucidation/Advanced/Search Functional Groups**. Functional Groups ranked in the order of the decreasing numbers of fragments containing given functional groups


























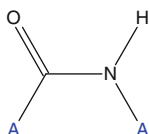
1 (ID:80)  GoodList (638 of 1029)	2 (ID:25)  GoodList (399 of 1029)	3 (ID:22)  Undefined (290 of 1029)	4 (ID:21)  Undefined (196 of 1029)	5 (ID:24)  Undefined (150 of 1029)
6 (ID:23)  Undefined (75 of 1029)	7 (ID:29)  Undefined (51 of 1029)	8 (ID:77)  Undefined (15 of 1029)	9 (ID:79)  Undefined (7 of 1029)	10 (ID:88)  Undefined (6 of 1029)
11 (ID:10)  Undefined (4 of 1029)	12 (ID:19)  Undefined (4 of 1029)	13 (ID:41)  Undefined (2 of 1029)	14 (ID:26)  Undefined (2 of 1029)	15 (ID:37)  Undefined (2 of 1029)
16 (ID:38)  Undefined (1 of 1029)	17 (ID:43)  Undefined (1 of 1029)	18 (ID:40)  Undefined (1 of 1029)	19 (ID:27)  Undefined (1 of 1029)	20 (ID:42)  BadList (0 of 1029)
41 (ID:39)  BadList (0 of 1029)	42 (ID:45)  BadList (0 of 1029)	43 (ID:46)  BadList (0 of 1029)	44 (ID:47)  BadList (0 of 1029)	45 (ID:48)  BadList (0 of 1029)

Fig. 5.60 Protuboxepin A: The top of the Functional Group Library where functional groups are ranked in the order of decreasing numbers of fragments containing given substructures. The last row shows functional groups assigned by the program to the BadList

were displayed on the screen when the command had been executed. Figure 5.60 shows the top 20 structures and a row (#41–#45), the latter containing functional groups whose presence in the molecule was rejected by the program.

The first two functional groups are the ones that are most frequently presented in the Found Fragments and are recommended by the program to be included into the User GoodList. Because both carbonyl and secondary amine groups are embedded into the secondary amide group (item 3), the fragment

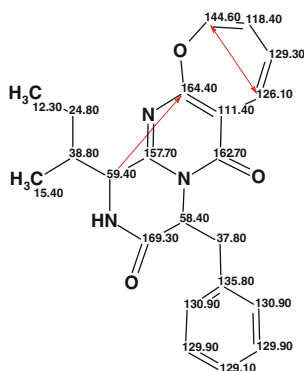


was placed into the User GoodList of Structure Elucidator. Note that only 2 of 1,029 fragments contain a tertiary alcohol functionality (item 14) and one fragment $\rightarrow\text{CH-OH}$ group (item 19), which can be considered as additional confirmation of the hypothesis about the presence of an NH (not OH!) in the molecule. Moreover,

the last row of the displayed Functional Groups can serve as a reason to exclude the presence of ester groups in the molecule.

MCD checking determined that the 2D NMR data contain at least one non-standard connectivity and therefore FSG was initiated with the following parameters: $m = 1-20$, $a = 1$, “Stop Generation when Structures Generated”, calculate ^{13}C chemical shifts during the structure generation. Result: $k = 1,928 \rightarrow 1$, $t_g = 5$ min 40 s, 2 from 35 connectivities have been extended during generation, 160 from 595 possible connectivity combinations were used, $d_A = 2.37$, $d_N = 1.82$, $d_I = 3.5$.

The single found structure (**5.47**) identical to protuboxepin A is presented below along with the ^{13}C chemical shift assignments and nonstandard connectivities (arrows).

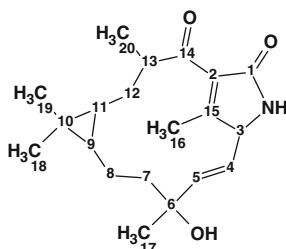


5.47

Therefore utilization of the “generalized portrait” methodology provided a single and correct solution to the problem.

5.24 Jatrophalactam

Jatropha curcas Linn. (Euphorbiaceae) is a native of tropical America but now thrives in many parts of the tropics and subtropics in Africa/Asia. It is a plant with many attributes, multiple uses, and considerable potential. The oil from *J. curcas* seeds can be used externally for the treatment of sciatica, dropsy, paralysis, rheumatism, and certain skin diseases. The seeds contain 30 % oil that can be processed to produce a high-quality biodiesel fuel, usable in a standard diesel engine. Previous investigations of the genus *Jatropha* had revealed that diterpenoid was their major secondary metabolite, and some of the diterpenoids were cytotoxic and tumor-inhibitory constituents. Wang and co-workers [27] isolated Jatrophalactam (**5.48**), a novel diterpenoid lactam possessing an *unprecedented skeleton* from the roots of *Jatropha curcas*.



5.48

Compound **5.48** was obtained as a colorless crystal. Its molecular formula, $C_{20}H_{29}NO_3$, was established on the basis of HRESIMS for the $[M]^+$ ion at m/z 331.2153, which indicated seven degrees of unsaturation. The IR spectrum showed characteristic absorptions for OH/NH ($3,392\text{ cm}^{-1}$, $3,266\text{ cm}^{-1}$) and C = O ($1,705\text{ cm}^{-1}$) functions. The ^{13}C NMR spectrum of **5.48** showed 20 carbon signals that were sorted by DEPT experiments into two carbonyls, four olefinic carbons, five methyls, three methylenes, four methines, and two quaternary carbons.

In the HSQC spectrum, a proton at δH 7.00 (brs), which had no correlations with the C-atom, was assigned by the authors [27] to be an active hydrogen of the NH group.

1D and HSQC data, as well as key COSY and HMBC correlations which were displayed graphically in article [27] are presented in Table 5.24.

The MCD which visualizes 1D and 2D NMR data collected in Table 5.24 is presented in Fig. 5.61.

No edits of the initial MCD were made. Checking the MCD for contradictions revealed the presence of at least one nonstandard connectivity in the COSY and HMBC data. The first run of the program was performed in the FSG mode, **Determine Option Automatically**. Results: $k = 74 \rightarrow 74 \rightarrow 41$, $t_g = 0.8\text{ s}$, 1 from 27 connectivities has been extended during generation, 4 from 27 possible connectivity combinations were used. The two top structures of the ranked output file are presented in Fig. 5.62.

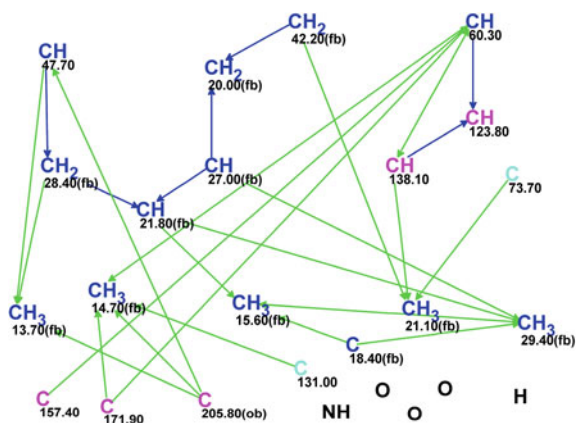
Comparison of structure #1 with the target structure shows that the best solution is correct. The red arrow corresponds to the HMBC connectivity which was automatically elongated by one chemical bond during the FSG process. However, the average and maximum deviations are quite large which, in general, could be explained by an unprecedented skeleton of the molecule. Other causes can be in an incorrect ^{13}C chemical shift assignment accounted for by the possible presence of more than one nonstandard connectivity, as well as by an NSC length of 4 or 5 skeletal bonds (parameter $a = 2$ or 3). Particularly, it is conceivable, from the common characteristic chemical shifts, that the carbon atoms with the chemical shifts of 157.4 and 171.9 should be exchanged in structure #1. To check this hypothesis, FSG was run again with the following parameters: $m = 2$, $a = 1$. Results: $k = 330 \rightarrow 330 \rightarrow 142$, $t_g = 0.7\text{ s}$, 2 from 27 connectivities have been extended during generation, 98 from 351 possible connectivity combinations were used. The two top structures of the ranked output file are presented in Fig. 5.63.

Table 5.24 Jatrophalactam: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	C HMBC
C1	171.9	171.18	C	–	–	–	–
C2	131	132.8	C	–	–	–	–
C3	60.3	62.62	CH	4.56	u	5.75	C1, C5, C15, C4
C4	123.8	124.74	CH	5.75	u	5.84, 4.56	–
C5	138.1	142.02	CH	5.84	u	5.75	C3
C6	73.7	71.03	C	–	–	–	–
C7	42.2	40.78	CH ₂	1.77	u	–	–
C7	42.2	40.78	CH ₂	1.33	u	1.58	–
C8	20	16.68	CH ₂	1.58	u	1.33, 0.53	–
C9	27	20.91	CH	0.53	u	0.43, 1.58	–
C10	18.4	18.79	C	–	–	–	–
C11	21.8	20.76	CH	0.43	u	1.42, 0.53	–
C12	28.4	27.95	CH ₂	1.42	u	0.43, 2.45	–
C12	28.4	27.95	CH ₂	1.48	u	–	–
C13	47.7	45.3	CH	2.45	u	1.42	C14
C14	205.8	201.39	C	–	–	–	–
C15	157.4	163.44	C	–	–	–	–
C16	14.7	14.16	CH ₃	2.01	u	–	C1, C14, C3, C2
C17	21.1	28.82	CH ₃	1.28	u	–	C6, C5, C7
C18	29.4	23.04	CH ₃	0.98	u	–	C9, C19, C10, C11
C19	15.6	9.64	CH ₃	0.89	u	–	C11, C18, C10
C20	13.7	17.07	CH ₃	1.26	u	–	C13, C14, C12
N1	100 ^a	–	NH	7.00	u	–	–

^a Fictitious ¹⁵N chemical shift

Fig. 5.61 Jatrophalactam:
Molecular connectivity
diagram



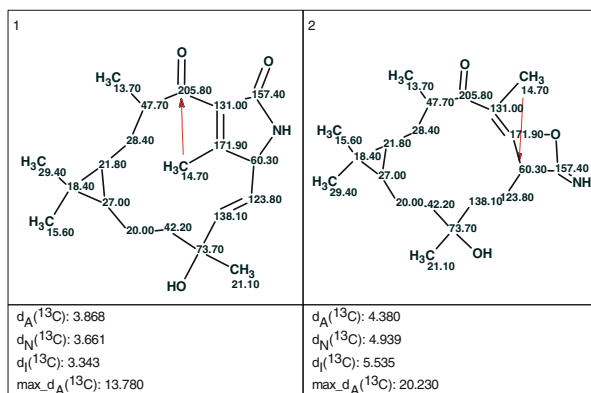


Fig. 5.62 Jatrophalactam: Two top structures of the ranked output file. One HMBC connectivity was extended during the structure generation process. The arrow indicates a nonstandard connectivity

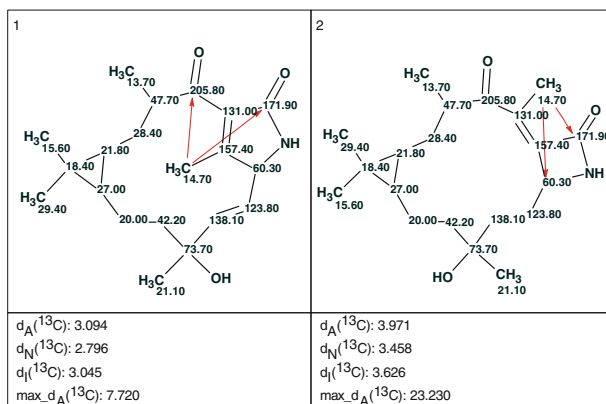


Fig. 5.63 Jatrophalactam: The two top structures of the ranked output file. Two connectivities were extended during structure generation ($m = 2$). The arrows indicate nonstandard HMBC connectivities

We see that in this case the best structure is the same, but the deviations are smaller and indeed carbon atoms 157.4 and 171.9 were automatically permuted. A question arises whether the signals corresponding to the nonstandard connectivities in the HMBC spectrum could be detected from the very beginning. The answer is obtained by inspecting the HMBC spectrum (Fig. 5.64) adopted from the Supporting Information to the article [27].

Figure 5.64 shows that the peak intensities corresponding to $^4J_{\text{CH}}$ correlations are the same as were measured for standard correlations $^{2-3}J_{\text{CH}}$.

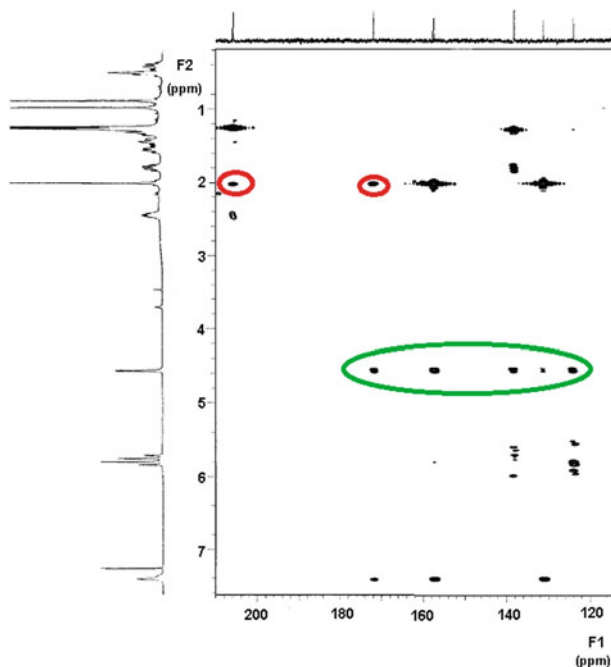


Fig. 5.64 Jatrophalactam: An expansion of part of the HMBC spectrum. $^4J_{\text{CH}}$ correlations from 2.01 to 171.9 and 205.8 are shown by *red circles*. The standard length correlations from 4.56 to C1, C4, C5, and C15 of the same intensity are highlighted by the *green ellipse* for comparison

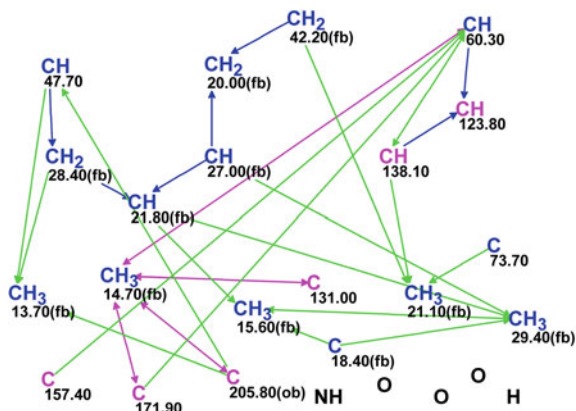
Therefore the mechanism of FSG allowed us to instantaneously and automatically reveal the contradictory connectivities, elongate them, and find the correct structure.

For completeness, we repeated checking of the MCD for contradictions with the option **Automatically Resolve Contradictions**. The program distinguished carbon atom C 14.70 as a potent carrier of connectivity set among which at least one NSC may present. The automatically modified MCD is displayed in Fig. 5.65.

Figure 5.65 shows that both $^4J_{\text{CH}}$ correlations from 2.01 to 171.9 and 205.8 were automatically elongated by one chemical bond, and consequently we can expect that the correct structure with the correct ^{13}C and ^1H chemical shift assignments will be generated. As expected, the edited MCD does not contain contradictions now, so *Strict Structure Generation* was initiated. Results: $k = 147 \rightarrow 147 \rightarrow 54$, $t_g = 0.2$ s. The top structures and their chemical shift assignments fully coincided with those presented in Fig. 5.63.

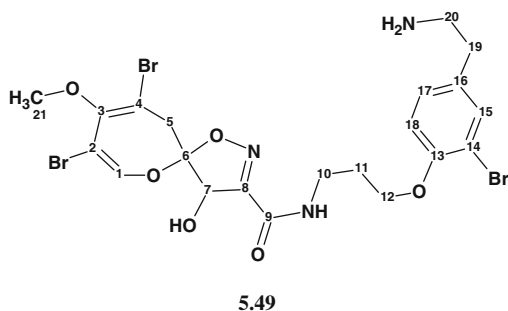
The considered example demonstrates two different methods of resolving the problem of contradictions in the 2D NMR data. As was mentioned previously (see Sect. 2.3), FSG is a more general approach, but there are such 2D NMR data sets for which the program manages to find and resolve contradictions at the MCD checking stage.

Fig. 5.65 Jatrophalactam: Automatically modified MCD. Connectivities elongated by one chemical bond were marked by the program with a *violet color*



5.25 Psammamplysin I

Screening of extracts derived from twilight or disphotic zone organisms, those living between 50 and 1,000 m depth, was undertaken by Right and co-workers [28] to test the theory that those organisms from relatively extreme environments would be at least as chemically productive as, if not more than, their counterparts found in more accessible regions of the oceans. On the basis of these screening results, two of 15 sponges were chosen at random for further chemical investigation. Each sponge extract was retested to confirm their bioactivity profiles and subsequently chemically screened for the presence of highly toxic compounds. During the course of this research, the team isolated five bromotyramine derivatives from the organic extract of the *Suberea* sp. sample, two of which were new compounds. The structure of psammamplysin I, compound **5.49**, characterized by the molecular formula $C_{21}H_{24}Br_3N_3O_6$, was elucidated by NMR and mass spectrometry.



The spectroscopic ^{13}C , 1H , HSQC, HMBC, and COSY data used for the structure elucidation are presented in Table 5.25. The authors postulated that “it was evident the molecule contained 12 sp^2 -hybridized carbons ... and it was also clear

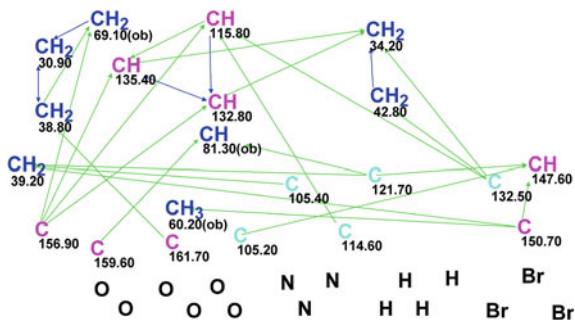
Table 5.25 Psammaplysin I: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	C HMBC
C1	147.6	146.1	CH	7.17	u	–	C2, C6, C3
C2	105.2	104.9	C	–	–	–	–
C3	150.7	149.3	C	–	–	–	–
C4	105.4	104	C	–	–	–	–
C5	39.2	40.89	CH_2	3.42	u	–	–
C5	39.2	40.89	CH_2	3.09	u	–	C6, C3, C4
C6	121.7	121.5	C	–	–	–	–
C7	81.3	79.9	CH	5.01	u	–	C8, C6
C8	159.6	157.73	C	–	–	–	–
C9	161.7	152.32	C	–	–	–	–
C10	38.8	37.55	CH_2	3.59	u	2.13	C11, C9
C11	30.9	30.05	CH_2	2.13	u	3.59, 4.14	C10, C12
C12	69.1	68.23	CH_2	4.14	u	2.13	C13, C10, C11
C13	156.9	151.71	C	–	–	–	–
C14	114.6	112.21	C	–	–	–	–
C15	135.4	134.98	CH	7.52	u	7.24	C13, C18
C16	132.5	133.97	C	–	–	–	–
C17	132.8	130.08	CH	7.24	u	7.52, 7.05	C18, C15, C13
C18	115.8	114.22	CH	7.05	u	7.24	C13, C17, C16, C14
C19	34.2	33.6	CH_2	2.91	u	3.17	C17, C20, C16, C15
C20	42.8	41.88	CH_2	3.17	u	2.91	C19, C16
C21	60.2	59.23	CH_3	3.68	u	–	C3

that there were *nine* sp^3 -hybridized carbons, indicating one methyl, six methylenes, one methine, and one quaternary carbon.” The MCD mapping the spectroscopic data is shown in Fig. 5.66.

MCD overview Figure 5.66 clearly demonstrates that there are eight (not nine) definite sp^3 -hybridized carbons, while the ninth sp^3 -hybridized carbon postulated by the authors could be identified as one of the four light blue-colored atoms

Fig. 5.66 Psammaplysin I:
Initial MCD



(C 105.4–C 121.7). Hybridization of C 132.5 can be assigned to sp^2 . We suggest that the authors' postulation would be reasonable only if it was derived on the basis of comparing the experimental ^{13}C NMR spectra with those of relative compounds or posteriori (from spectrum-structure correlations, atom C 121.7 would be assigned as sp^2 -hybridized rather than as sp^3). Therefore, if the problem is solved as "ab initio," there are no grounds for any editing of the set of light blue atoms. On the basis of the chemical composition of the unknown, it is easy to predict that the structure generation will be too time-consuming. Indeed, the molecule contains six oxygens, three nitrogens, three bromines, and four exchangeable hydrogen atoms, which means that the number of possible atomic combinations will be huge. In addition, the presence of contradictions in the 2D NMR data was detected by MCD checking. The expectation was confirmed by attempts to carry out FSG: it was stopped by the user in an hour when more than a million structures were generated. It was suggested that a fragment search against the Fragment Library using the experimental ^{13}C NMR spectral data can be helpful.

The command **Structure Elucidation/Search Fragments by CNMR spectrum** was activated with the option **Use Filter during Search**. As a result 4,887 Found Fragments were selected (they can be displayed by the command contained within the main menu **View/Structure Lists/Found Fragments**). The fragments are ranked in descending order of the number of carbon atoms in the chemical composition of a fragment, so the largest fragments are ranked first. Four top ranked fragments are shown in Fig. 5.67.

In the **Found Fragments** window, experimental ^{13}C chemical shifts are graphically displayed below the structure of a Found Fragment in comparison with the chemical shifts of a fragment subspectrum.

The first ranked fragment and its ^{13}C subspectrum in comparison with the experimental ^{13}C NMR spectrum are presented in Fig. 5.68.

Figure 5.68 shows that a subspectrum of the fragment (upper) is successfully "projected" onto the experimental spectrum (lower). It is not an accidental coincidence in this case because fragment #1 is a part of structure **5.49** (it is noteworthy that structure **5.49** is deliberately not present in the ACD/CNMR database). Therefore, an attempt to create the MCD(s) from this fragment can be undertaken. With fragment #1 being displayed in the window, the command **Structure/Create from Current Structure/MCDs for CSB Generator** was initiated. Most options associated with the MCD creation were set automatically. Figure 5.69 shows the right part of the

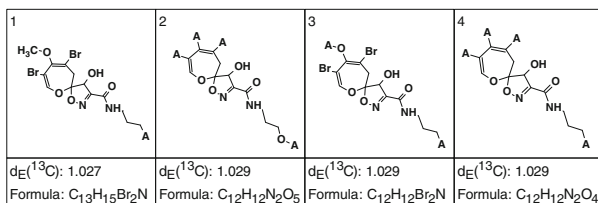


Fig. 5.67 Psammalyisin I: Four top ranked fragments

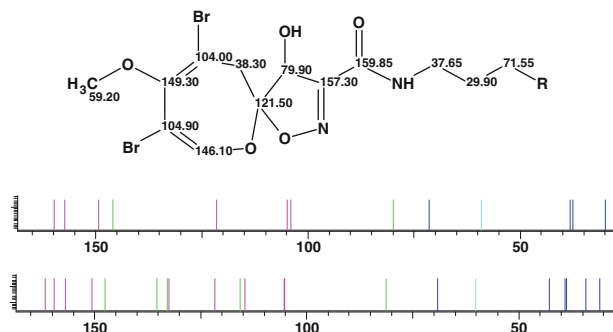


Fig. 5.68 Psammaplysin I: Fragment #1 and its ^{13}C sub-spectrum (*upper*) in comparison with the experimental ^{13}C NMR spectrum (*lower*)

Fig. 5.69 Psammaplysin I:
Right part of the window
Create MCD Options

Restrictions	
<input type="checkbox"/> Interrupt process after	10 Minutes
Maximum Number of Stored MCDs (for each MF)	1000
Assignment Options	
Search (User)	
<input type="checkbox"/> Determine Automatically	<input type="checkbox"/> Use Group Assignment
Minimum CNMR Chemical Shift Tolerance (ppm)	3.5
Tolerance for "First Sphere" Atoms (ppm)	12
Tolerance for "Second Sphere" Atoms (ppm)	6
Check (User)	
<input type="checkbox"/> Determine Automatically	
Maximum Allowed Number of Extended Correlations	2
Allow the Correlation Length to be Extended by up to	1 Bonds
Combine (User)	
<input type="checkbox"/> Determine Automatically	
Maximum Number of Fragments Included in MCD	1
Minimum Number of Fragments Included in MCD	1
Minimum Percent of Replaced MCD's Atoms	25
<input type="checkbox"/> Check MCDs by Generator	Checking Options...
<input type="button" value="OK"/> <input type="button" value="Cancel"/> <input type="button" value="Help"/>	

Create MCD Options window where settings used for MCD creation are displayed. Here the maximum and minimum of fragments included in the MCD were set by the user, while the minimum CNMR chemical shift tolerance 3.5 ppm was found by the trial method. As a result, two MCDs were created (Figs. 5.70 and 5.71).

Comparison of the two MCDs shows that the difference between them is the exchanging of the carbons with chemical shifts 105.2 and 105.4 ppm.

Fig. 5.70 Psammaplysin I:
MCD #1 created from the
Found Fragment #1

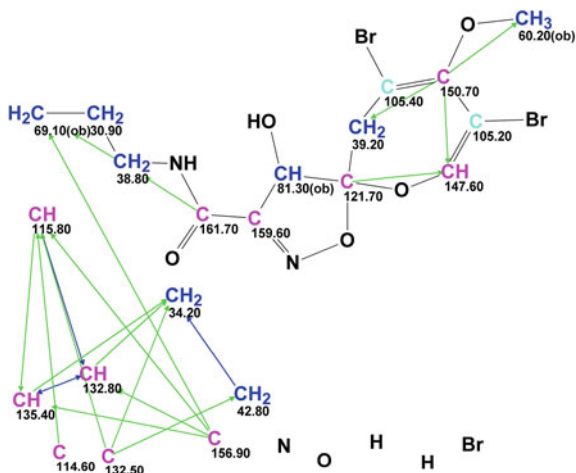
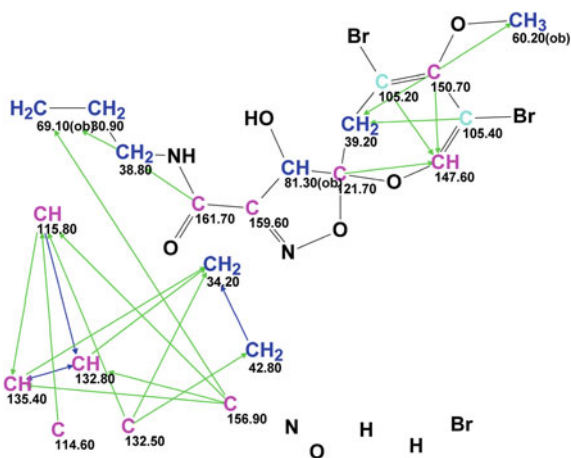


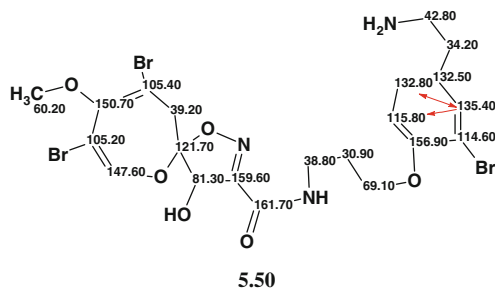
Fig. 5.71 Psammaplysin I:
MCD #2 created from the
Found Fragment #1



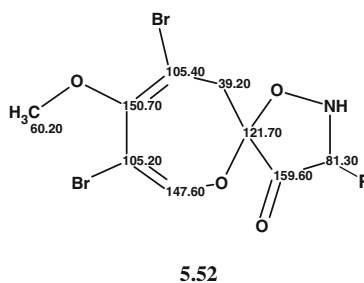
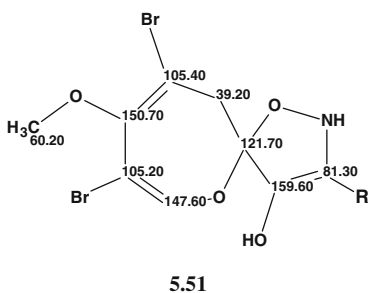
FSG was initiated using the command **Structure Elucidation/Run CSB Generator** which automatically performs the structure generation first from the MCD #1 and then from the MCD #2. Both files of generated structures are combined. For FSG, the options used are those shown in Fig. 5.72.

FSG was completed with the following results: $k = 27 \rightarrow 24 \rightarrow 24$, $t_g = 1$ s. The three top structures of the ranked output file are presented in Fig. 5.73.

It is easy to see that structure #1 coincides with the structure of psammaplysin I determined by Wright and co-workers. The chemical shift assignment, as well as two nonstandard connectivities (135.4–132.8, COSY, and 135.4–115.8, HMBC), are shown in structure 5.50:



Note that the elucidated structure may have two additional tautomeric forms partly presented by the substructures **5.51** and **5.52**:



The ^{13}C chemical shift prediction for the full tautomeric structures confirmed structure **5.50**: it turned out that average deviation values calculated for two additional tautomeric forms were 5–5.5 ppm, while the maximum deviations reached 45 ppm.

The found solution was based on the fact that fragment #1 was successfully selected by visual comparison of the graphical representation of its ^{13}C sub-spectrum with the experimental spectrum of the unknown. This approach can be useful when a large fragment of the molecule which is related to the unknown is present in the Fragment Library of Structure Elucidator.

Fig. 5.72 Psammaplysin I:
Options of FSG

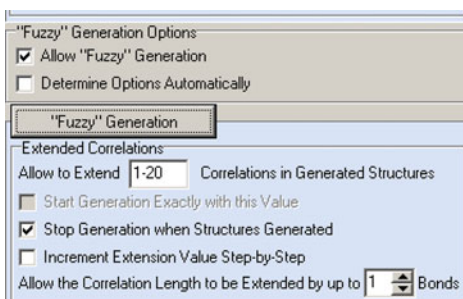
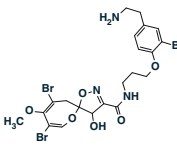
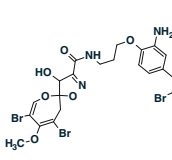
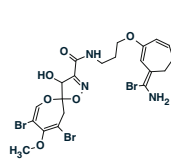
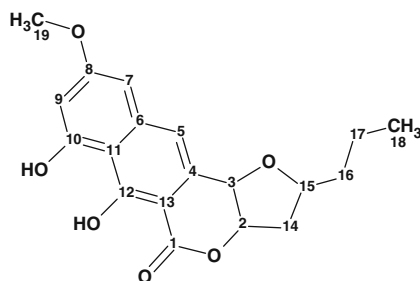


Fig. 5.73 Psammamplysin I:
Three top structures of the
ranked output file

		
$d_A(^{13}\text{C}): 1.827$ $d_N(^{13}\text{C}): 2.587$ $d_I(^{13}\text{C}): 2.447$ $\text{max}_dA(^{13}\text{C}): 9.380$	$d_A(^{13}\text{C}): 4.879$ $d_N(^{13}\text{C}): 5.648$ $d_I(^{13}\text{C}): 5.676$ $\text{max}_dA(^{13}\text{C}): 18.710$	$d_A(^{13}\text{C}): 5.045$ $d_N(^{13}\text{C}): 5.660$ $d_I(^{13}\text{C}): 5.670$ $\text{max}_dA(^{13}\text{C}): 19.270$

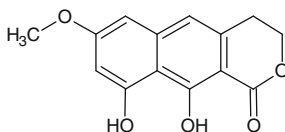
5.26 Lasionectrin

El Aouad et al. [29] isolated a new naphthopyrone (**5.53**) with antiplasmodial properties from fermentation broths of a *Lasionectria* species. This compound was named lasionectrin. The article is the first account of isolating a natural product from fungi of this genus.



5.53

A pseudomolecular ion at m/z 345.1335 by (+)-ESI-TOFMS and the presence of 19 signals in the ^{13}C NMR spectrum provided a molecular formula of $\text{C}_{19}\text{H}_{20}\text{O}_6$ for compound **5.53**. The UV spectrum displayed absorptions at 216, 260, 306, and 363 nm characteristic of a 3,4-dihydro-9,10-dihydroxy-7-methoxynaphtho[2,3-c]pyran-1-one moiety **5.54** (naphthopyrone)



5.54

Table 5.26 Lasionectrin: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	COSY	HMBC
C1	171.2	167.97	C	–	–	–	–
C2	84.2	78.04	CH	5.24	u	2.58, 4.93	C15, C3
C3	75	72.29	CH	4.93	d(1.6)	5.24	C12 , C13, C5, C4, C2
C4	131.9	137.21	C	–	–	–	–
C5	120.7	115.69	CH	7.23	s	–	C1 , C3, C12 , C11, C7, C6, C13
C6	142.1	141.41	C	–	–	–	–
C7	101.1	98.65	CH	6.8	d(2.1)	6.51	C5, C8, C11, C9
C8	164.4	160.31	C	–	–	–	–
C9	103.5	99.17	CH	6.51	d(2.1)	6.8	C8, C7, C10, C11
C10	159.5	159.21	C	–	–	–	–
C11	110.3	105.04	C	–	–	–	–
C12	163.6	163.57	C	–	–	–	–
C13	98.5	101.85	C	–	–	–	–
C14	40.8	38.63	CH ₂	2.12	u	–	C15, C16
C14	40.8	38.63	CH ₂	2.58	u	5.24, 4.33	C4 , C3, C2, C16
C15	80.4	79.05	CH	4.33	u	2.58, 1.70	C17, C14, C3
C16	39.4	37.09	CH ₂	1.57	u	–	–
C16	39.4	37.09	CH ₂	1.7	u	4.33, 1.49	C15, C14, C18, C17
C17	20.3	18.27	CH ₂	1.49	u	1.7	C15, C16, C18
C17	20.3	18.27	CH ₂	1.39	u	0.97	–
C18	14.4	14.04	CH ₃	0.97	t(7.3)	1.39	C16, C17
C19	56	55.31	CH ₃	3.89	s	–	C8

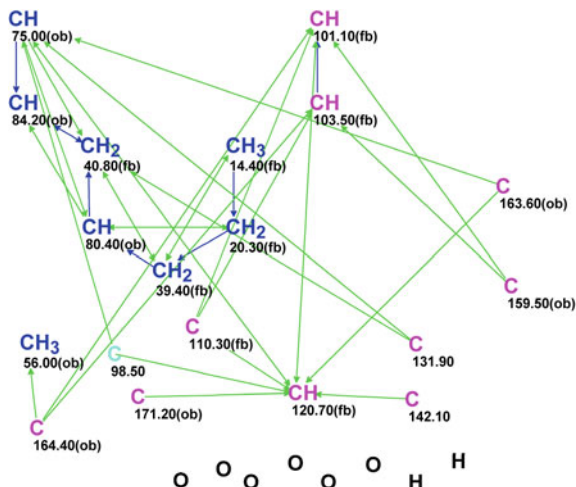
and are similar to those observed in structurally related compounds. The authors [29] used this substructure as a starting point to elucidate the full structure of the new compound by the application of 1D and 2D NMR data.

We will treat the compound as a real “unknown” and will try to solve the problem “ab initio”. 1D NMR spectra together with the 42 HMBC and 8 COSY correlations derived by the authors [29] from the 2D NMR data were input into Structure Elucidator (Table 5.26).

Using these data, an MCD was created. Slightly edited MCD is shown in Fig. 5.74.

MCD overview The displayed atom properties were assigned by the program and partly by the user considering the molecular composition (particularly the absence of nitrogen atoms in the molecular formula) and the characteristic chemical shifts [18]. The hybridization of C(98.5) was set by the program “*sp² or sp³*” (atom C is marked with a light blue color) to take into account the possibility of a carbon double bond C=C or O–C–O fragment in the structure.

Fig. 5.74 Lasionectrin:
Slightly edited MCD



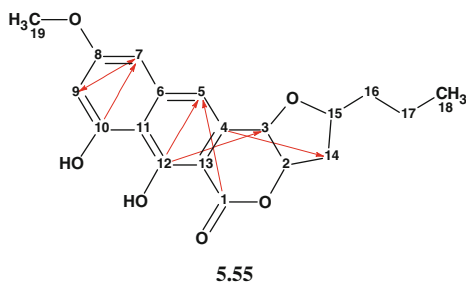
¹H multiplicities were input into the program in accordance with the coupling constant values presented in Table 5.26, column M(J). Table 5.26 shows that CH₃ (C18) produces a triplet with $J = 7.3$ Hz at 0.97 ppm and hence the number of hydrogen atoms in the first sphere of atom C18 environment is expected to be 2. This number is typed in the field corresponding to the option “Number of hydrogens on neighbor atoms” (Fig. 5.75).

Fig. 5.75 Lasionectrin:
The Structure Elucidator
window intended for setting
and editing atom properties

Edit Properties of Atom # 1	
Number of Current Atom = 1	
<input style="float: left;" type="button" value=" << Previous (27) "/> <input style="float: right;" type="button" value=" Next (2) >> "/>	
Assigned Shift(s)	
Atom's NMR Shift (nucleus 13C, shifts from -300 to 300 ppm)	
Experimental	14.4 ± 0 <input type="button" value="Clear"/>
Calculated	n/a ± n/a
Atom's QM calculated NMR Shift (ppm)	
QM (GIAO/DFT)	± <input type="button" value="Clear"/>
Attached Hydrogen's NMR Shift(s) (-2 - 20 ppm)	
Experimental	0.97 ± 0 <input type="button" value="Clear"/>
Calculated	n/a ± n/a
Atom Properties	
Connection with Heteroatoms	forbidden
Number of Hydrogens on Neighbor Atoms	2
Hybridization State	sp3
Charge	0
Valency	not defined
<input type="checkbox"/> Allow Non-default Valences	
<input type="button" value="OK"/> <input type="button" value="Cancel"/> <input type="button" value="Help"/>	

As stated earlier, the procedure of setting n values is very delicate and should be performed with great caution. As a matter of fact, when assigning an n number to some carbon atom, we introduce a new “axiom” and if the axiom turns out to be false then we will never obtain a valid solution. In order to make a decision, the values of the J_{HH} coupling constants are usually analyzed. If $J_{\text{HH}} < 2.0\text{--}2.5$ Hz then it is usually assumed that a signal is split due to a long-range coupling and the corresponding n value is equal to zero. Experience shows that the safest approach is to use only clear and obvious multiplets (for instance, singlets, doublets, and triplets of methyl groups). With the data presented in Table 5.26, the following $n(\text{H})$ values were introduced: $n(\text{H}) = 0$ for the atoms C3, C5, C7, C9, and C19, while $n(\text{H}) = 2$ for C18. We should remember that here we are modeling a situation where we know nothing about the real structure of the molecule under investigation and we cannot guess that in structure 5.53 the coupling constant $J_{\text{HH}} = 1.6$ Hz (see Table 5.26, the line related to C3) corresponds to the $^3J_{\text{HH}}$ interaction of vicinal protons.

Checking the MCD for contradictions, the program detected the presence of NSCs. Since both the structure and shift assignment were known from the work [18] the checking of the “proposed” structure 5.53 using the 2D NMR data allowed us to establish 6 NSCs ($^nJ_{\text{CH}}$, $n > 3$) in the HMBC spectrum and one NSC ($^nJ_{\text{HH}}$, $n > 3$) in the COSY spectrum. The NSCs are marked by underlined atom labels in Table 5.26 and drawn as arrows in structure 5.55, where the connectivity C7–C9 belongs to the COSY spectrum.



FSG was initiated with the following parameters:

$m = 0\text{--}20$, $a = 16$, “Stop generation when structure generated”, “Calculate carbon spectra during generation,” “Reject structures with $d > 4$ ppm and $d(\text{max}) > 20$ ppm”.

The following FSG runs were performed:

Run 1 Generated structures were stored only at $m_g = 5$. Results: $k = 74,223 \rightarrow 4$, $t_g = 4$ min 12 s. The three first structures are shown in Fig. 5.76.

Large deviations and “exotic” structures are generally observed for an incorrect solution. Therefore, it is necessary to repeat structure generation with $m = 6$.

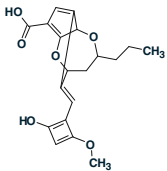
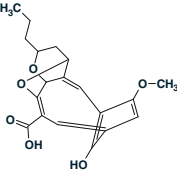
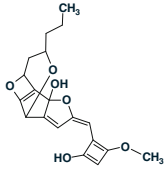
		
$d_A(^{13}\text{C}): 4.336$ $d_N(^{13}\text{C}): 4.805$ $d_I(^{13}\text{C}): 3.908$ $\text{max_}d_A(^{13}\text{C}): 9.770$	$d_A(^{13}\text{C}): 5.082$ $d_N(^{13}\text{C}): 5.816$ $d_I(^{13}\text{C}): 3.991$ $\text{max_}d_A(^{13}\text{C}): 21.550$	$d_A(^{13}\text{C}): 5.251$ $d_N(^{13}\text{C}): 4.899$ $d_I(^{13}\text{C}): 3.639$ $\text{max_}d_A(^{13}\text{C}): 16.130$

Fig. 5.76 Lasioneitrin: The three first resultant structures obtained from the first program run

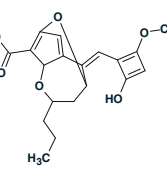
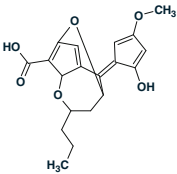
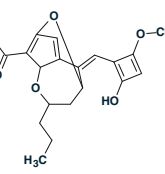
		
$d_A(^{13}\text{C}): 3.996$ $d_N(^{13}\text{C}): 4.124$ $d_I(^{13}\text{C}): 3.759$ $\text{max_}d_A(^{13}\text{C}): 13.710$	$d_A(^{13}\text{C}): 4.018$ $d_N(^{13}\text{C}): 4.519$ $d_I(^{13}\text{C}): 3.739$ $\text{max_}d_A(^{13}\text{C}): 11.780$	$d_A(^{13}\text{C}): 4.153$ $d_N(^{13}\text{C}): 4.183$ $d_I(^{13}\text{C}): 3.524$ $\text{max_}d_A(^{13}\text{C}): 13.710$

Fig. 5.77 Lasioneitrin: Top structures of the file obtained from the second program run

Run 2 The generation option $m = 6$ was set. Results: $k = 1,783,619 \rightarrow 158 \rightarrow 94$, $t_g = 1 \text{ h } 56 \text{ min}$. The correct structure was not generated. The three “best” structures of the ranked file are shown in Fig. 5.77.

We see again that the solution found is not valid, which should be expected because we erroneously set the number of hydrogen atoms equal to zero in the first sphere of carbon C3.

The computational experiments performed so far have shown that the initial data used contained a wrong assumption (a false “axiom”). Looking for the false axiom is usually initiated with the removal of user constraints that can lead to the loss of the correct structure. In the current situation, the most suspicious is the constraint $n(\text{H}) = 0$ introduced for the sp^3 carbon C(75.00) because other similar constraints (n -values) are related to methyl groups and to protons which are supposed to be aromatic. The $n(\text{H}) = 0$ constraint was removed on the MCD at C(75.00) and the FSG was repeated under the same conditions.

Run 1a Two “exotic” molecules were stored at $m_g = 4$ ($t_g = 1 \text{ min } 23 \text{ s}$), both with large calculated deviations.

Run 2a At $m = 5$ the result was: $k = 337,637 \rightarrow 47 \rightarrow 28$, $t_g = 51 \text{ min}$ and all structures (rather exotic) and having large deviations (4–5.5 ppm) were again declined.

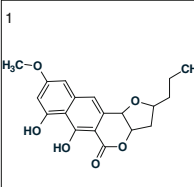
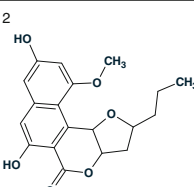
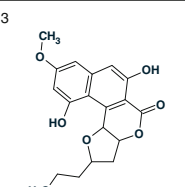
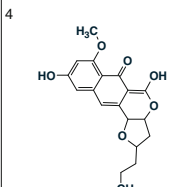
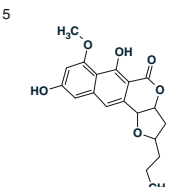
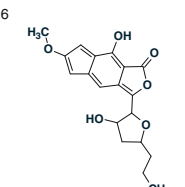
		
$d_A(^{13}\text{C}): 2.727$ $d_N(^{13}\text{C}): 2.690$ $d_I(^{13}\text{C}): 3.187$ $\text{max}_dA(^{13}\text{C}): 6.160$	$d_A(^{13}\text{C}): 3.042$ $d_N(^{13}\text{C}): 3.175$ $d_I(^{13}\text{C}): 3.400$ $\text{max}_dA(^{13}\text{C}): 10.630$	$d_A(^{13}\text{C}): 3.069$ $d_N(^{13}\text{C}): 3.085$ $d_I(^{13}\text{C}): 3.106$ $\text{max}_dA(^{13}\text{C}): 10.630$
		
$d_A(^{13}\text{C}): 3.179$ $d_N(^{13}\text{C}): 2.586$ $d_I(^{13}\text{C}): 3.887$ $\text{max}_dA(^{13}\text{C}): 13.570$	$d_A(^{13}\text{C}): 3.306$ $d_N(^{13}\text{C}): 2.954$ $d_I(^{13}\text{C}): 3.372$ $\text{max}_dA(^{13}\text{C}): 7.450$	$d_A(^{13}\text{C}): 3.354$ $d_N(^{13}\text{C}): 3.756$ $d_I(^{13}\text{C}): 4.007$ $\text{max}_dA(^{13}\text{C}): 14.590$

Fig. 5.78 Lasionelectrin: The top structures obtained from the program Run 3a

Run 3a Generation was performed with $m = 6$. Results: $k = 6,687,699 \rightarrow 822 \rightarrow 385$ and $t_g = 10$ h 51 min.

The top six structures of the file ranked by the deviation d_A (calculated using the HOSE code approach) are shown in Fig. 5.78.

The best structure #1 coincides with structure 5.53 and the average deviations are acceptable. Hence the valid solution was found in the presence of six nonstandard 2D NMR correlations of unknown length. The relatively long time for structure generation is accounted for not only by the presence of six NSCs, but also by a specific peculiarity of structure 5.53: it contains a sequence of six skeletal atoms which do not have attached hydrogen atoms (a “silent fragment”, see Sect. 1.2.2).

On reason NSCs can appear in 2D NMR data is because peak intensities corresponding to correlations of different lengths can be of the same volume. To understand the reasons for NSCs showing up in the 2D NMR data for structure 5.55 we inspected the HMBC and COSY spectra presented in the supporting information related to the article [29]. Inspection showed that the intensity of the $^4J_{\text{HH}}$ COSY peak (6.8–6.51) is of an order similar to the other “standard” COSY peaks. However, inspection of the HMBC spectrum showed that three NSC peaks (6.8–159.5, 4.93–163.5, and 2.58–131.9) were actually absent from the spectra.¹

¹ The authors informed us that the correlations which are invisible in Fig. 5.7 were actually observed, but with very low peak intensities, which frequently occurs for NSCs.

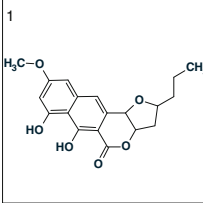
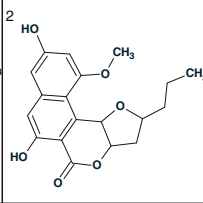
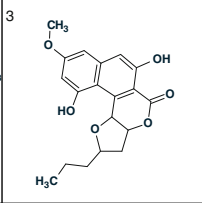
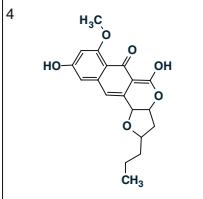
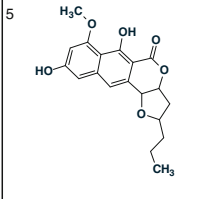
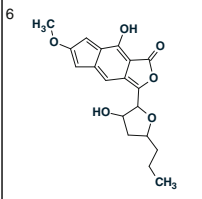
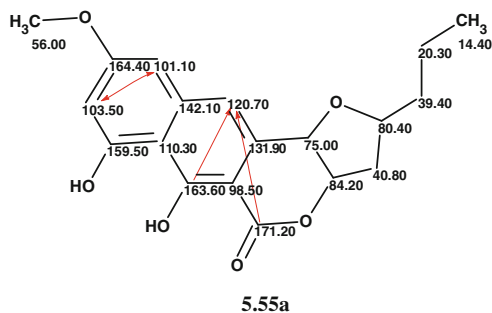
		
$d_A(^{13}\text{C}): 2.727$ $d_N(^{13}\text{C}): 2.690$ $d_l(^{13}\text{C}): 3.187$ $\text{max_}d_A(^{13}\text{C}): 6.160$	$d_A(^{13}\text{C}): 3.042$ $d_N(^{13}\text{C}): 3.175$ $d_l(^{13}\text{C}): 3.400$ $\text{max_}d_A(^{13}\text{C}): 10.630$	$d_A(^{13}\text{C}): 3.069$ $d_N(^{13}\text{C}): 3.085$ $d_l(^{13}\text{C}): 3.106$ $\text{max_}d_A(^{13}\text{C}): 10.630$
		
$d_A(^{13}\text{C}): 3.179$ $d_N(^{13}\text{C}): 2.586$ $d_l(^{13}\text{C}): 3.887$ $\text{max_}d_A(^{13}\text{C}): 13.570$	$d_A(^{13}\text{C}): 3.306$ $d_N(^{13}\text{C}): 2.954$ $d_l(^{13}\text{C}): 3.372$ $\text{max_}d_A(^{13}\text{C}): 7.450$	$d_A(^{13}\text{C}): 3.354$ $d_N(^{13}\text{C}): 3.756$ $d_l(^{13}\text{C}): 4.007$ $\text{max_}d_A(^{13}\text{C}): 14.590$

Fig. 5.79 Lasionectionin: The top structures obtained from program Run 2b

Formally speaking the mentioned three absentee peaks that were included into the HMBC data carry incorrect *structural information*, but the program nevertheless demonstrated the capability of finding a valid solution to the problem even when the initial information was false. The cost of getting a valid solution from the false information is a long FSG time. It should be noted that if structure generation is expected to be time-consuming it is performed overnight, similar to the acquisition of 2D NMR data.

For the sake of completeness, we excluded three false HMBC correlations and repeated the solution with the true experimental data containing only three NSCs—two in the HMBC and one in the COSY spectrum. The numbers of hydrogen atoms on neighboring carbons were introduced. The parameters for FSG were set as $m_g = 3$, $a = 16$, and ^{13}C calculation was performed during structure generation. The results are: $k = 21,837 \rightarrow 10 \rightarrow 6$, $t_g = 7$ min 11 s. The top of the output file ranked with $d_A(^{13}\text{C})$ values is shown in Fig. 5.79.

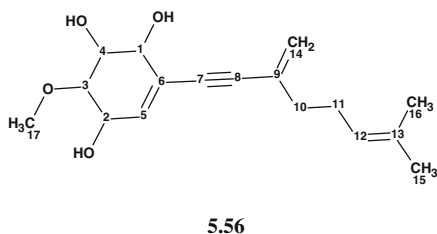
We see that ranking structures with $d_A(^{13}\text{C})$ deviations again allows us to select the best structure which is identical to structure 5.53. Structure 5.55a shows the carbon chemical shift assignments and nonstandard connectivities (double-headed arrow—COSY, single-headed arrows—HMBC).



We emphasize that the application of ^1H multiplicities requires great care. Constraints based on the multiplicities of methyl groups seem to be the most reliable.

5.27 Phomentrioloxin

Cimmino et al. [30] reported the isolation and the chemical characterization of a novel metabolite phomentrioloxin (**5.56**), the main phytotoxin produced by one strain of *Phomopsis* sp. isolated from *Carthamus lanatus*.



Phomentrioloxin has a molecular weight of 292.1675 and a molecular formula $\text{C}_{17}\text{H}_{24}\text{O}_4$ deduced by its HRESIMS, consistent with six hydrogen deficiencies. IR bands observed at 3385, 2186, 1666, 1628, and 1604 cm^{-1} are typical for hydroxyl (3,385), alkyne (2,185), and olefinic (1666, 1628, 1604) groups [18, 31]. The UV spectrum showed absorption maxima associated with an extended conjugated chromophore [32]. For structure elucidation, 1D NMR, HSQC, and HMBC spectra were used in the work [30]. The structure assigned to phomentrioloxin was further confirmed by preparing two key derivatives.

The spectroscopic data (see Table 5.27) presented in the work [30] was input into the Structure Elucidator.

An MCD (Fig. 5.80) was created by the program with options in which the presence of sp (evident from IR spectrum), sp^2 , and sp^3 hybridized carbon atoms in the molecule was allowed.

Table 5.27 Phomentrioloxin: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH shift	HMBC
C1	69.1	76.75	CH	4.35	C6, C7
C2	67.9	72	CH	4.21	–
C3	79.1	77.91	CH	3.7	C1
C4	64.8	69.53	CH	4.51	–
C5	135.3	136.86	CH	6.16	C7, C3, C1
C6	135.4	124.76	C	–	–
C7	87.3	88.35	C	–	–
C8	92.4	94.62	C	–	–
C9	131.5	130.36	C	–	–
C10	37.9	36.52	CH ₂	2.22	–
C11	18.4	26.18	CH ₂	1.64	C13
C12	124.4	123.48	CH	5.13	C15
C13	123.1	132.75	C	–	–
C14	123.8	123.57	CH ₂	5.31	–
C14	123.8	123.57	CH ₂	5.4	C9, C11, C8
C15	27.4	25.69	CH ₃	2.22	C13, C12, C9
C16	26.4	17.67	CH ₃	1.85	C12
C17	59.3	57.17	CH ₃	3.55	C3
O1	100 ^a	–	OH	2.65	C1
O2	110 ^a	–	OH	2.68	C3
O3	120 ^a	–	OH	2.64	–

^a Fictitious ¹⁷O chemical shifts used to distinguish the different atoms of oxygen

MCD overview Seven carbon atoms with chemical shifts in the range 64.80–123.10 are automatically marked in black indicating that their hybridizations are not defined. This uncertainty, which is accounted for by the presence of the *sp*-hybridized carbon atoms in a molecule, can be removed if we use the characteristic chemical shifts of hydrogens attached to four carbon atoms with ¹³C chemical shifts in the range of 64.80–79.10 ppm. Table 5.27 shows that the corresponding ¹H chemical shifts occur in the range of 3.70–4.51 ppm, while characteristic ¹H chemical shifts of $\equiv C-H$ vary between 2.2 and 3.2 ppm [18]. Therefore, the mentioned four carbon atoms cannot be of alkyne type. It is most probable that they have oxygen atoms as neighbors and hence their properties should be marked as *sp*³/*ob*. As the number of olefinic carbons must be even (six in the current case), the carbons C(123.10)–C(135.4) should be marked with the property *sp*²/*fb* [18]. Using an exclusion method we can conclude that carbons C (87.3) and C(92.4) belong to a $C\equiv C$ bond. Analysis of the MCD also shows that there are two carbon atoms—C 64.80 and C 67.90—which have no connectivity at all, which can lead, in general, to an increase in the generation time. However, in this case four carbon atoms—64.8, 67.90, 69.10, and 79.1—must be connected to oxygens (see also the chemical shifts of the corresponding protons [18]), and this

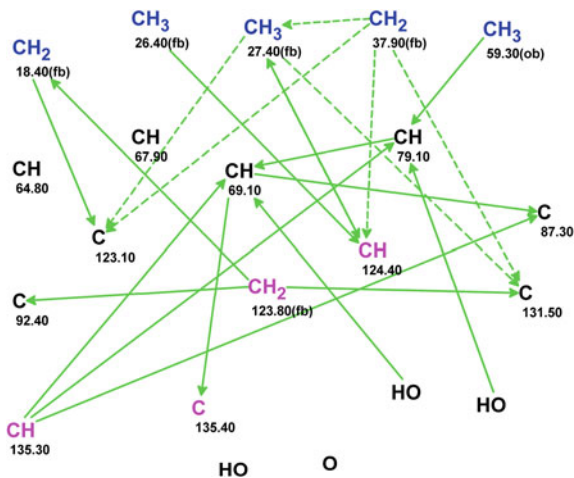


Fig. 5.80 Phomentrioloxin: An MCD created under the condition that the analyzed molecule can contain an alkyne group. Carbon atoms whose hybridization state are not defined are marked with a *black color*

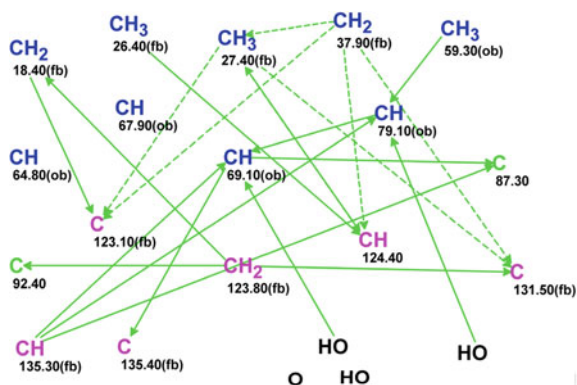


Fig. 5.81 Phomentrioloxin: Modified MCD. The *sp*-hybridized carbon atoms are marked with a *green color*

reduces the isomer space. The MCD modified according to the considerations given above is presented in Fig. 5.81.

Checking the modified MCD did not detect any contradictions and strict structure generation followed by ¹³C chemical shift prediction was run. Results: $k = 20 \rightarrow 3$, $t_g = 0.2$ s and the structures obtained are shown in Fig. 5.82:

The large deviations for the first structure suggest unrevealed contradictions in the HMBC data. FSG was initiated with the option parameters $m = 1$, $a = 16$ (any length of nonstandard correlation is allowed). The results: $k = 630 \rightarrow 476 \rightarrow 61$, $t_g = 1$ s, with 1 from 18 correlations extended and 18 from 18 possible connectivity

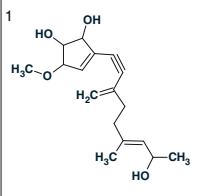
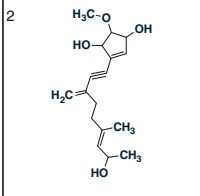
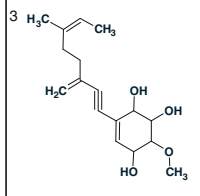
		
$d_A(^{13}\text{C}): 4.867$ $d_N(^{13}\text{C}): 5.380$ $d_I(^{13}\text{C}): 5.417$ $\text{max}_dA(^{13}\text{C}): 14.890$	$d_A(^{13}\text{C}): 4.926$ $d_N(^{13}\text{C}): 5.341$ $d_I(^{13}\text{C}): 5.446$ $\text{max}_dA(^{13}\text{C}): 14.890$	$d_A(^{13}\text{C}): 5.695$ $d_N(^{13}\text{C}): 4.965$ $d_I(^{13}\text{C}): 5.179$ $\text{max}_dA(^{13}\text{C}): 15.470$

Fig. 5.82 Phomentrioxin: Structures obtained as a result of strict structure generation

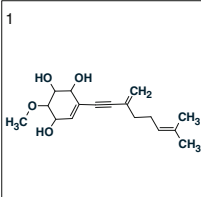
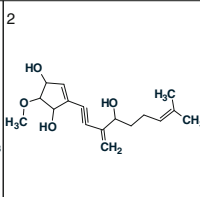
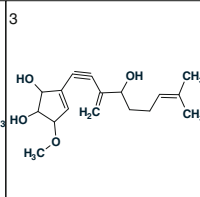
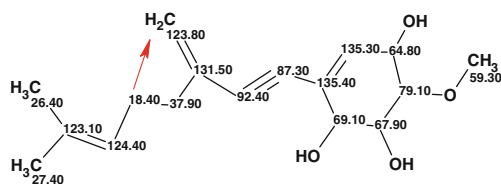
		
$d_A(^{13}\text{C}): 3.930$ $d_N(^{13}\text{C}): 3.376$ $d_I(^{13}\text{C}): 3.502$ $\text{max}_dA(^{13}\text{C}): 10.640$	$d_A(^{13}\text{C}): 4.335$ $d_N(^{13}\text{C}): 4.303$ $d_I(^{13}\text{C}): 5.172$ $\text{max}_dA(^{13}\text{C}): 9.440$	$d_A(^{13}\text{C}): 4.252$ $d_N(^{13}\text{C}): 4.339$ $d_I(^{13}\text{C}): 5.143$ $\text{max}_dA(^{13}\text{C}): 9.440$

Fig. 5.83 Phomentrioxin: The top structures of the ranked output file obtained as a result of FSG

combinations were used during generation. The top three structures of the ranked output file are shown in Fig. 5.83.

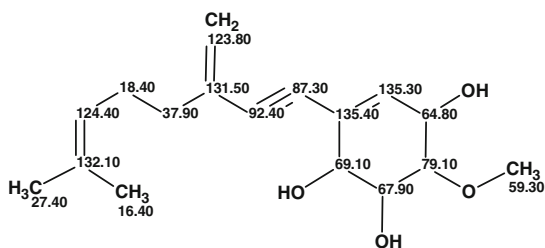
The first ranked structure coincided with the structure of phomentrioxin and the ^{13}C chemical shift assignments were the same as in article [30] (shown in 5.57):



5.57

The HMBC data contain a single nonstandard connectivity which was revealed between carbons C 123.80 and C 18.40. The large deviation values are accounted for by poor prediction of two of the ^{13}C chemical shifts: 123.1 (132.8 calculated)

and 26.4 (17.7 calculated). Because both chemical shifts belong to a common structural fragment, we performed a search for the fragment $(\text{CH}_3)_2\text{C}=\text{CH}-\text{CH}_2\text{CH}_2-$ in the ACD/DB containing 425,000 chemical structures. As a result 1,931 structures were found. In *all* structures, experimental chemical shifts of interest were equal to 132–133 and 16–18 ppm correspondingly, which allows us to suggest that there are two misprints in the table of NMR data published in article [30]: 123.1 instead of 132.1 and 26.4 instead of 16.4 ppm. When new chemical shifts were assigned to structure **5.57**, the chemical shift recalculation was resulted with the following deviations: $d_A = 2.96$, $d_I = 2.64$ and $d_N = 2.5$ ppm. Finally, structure **5.58** shows the phomentrioloxin molecule with new (supposedly correct) chemical shift assignment:



5.58

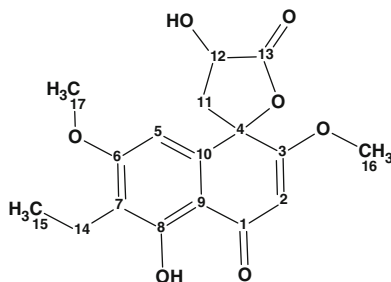
The example shows how the application of Structure Elucidator in combination with chemical shift prediction as a support, together with ACD/DB, allows one not only to quickly establish the structure of a new compound, but also to correct incorrect chemical shift assignments presented in the published materials.²

5.28 Perenniporide A

During a search for new bioactive natural products from the fungi of unique types, Feng et al. [33] chemically investigated a strain of *Perenniporia* sp. *Perenniporia* is a mostly perennial genus found on dead and living hardwood and conifers. To date only a few bioactive secondary metabolites have been isolated from the species of this genus. Fractionation of an ethyl acetate extract prepared from a solid-substrate fermentation of the fungus afforded four *new* naphthalenone derivatives, named perenniporides A–D and several known compounds.

Here we will describe the structure elucidation of perenniporide A (**5.59**) whose structural assignment was confirmed by X-ray crystallography [33].

² We contacted the authors about our attempts to correct the chemical shift assignment, but unfortunately without any response.



5.59

Perenniporide A was assigned the molecular formula $C_{17}H_{18}O_7$ (nine degrees of unsaturation) on the basis of HRESIMS m/z 335.1124 (calculated for $C_{17}H_{19}O_7$, 335.1125). An IR broad absorption band with maximum at $3,330\text{ cm}^{-1}$ is obviously associated with the stretching vibrations of hydroxyl groups. An IR band observed at $1,792\text{ cm}^{-1}$ is characteristic for a carbonyl group of strained lactones, the band near $1,645\text{ cm}^{-1}$ can be associated with the presence of an α, β -conjugated ketone and the absorption band at $1,608\text{ cm}^{-1}$ is typical for a benzene ring [18]. For structure elucidation, the authors [33] used 1D NMR spectra in combination with HSQC and HMBC correlations tabulated in the article.

The spectroscopic data used for the structure determination with the assistance of StrucEluc are presented in Table 5.28 and the MCD is shown in Fig. 5.84.

MCD overview Because the molecule contains oxygen atoms, five carbon atoms whose ^{13}C chemical shifts fall in the interval 78–120 ppm are of light blue color, i.e., their hybridization states are defined as “ sp^3 or sp^2 ”. The quaternary sp^2 carbons did not receive any label defining the possibility of an oxygen atom neighborhood. Though the carbons C 176.9 and C 190.0 are most probably related to carbonyl groups (and hints regarding to the presence of ester and ketone groups were obtained from the IR spectrum), no corresponding labels “*ob*” were set by the user as the molecule under investigation is relatively small, while the number of HMBC connectivities (ca. 30) is quite large. Four distinct singlets and two doublets observed in the ^1H NMR spectrum (see Table 5.28) allowed us to introduce the numbers of hydrogen atoms to be attached to the corresponding neighboring atoms.

First run Checking the MCD showed the absence of contradictions in the 2D NMR data, and the structure generation process was run from the automatically created MCD with the results: $k = 4 \rightarrow 2 \rightarrow 2$, $t_g = 0.003\text{ s}$. Both generated structures are shown in Fig. 5.85.

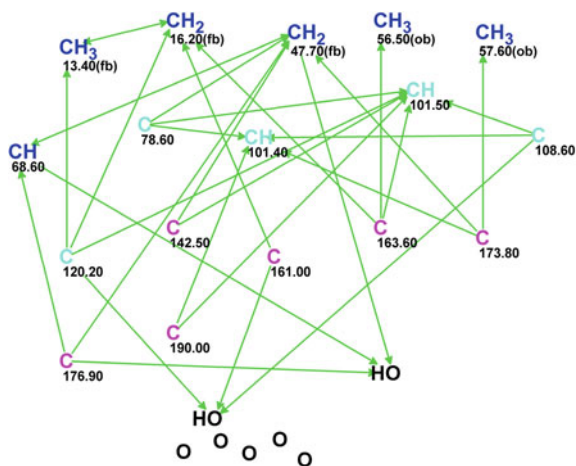
The huge values of the deviations and the exotic nature of the compound unambiguously indicate that the structures are wrong. Assuming that the axiomatic knowledge of the system is inconsistent, the next attempt should be done using FSG.

Table 5.28 Perenniporide A: Spectroscopic NMR data

Label	δX	δC_{calc}	XH_n	δH	M(J)	HMBC
C1	190	187.72	C	–	–	–
C2	101.4	104.44	CH	5.77	s	C4, C1, C9, C3
C3	173.8	172.45	C	–	–	–
C4	78.6	84.75	C	–	–	–
C5	101.5	103.95	CH	6.81	s	C1, C7, C9, C4, C6, C10
C6	163.6	164.27	C	–	–	–
C7	120.2	120.16	C	–	–	–
C8	161	160.17	C	–	–	–
C9	108.6	106.5	C	–	–	–
C10	142.5	142.05	C	–	–	–
C11	47.7	41.67	CH ₂	2.5	u	–
C11	47.7	41.67	CH ₂	3.17	u	C13, C10, C12, C3, C4
C12	68.6	67.47	CH	4.98	u	C11, C13
C13	176.9	177.28	C	–	–	–
C14	16.2	15.73	CH ₂	2.67	q(7.5)	C7, C6, C8, C15
C15	13.4	13.19	CH ₃	1.09	t(7.5)	C14, C7
C16	57.6	57.13	CH ₃	3.99	s	C3
C17	56.5	55.65	CH ₃	3.98	s	C6
O1	100 ^a	–	OH	13.14	u	C9, C7, C8
O2	110 ^a	–	OH	5.3	u	C11, C12, C13

^a Fictitious ¹⁷O chemical shifts

Fig. 5.84 Perenniporide A:
The MCD



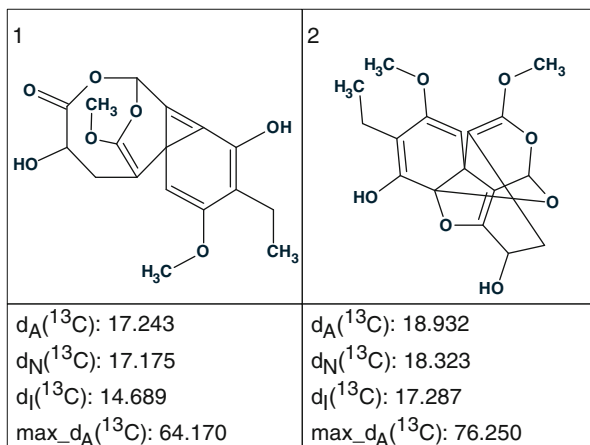


Fig. 5.85 Perenniporide A: The structures generated as a result of the first program run

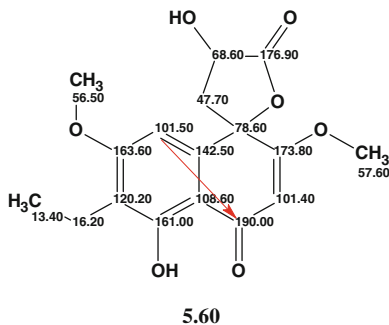
Second run FSG was performed using options $m = 1$, $a = 16$, and the following result was obtained: $k = 133 \rightarrow 89 \rightarrow 89$, $t_g = 3$ s, with the average deviations calculated for the best structure as 7–11 ppm, which allowed us to suggest that it is worth to try repeating the FSG with the parameters $m = 2$, $a = 16$.

Third run The following result was obtained: $k = 4,697 \rightarrow 3,322 \rightarrow 2,981$, $t_g = 3$ min, and again the deviations turned out to be fairly large (4.5–5.5 ppm), while the best structures are again rather exotic. A number of the structures contain a benzene ring whose presence is rather probable on the basis of the IR spectrum.

At this stage we suspected that a potentially correct structure was being rejected by the spectral filter. As noted in Sect. 1.3, the spectral filter was checked using hundreds of thousands of NMR spectra, but there is no guarantee that an exception will not be met in some rare cases. This means that an experimentally measured chemical shift or vibrational frequency does not fall into the interval of spectral features characteristic for some fragments existing in the verified molecule. To check this hypothesis structure generation should be carried out with the spectral filter switched off.

Fourth run FSG was repeated with the options used for the second run ($m = 1$, $a = 16$), but the option “Allow Filter during Generation” was disabled. Results: $k = 133 \rightarrow 109 \rightarrow 107$, $t_g = 3$ s, and the two top ranked structures are presented in Fig. 5.86.

The ^{13}C chemical shift deviation values and the difference $\Delta = d(\#2) - d(\#1) \cong 6$ ppm, as well as substructures included into structure #1, allow us to unambiguously distinguish this structure as correct. The chemical shift assignment which coincided with that suggested by the authors [33] is shown in structure 5.60 (an arrow displays the nonstandard connectivity).



It was interesting to identify the reason for rejecting the correct structure by the spectral filter. For this goal, the command Structure/Check by Filter was executed as applied to the answer structure. The following program message was received:

“Structure has been rejected by the following tests:
 NMR: **CH₂** Groups
 Structure contains contradictory group
 [65] CH₃-CH₂-[Ar]
 Characteristic features of group in C13 NMR:

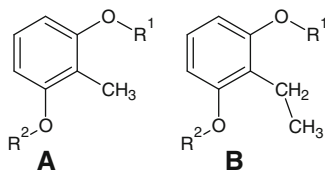
Chem. Shift	J	Multiplicity	Integral
18.8–38.2	120–150	3	1

This message can be interpreted in the following way: the CH₂ of the ethyl group attached to a benzene ring possesses a characteristic ¹³C chemical shift interval of between 18.8 and 38.2 ppm. The corresponding group of structure **5.60** had δC 16.2 ppm, which is outside the limits of the characteristic interval. This is the reason why the real structure of perenniporide A was rejected by the spectral filter.

Fig. 5.86 Perenniporide A:
 The top ranked structures of
 the output file

1		2	
	$d_A(^{13}\text{C}): 1.699$ $d_N(^{13}\text{C}): 1.813$ $d_I(^{13}\text{C}): 1.443$ $\text{max}_dA(^{13}\text{C}): 6.150$		$d_A(^{13}\text{C}): 7.901$ $d_N(^{13}\text{C}): 7.018$ $d_I(^{13}\text{C}): 7.769$ $\text{max}_dA(^{13}\text{C}): 46.210$

In this connection a question arises: Is the chemical shift of 16.2 ppm a rare case specific for structure **5.60** or can some regularity be revealed? In the course of challenging the Structure Elucidator program it was revealed that the methyl group included in fragment A between the two O–R substituents has an unexpectedly small ^{13}C chemical shift around 6–7 ppm (the typical value is of ~ 20 ppm for Ar–CH₃ [18]).

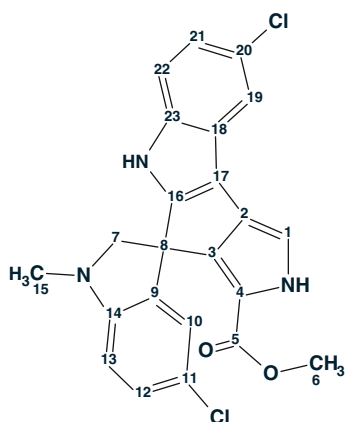


With this in mind fragment A was placed in a special Library of Exceptions (see Sect. 1.3.3.6). If such a fragment is detected in the verified structure the algorithm takes into account the specific property of the fragment and the structure is not rejected by the program due to the small value of the CH₃ chemical shift. Fragment B existing in structure **5.60** was searched in the ACD/CNMR database and as a result 25 structures were found in which the chemical shift of the CH₂ group varies between 16 and 18 ppm. Therefore, fragment B also should be included into the Library of Exceptions. This example illustrates how continuous optimization of the knowledge base of the system increases its strength and flexibility. When a specific fragment (an exception) that was not included into the Filter Library is present in the verified structure the program fails to determine the correct structure. However, the Spectral Filter failure does not mean that it is a failure of the CASE approach, because any knowledge (including the system knowledge) is restricted and is constantly growing.

As ^{13}C chemical shift prediction is fast and accurate now, we suggest that spectral filtering based on characteristic spectral intervals can be replaced by ^{13}C chemical shift prediction during the structure generation process. This change in the general CASE strategy will raise the reliability of the structure elucidation and exclude program failures caused by filter imperfection. At the same time, one should take into account that the number of generated structures will increase in this case.

5.29 Spiroindimicin B

Zhang et al. [34] isolated from a deep-sea-derived material new bisindole alkaloids spiroindimicins A–D. Structural elucidation of these compounds revealed the presence of unusual [5, 6] or [5] spiro-ring containing skeletons. Spiroindimicin A possesses a [5, 6] spiro-ring system, whereas spiroindimicins B–D contain a distinct [5] spiro-ring system. Spiroindimicins represent *the first example of an unusual spiro-ring containing bisindole alkaloids* of bacterial origin. Spectroscopic data (1D NMR, HSQC, HMBC) for spiroindimicin B (**5.61**) whose structure was confirmed by crystallographic analysis [34] were available from the original article and they were used for challenging StrucEluc.



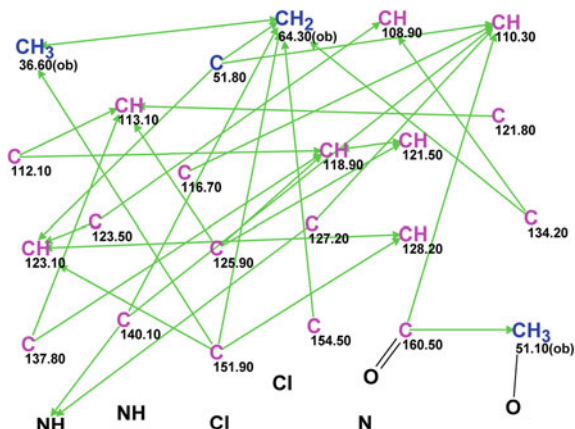
5.61

Table 5.29 Spiroindimicin B: Spectroscopic data

Label	δX	δC_{calc}	XH_n	δH	M(J)	C HMBC
C1	110.3	109.68	CH	6.91	u	C2, C5, C4, C3, C8
C2	127.2	123.35	C	–	–	–
C3	140.1	140.78	C	–	–	–
C4	116.7	122.36	C	–	–	–
C5	160.5	164.51	C	–	–	–
C6	51.1	51.51	CH ₃	3.61	u	C5
C7	64.3	73.67	CH ₂	4.07	u	C16, C3, C8, C15
C7	64.3	73.67	CH ₂	3.64	u	C16, C8, C9, C14
C8	51.8	59.77	C	–	–	–
C9	134.2	126.93	C	–	–	–
C10	123.1	125.44	CH	6.54	u	C11, C14, C12, C8
C11	123.5	127.52	C	–	–	–
C12	128.2	128.9	CH	7.07	u	–
C13	108.9	114.04	CH	6.57	u	C9, C11
C14	151.9	151.54	C	–	–	–
C15	36.6	40.62	CH ₃	2.88	u	C7, C14
C16	154.5	155.01	C	–	–	–
C17	112.1	112.06	C	–	–	–
C18	121.8	123.59	C	–	–	–
C19	118.9	121.17	CH	7.62	u	C23, C17, C21, C20
C20	125.9	122.49	C	–	–	–
C21	121.5	124.06	CH	7.06	u	C20, C19
C22	113.1	113.75	CH	7.15	u	C17, C18, C20, C23
C23	137.8	136.06	C	–	–	–
N1	100 ^a	–	NH	8.88	u	C2, C3
N2	200 ^a	–	NH	8.45	u	–

^a Fictitious values of ¹⁵N chemical shift

Fig. 5.87 Spiroindimicin B: The edited MCD



Spiroindimicin B was assigned a molecular formula of $C_{23}H_{17}Cl_2N_3O_2$ (sixteen degrees of unsaturation) by HRESIMS (m/z 436.0602, $[M-H]^-$, calculated 436.0620). As the ratio $n(\text{skeletal})/n(\text{hydrogens})$ is close to 2, we can expect that the structural elucidation will be challenging. The 1D and 2D NMR data of compound **5.61** are presented in Table 5.29.

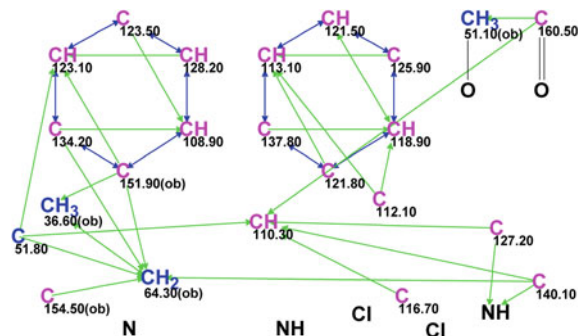
The inspection of the 1H , ^{13}C , and HSQC NMR spectroscopic data of **5.61** allowed the authors [34] to reveal two exchangeable protons, one *N*-methyl [δC 36.6, δH 2.88 (3H, s)], one methoxyl [δC 51.1, δH 3.61 (3H, s)], one methylene, 18 olefinic and/or aromatic carbons (seven of which are protonated), and one unprotonated sp^3 carbon. Carbon C 160.5 was assigned to an ester carbonyl (though this chemical shift is also characteristic of an amide). An MCD, where some of the most convincing assumptions by the authors were taken into account, is shown in Fig. 5.87.

MCD overview The presence of CH_3-O (51.10) and CH_3-N (36.60) groups is obvious. The bond CH_3-O can be drawn without question while there is a question regarding which of the three nitrogen atoms is really bonded to the methyl group. The single CH_2 group at δC 64.3 is automatically assigned as bonded to oxygen or nitrogen atoms. It is possible to assume that carbons C(151.9), C(154.5), and C(160.5) are chemically connected with heteroatoms, and a connectivity between C(51.10) and C(160.5) suggests that $MeO-CO$ exists in the unknown compound.

MCD checking for the presence of contradictions produced the message: "Current Molecular Connectivity Diagram passed all tests." As the NSCs were not detected by the program, strict structure generation was initiated. It was stopped after 1.5 h as no generated structure was stored within this time. This suggests that there is a lack of structural constraints (independent of the presence or absence of NSCs) and therefore the user should introduce new "axioms" to allow for the problem to be solved in a reasonable time.

As mentioned earlier, 18 olefinic and/or aromatic carbons can be present in the molecule in accordance with the authors' [34] suggestion. In this situation, it is worth trying to distinguish at least one set of sp^2 hybridized carbon atoms that can

Fig. 5.88 Spiroindimicin B: A modified MCD containing two benzene rings selected manually



be arranged to form an aromatic ring. As mentioned earlier (Sect. 5.21, see Fig. 5.54), the most convincing atom selection is proven by the “characteristic” geometric figure formed on the MCD by ${}^3J_{\text{CH}}$ HMBC connectivities between carbons of a given benzene ring.

If in addition COSY and ${}^2J_{\text{CH}}$ HMBC correlations are observed for some atom pairs of the set of examined atoms, the recognition of benzene rings on MCD is alleviated.

Manual analysis of the MCD displayed in Fig. 5.87 allowed us to distinguish two groups of sp^2 carbons which (groups) make up “closed” sets of connectivities. These sets are subsets of connectivities forming the “ideal” figure shown above. The modified MCD is shown in Fig. 5.88 where all assumptions are presented graphically.

Figure 5.88 shows two groups of carbon atoms which are assumed to form two benzene rings. These atoms are connected by 1–1 connectivities (blue lines) in each ring. Such representation of the connected carbon atoms allows us to depict the assumed benzene ring without using alternating single and double bonds.

The next attempt to generate structures was then made. Strict structure generation under significantly enhanced constraints instantly resulted in zero output structures.

In such a situation, the only way to solve the problem is to successively try different values of nonstandard connectivities m ($m = 1, 2, 3, \dots$) until an acceptable solution to the problem is found. This procedure can be carried out either automatically by running FSG with appropriate options ($m = 1-20$, $a = 16$, “stop generation when structures generated”) or manually in a “step by step” mode. To clarify the “mechanism” of the procedure we will use here the latter mode.

FSG (settings: $m = 1$, $a = 16$, “stop generation when structures generated”) was initiated and instantly completed, but again no structure was generated. The next run was performed with parameters $m = 2$, $a = 16$. In so doing, ${}^{13}\text{C}$ NMR spectrum calculation and structural and spectral filtration were used. The filtration using predicted chemical shifts was controlled by the following parameters (see Fig. 2.10): only those structures were saved for which an average deviation $d < 6$ ppm, and a maximum deviation $d_{\text{max}} < 25$ ppm.

Results: $k = 52,950 \rightarrow 0$, 153 from 153 connectivity combinations were tried, $t_g = 4$ min.

1	2	3	4
$d_A(^{13}\text{C}): 3.104$ $d_N(^{13}\text{C}): 2.643$ $d_I(^{13}\text{C}): 4.933$ $\text{max_}d_A(^{13}\text{C}): 9.350$	$d_A(^{13}\text{C}): 3.263$ $d_N(^{13}\text{C}): 2.970$ $d_I(^{13}\text{C}): 3.453$ $\text{max_}d_A(^{13}\text{C}): 15.050$	$d_A(^{13}\text{C}): 3.356$ $d_N(^{13}\text{C}): 3.173$ $d_I(^{13}\text{C}): 4.640$ $\text{max_}d_A(^{13}\text{C}): 9.010$	$d_A(^{13}\text{C}): 3.478$ $d_N(^{13}\text{C}): 2.941$ $d_I(^{13}\text{C}): 4.542$ $\text{max_}d_A(^{13}\text{C}): 13.060$
5	6	7	8
$d_A(^{13}\text{C}): 3.582$ $d_N(^{13}\text{C}): 3.019$ $d_I(^{13}\text{C}): 4.793$ $\text{max_}d_A(^{13}\text{C}): 17.360$	$d_A(^{13}\text{C}): 3.597$ $d_N(^{13}\text{C}): 2.893$ $d_I(^{13}\text{C}): 5.417$ $\text{max_}d_A(^{13}\text{C}): 9.460$	$d_A(^{13}\text{C}): 3.706$ $d_N(^{13}\text{C}): 2.932$ $d_I(^{13}\text{C}): 3.956$ $\text{max_}d_A(^{13}\text{C}): 18.390$	$d_A(^{13}\text{C}): 3.904$ $d_N(^{13}\text{C}): 3.043$ $d_I(^{13}\text{C}): 5.318$ $\text{max_}d_A(^{13}\text{C}): 14.910$

Fig. 5.89 Spiroindimicin B: Top structures of the ranked output file

Then FSG was continued with $m = 3$ and completed with a tangible result this time:

$k = 1,108,478 \rightarrow 3,311 \rightarrow 1,891$, $t_g = 1$ h 45 min, 816 from 816 possible connectivity combinations were tried. The top ranked structures in the structure file ranked by $d_A(^{13}\text{C})$ values is shown in Fig. 5.89.

It is immediately obvious all distinguished structures are very similar. At the same time, the deviation values calculated by different methods are close within the series corresponding to the same methods. The values of the deviations are large due to the novelty of the target structure. The carbon atom chemical shift assignments for spiroindimicin B and three nonstandard connectivities are shown on structure 5.62.

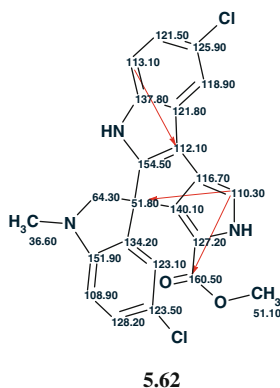
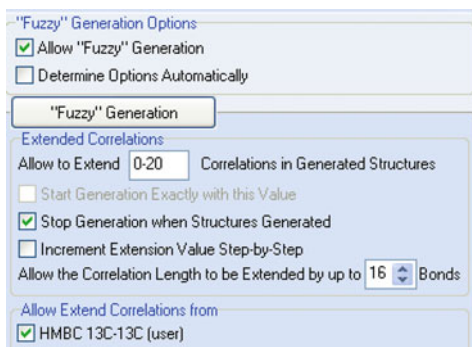


Fig. 5.90 Spiroindimicin B:
The options of FSG

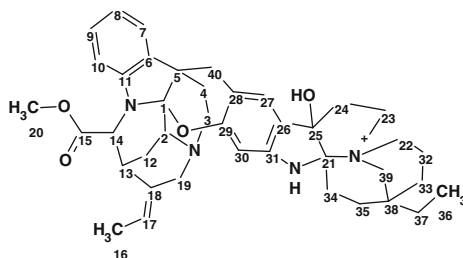


Therefore, in spite of some fuzziness of the analysis results, the correct structure was distinguished among ca. 2,000 candidate structures. Note that we used the most reliable option for FSG—the lengths of NSCs existing in the 2D NMR data were assumed to be unknown. As mentioned above, the problem could be solved with the same results during one program run with the FSG options shown in Fig. 5.90.

Allowing for any lengths for NSCs, i.e., choosing $a = 16$, provides higher reliability for the structure elucidation but the result is the relatively long time required for structure generation. For comparison, when the problem was resolved automatically with the following parameter set: $m = 0-20$, $a = 1$ (i.e. all NSCs were suggested to be of $^4J_{\text{CH}}$ type), similar results were obtained in 23 m, i.e. five times faster.

5.30 Goniomedine A

Beniddir et al. [35] chemically investigated the stem bark of *Gonioma malagasy* Mgf and P.Bt (Apocynaceae), a tree collected in the forest of Toliara in the southwest of Madagascar. They reported the isolation and structure elucidation of Goniomedines A and B representing the first example of an *unprecedented class* of quebrachamine-pleiocarpamine bisindole alkaloids which possess a unique ring system in natural products characterized by an unprecedented linkage between the two dihydroindole moieties. Here we discuss the computer-based structure elucidation of Goniomedine A (**5.63**).



5.63

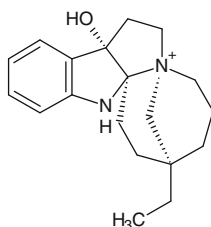
The authors established that the UV spectrum of goniomedine A was characteristic of a dihydroindole chromophore with absorption maxima at 254 and 296 nm. The IR spectrum of **5.63** showed absorption bands at $3,400\text{ cm}^{-1}$, for OH and NH elongation, and at $1,750\text{ cm}^{-1}$, admittedly for an aliphatic ester.

The HRESIMS of **5.63** showed an $[M]^+$ ion peak at m/z 649.3756 suggesting a molecular formula of $C_{40}H_{49}N_4O_4$. When an attempt to add this molecular formula to the StrucEluc program was made, the program displayed the following message: “Value of DBE of the current molecular formula $C_{40}H_{49}O_4N_4$ is fractional (18.5). Do you want to accept this MF?”

This message can be interpreted as an indication of the presence of a positively charged atom in the molecule. Considering the origin of the compound we can expect that the molecule contains a charged nitrogen N^+ atom.

The authors reported that the 1H and ^{13}C NMR data and HSQC spectrum suggest the presence of five sp^3 quaternary carbons, 13 sp^3 methylenes, three sp^3 methines, three methyl groups, seven sp^2 methines, and eight sp^2 quaternary carbons. We believe that the cited unambiguous classification of carbon atoms with the various types of hybridization is possible only on the basis of preliminary investigations using related compounds or from an already identified molecule. 1H – 1H COSY (13), HSQC, TOCSY, and HMBC (121) correlations were used in [35] to perform the structure elucidation. The spectral data utilized for the CASE analysis are presented in Table 5.30.

The authors revealed that the 1H and ^{13}C NMR signals of the righthand part of the structure **5.63** shown below were consistent with a rhazidine moiety, a quebrachamine-type alkaloid **5.64** that was previously isolated from *Gonioma malagasy*.



5.64

This valuable observation confirmed the righthand part of the molecule. For the lefthand part, the downfield shift of C1 (93.3) was considered as an indication of the fact that it is linked to both nitrogen and oxygen atoms [35]. However, this conclusion is not absolutely evident as a chemical shift in a range of 93–95 ppm can also be related to a sp^3 hybridized carbon connected with one or two oxygens [18]. The CASE application presupposes performing the structure elucidation process using a minimal number of user assumptions (“axioms” postulated by a researcher). An MCD containing more than a hundred correlations (see Table 5.30) would look way too overcrowded, so Fig. 5.91 displays only atoms without any connectivities to provide analysis of the atom properties, which were slightly edited in accordance

Table 5.30 Goniomedine A: Spectroscopic data

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C1	93.3	100.68	C	–	–	–	–
C2	50.4	52.37	CH	4.15	u	2.14	C4, C1, C12, C13, C19
C3	46.2	52.03	CH ₂	3.38	u	2.22	C5, C40, C2
C3	46.2	52.03	CH ₂	3.86	u	–	–
C4	30.7	28.78	CH ₂	2.22	u	3.38	C40, C3, C5
C5	43.1	46.11	C	–	–	–	–
C6	134.3	130.99	C	–	–	–	–
C7	121.9	121.74	CH	7.33	d(7.2)	6.84	C10, C5, C11, C9
C8	120.6	119.91	CH	6.84	dd(7.2, 7.8)	7.10, 7.33	C7, C6, C11, C10, C9
C9	127.8	126.84	CH	7.1	t(7.8)	6.84, 6.42	C10, C8, C11, C7, C6
C10	110.8	109.67	CH	6.42	d(7.8)	7.1	C6, C7, C5, C11, C8
C11	145.5	147.51	C	–	–	–	–
C12	24.9	28.46	CH ₂	2.93	u	–	–
C12	24.9	28.46	CH ₂	2.14	u	3.59, 4.15	C14, C2, C18, C13, C1
C13	30.2	32.14	CH	3.59	u	–	–
C14	57.1	56.76	CH	4.66	u	3.59	C12, C13, C15, C18
C15	169.6	170.54	C	–	–	–	–
C16	11.9	12.25	CH ₃	1.65	u	5.65	C17, C19, C18
C17	124.2	118.95	CH	5.65	u	1.65	C16, C19, C13
C18	130.2	136.72	C	–	–	–	–
C19	52	47.49	CH ₂	3.82	u	–	–
C19	52	47.49	CH ₂	4.74	u	–	C4, C17, C18, C2
C20	52.2	51.35	CH ₃	3.83	u	–	C15
C21	100.8	100.82	C	–	–	–	–
C22	56.2	53.6	CH ₂	3.44	u	1.84	C23, C21, C32
C22	56.2	53.6	CH ₂	3.65	u	–	C39, C33
C23	60.3	60.95	CH ₂	3.31	u	2.41	C39, C24
C23	60.3	60.95	CH ₂	3.59	u	2.14, 4.66	C12, C17, C21, C18, C2, C25, C14
C24	37.8	39.11	CH ₂	2.91	u	–	C23
C24	37.8	39.11	CH ₂	2.41	u	3.31	C21, C26, C25
C25	88.9	85.61	C	–	–	–	–
C26	125	124.98	C	–	–	–	–
C27	124	128.37	CH	7.12	s	–	C29, C25, C31, C30, C40
C28	110.6	117.47	C	–	–	–	–
C29	154.6	153.83	C	–	–	–	–
C30	96.7	99.66	CH	6.05	s	–	C31, C29, C26, C28, C25, C40
C31	146.3	144.37	C	–	–	–	–

(continued)

Table 5.30 (continued)

Label	δX	δC_{calc}	XH_n	δH	M(J)	COSY	HMBC
C32	19.2	18.19	CH ₂	2.46	u	–	C22, C33, C35
C32	19.2	18.19	CH ₂	1.84	u	3.44, 1.55	C38
C33	30.3	31.57	CH ₂	1.55	u	1.84	C35, C38, C32, C37
C33	30.3	31.57	CH ₂	1.74	u	–	C39, C22
C34	28.3	27.64	CH ₂	2.47	u	–	C38
C34	28.3	27.64	CH ₂	2.17	u	1.73	C21, C35, C25
C35	30	27.74	CH ₂	1.73	u	2.17	C39, C37, C21, C34, C38
C36	6.4	8.28	CH ₃	0.86	u	1.31	C37, C38
C37	33.9	31.19	CH ₂	1.31	u	0.86	C38, C33, C36, C39
C38	31.6	35.24	C	–	–	–	–
C39	61.9	63.88	CH ₂	3.69	u	–	C23
C39	61.9	63.88	CH ₂	3.29	u	–	C22, C38, C21, C33
C40	25.7	35.44	CH ₂	3.53	u	–	–
C40	25.7	35.44	CH ₂	3.49	u	–	C1, C6, C4, C28, C29, C27, C5
N1	120 ^a	–	NH	7.16	u	–	C34, C21, C25, C31, C30

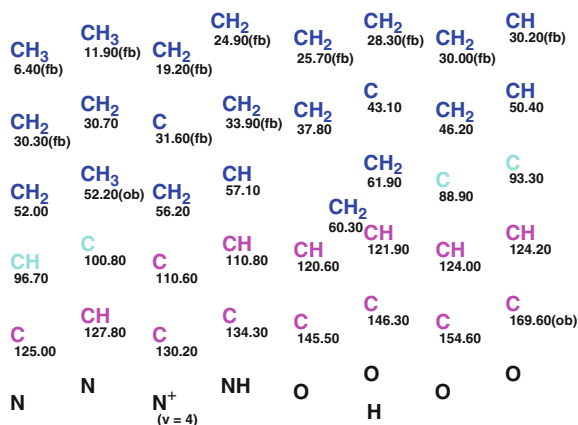
^a Fictitious value of ¹⁵N chemical shift

with the characteristic NMR chemical shifts. The most “critical” atoms C 88.9–C 100.8, which are light blue in color by the program (*sp*³ or *sp*²), were left without any edits.

MCD overview All connectivities are hidden on the MCD, but this does not prevent the program from checking the MCD for the absence of contradictions. MCD checking resulted in the following program message being displayed: “The minimum number of non-standard connectivities is 4.” Experience shows that the real number of NSCs can be significantly larger than the minimum value determined by the program. With this in mind, we cannot expect that problem solving will be simple and quick because FSG is unavoidable. Visual inspection of the HMBC spectral pattern (see SI to [35]) showed that all peaks, including those that have negligible intensities, were input by the authors [35] into the table of experimental data.

This allows us to decide to delete some of the HMBC correlations which are associated with the weakest peaks and potentially contributing to the set of NSCs. Therefore, the four weakest peaks were removed from the HMBC data (and the corresponding connectivities were deleted on MCD). It is probable, of course, that some of the HMBC correlations of standard (^{2–3}*J*_{CH}) length can be excluded from the data along with NSCs, but if the total number of correlations is large, as in this case (121 correlations are observed in the HMBC), then it likely will not influence the time associated with the program application.

Fig. 5.91 Goniomedine A: The initial MCD showing only atoms with assigned properties. All connectivities are hidden to ease the atom property analysis



Due to the presence of a positively charged nitrogen atom, the properties of one nitrogen atom were edited in the window “Edit properties of atom #” as displayed in Fig. 5.92. After editing, the charged nitrogen atom (with its valence marked as $v = 4$) appears in the MCD as shown in Fig. 5.91.

The speed of FSG obviously depends on the number and the strictness of the structural constraints imposed by the initial data. Therefore, to speed up structure generation it was necessary to further inspect the MCD and try to introduce additional constraints based on more or less reliable assumptions. Carbon atoms characterized by ^{13}C chemical shifts in the range between 46 and 62 ppm were supplied with the properties sp^3/ob , taking into account the characteristic ^{13}C and associated ^1H chemical shifts (see Table 5.30). The methine group at 96.7 (δH 6.05) and atoms with ^{13}C chemical shifts from δC 110 ppm and higher were assigned sp^2 hybridization. The carbon C(169.6) was assigned to a carbonyl group (IR band at $1,750\text{ cm}^{-1}$). Moving and grouping the sp^2 hybridized carbons in the MCD window (connectivities were displayed for this goal) allowed us to distinguish two benzene rings. The carbons belonging to each ring were connected by hand with “COSY like” connectivities (connectivities but not bonds!) of one bond length. After benzene ring selection the properties of the remaining carbon atoms were assigned as follows: C 88.90— sp^3/ob , C 93.30— sp^3/ob , C 100.80— $sp^3/2ob$. Here we took

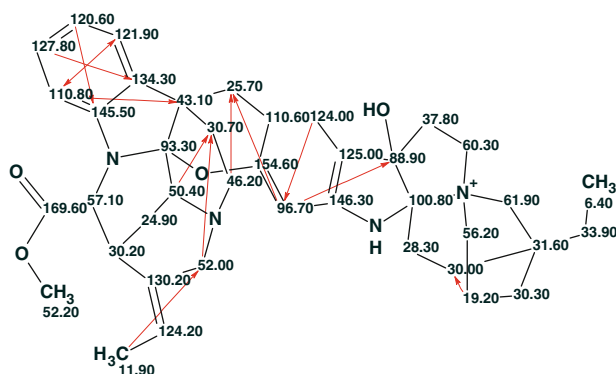
The 'Atom Properties' dialog box shows the following settings:

- Connection with Heteroatoms: auto
- Number of Hydrogens on Neighbor Atoms: (empty field)
- Hybridization State: auto
- Charge: 1
- Valency: 4
- Allow Non-default Valences

Fig. 5.92 Goniomedine A: Setting the properties of a charged nitrogen atom

into account that the total number of definitely sp^2 -hybridized carbon atoms involved both in aromatic rings and non-aromatic carbon double bonds should be even, hence one can suggest the carbon atom C(100.8) is sp^3 -hybridized and connected to two heteroatoms. A fragment $\text{CH}_3\text{-O-C=O}$ whose existence is evident from the carbon chemical shifts and HMBC correlations was drawn by hand. In addition, three weak HMBC connectivities were suggested to have a length of 1–3 C–C chemical bonds, i.e., no risky assumptions were introduced. The final MCD is shown in Fig. 5.93, where $^{2-3}J_{\text{CH}}$ connectivities are omitted to make the picture more understandable.

First run FSG was initiated from the MCD presented in Fig. 5.93 in the mode **Determine Options Automatically**. The results gave: $k = 368 \rightarrow 13 \rightarrow 13$, $t_g = 1$ h 28 min, 5 from 71 correlations were extended during generation, 614,566 from 13,019,909 possible fuzzy combinations were used during generation, i.e., only $\sim 5\%$ of all possible combinations of connectivities were involved in the generation process at $m = 5$. FSG was started with $m = 4$ (minimum number of NSCs) and completed at $m_g = 5$ when the resulting structure file was stored. Figure 5.94 which shows the three top structures of the ranked output file allows us to conclude that the best structure is Goniomedine A. Structure 5.65 displays the ^{13}C chemical shifts assigned to the identified molecule. The arrows show 12 NSCs included into the HMBC data set in the work [35].



5.65

This example shows that if the total number of 2D NMR correlations is large, there is no need to involve as many correlations as possible, including those which have very low intensity. The benefit of taking into account several weak 2D NMR peaks to increase the total number of constraints can actually be problematic: if these correlations prove to be of nonstandard lengths then the problem becomes more difficult to solve and requires many additional assumptions.

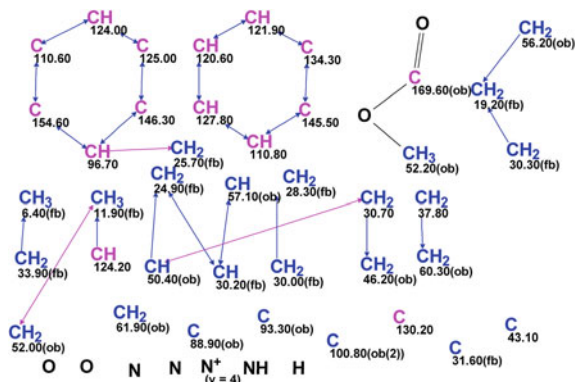
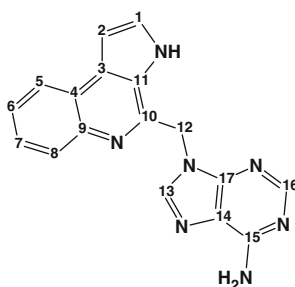


Fig. 5.93 Goniomedine A: The edited MCD. Connectivities of $2-3 J_{CH}$ lengths are hidden to make the picture more understandable. Connectivities of $2-4 J_{CH}$ lengths are marked with a *violet color*

5.31 Aplidiopsamine A

Carroll et al. [36] isolated a polyaromatic alkaloid, Aplidiopsamine A (**5.66**) from the temperate Australian ascidian, *Aplidiopsis confluenta*. Its structure was determined from the interpretation of the mass spectrometry, 1D and 2D NMR spectra. It is worthy noting that Aplidiopsamine A is the *first alkaloid* to possess the tricyclic aromatic substructure 3H-pyrrolo[2,3-c]quinoline conjugated to an adenine. The 3H-pyrrolo[2,3-c]quinoline ring system is *extremely rare* and has been reported only once before. Aplidiopsamine A therefore represents only the second example of a molecule containing this *unprecedented tricyclic ring system* and the spectroscopic properties of this ring system are reported for the first time in article [36].



5.66

Aplidiopsamine A was obtained as a yellow gum. A $[M+H]^+$ ion in the (+) HRESIMS at m/z 316.1316 (Δ 3.6 ppm) provided a molecular formula of C₁₇H₁₃N₇ with RDBE = 15. A high degree of unsaturation, as well as the ratio of skeletal to hydrogen atoms which is close to 2, make the problem challenging (see Sect. 1.2.2

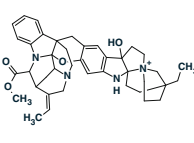
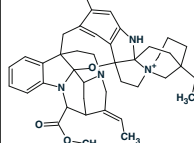
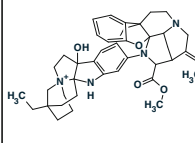
1	2	3
		
$d_A(^{13}\text{C}): 2.503$ $d_N(^{13}\text{C}): 2.509$ $d_I(^{13}\text{C}): 2.708$ $\text{max_}d_A(^{13}\text{C}): 9.740$	$d_A(^{13}\text{C}): 3.064$ $d_N(^{13}\text{C}): 2.661$ $d_I(^{13}\text{C}): 3.237$ $\text{max_}d_A(^{13}\text{C}): 10.400$	$d_A(^{13}\text{C}): 3.205$ $d_N(^{13}\text{C}): 3.352$ $d_I(^{13}\text{C}): 3.449$ $\text{max_}d_A(^{13}\text{C}): 12.080$

Fig. 5.94 Goniomedine A: The top structures of the ranked output file

and Ref. [37]). An absorption band at $3,485\text{ cm}^{-1}$ in the IR spectrum suggested that the molecule contained an amine functionality. The ^{13}C , ^1H and HSQC NMR spectroscopic data, as well as the data obtained from COSY and HMBC spectra are presented in Table 5.31.

The chemical shift assignment was carefully investigated by the authors [36]. They found that the hydrogen attached to carbon C2 (101.7) has the same chemical shift as the hydrogen atoms associated with the group NH_2 at 7.17 ppm. Reliably determined multiplicities and coupling constants which were used for setting additional constraints during the structure generation process (the numbers of hydrogens attached to neighbor atoms) are shown in the M(J) column of Table 5.31. The MCD mapping the data presented in Table 5.31 is shown in Fig. 5.95.

Table 5.31 Aplidiopsamine A: Spectroscopic NMR data

Label	δX	$\delta\text{C}_{\text{calc}}$	XH_n	δH	M(J)	COSY	C HMBC
C1	127.9	126.14	CH	7.72	u	12.33, 7.16	C3, C11, C2
C2	101.7	100.66	CH	7.17	u	7.72	C3, C4, C11, C1
C3	128.8	128.17	C	–	–	–	–
C4	123.3	119.59	C	–	–	–	–
C5	122.9	121.64	CH	8.22	d(8.2)	7.48	C9, C7, C3, C4
C6	126.1	126.72	CH	7.48	dd(7.1, 8.2)	8.22, 7.43	C4, C8
C7	126	128.24	CH	7.43	dd(7.1, 8.1)	7.48, 7.71	C5, C9
C8	128.9	125.44	CH	7.71	d(7.1)	7.43	C4, C9, C6
C9	141.8	140.12	C	–	–	–	–
C10	143.3	142.38	C	–	–	–	–
C11	126.5	130.09	C	–	–	–	–
C12	44.4	44.94	CH_2	5.90	S	–	C13, C11, C17, C10
C13	142.4	140.05	CH	8.28	S	–	C12, C17, C14
C14	118.3	119.43	C	–	–	–	–
C15	155.8	156.23	C	–	–	–	–
C16	153.7	152.98	CH	8.04	S	–	C14, C17, C15
C17	149.4	149.11	C	–	–	–	–
N1	100 ^a	–	NH	12.33	u	7.72	C2, C11, C3
N2	120 ^a	–	NH_2	7.17	u	–	C14

^a Fictitious ^{15}N chemical shifts

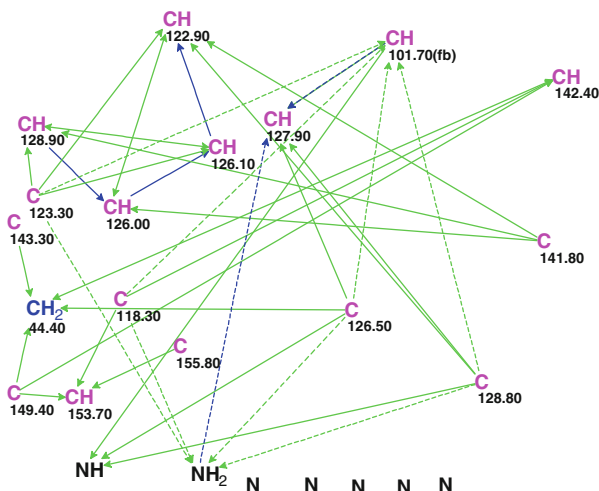
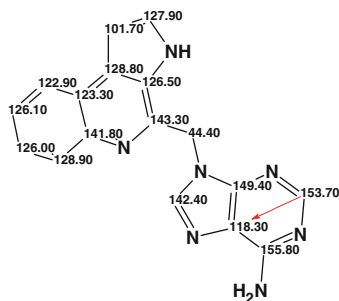


Fig. 5.95 Aplidiopsamine A: Molecular connectivity diagram

MCD overview Hybridization states were automatically assigned for all carbon atoms, while only one of them (C 101.7) was supplied with a label defining the possibility of the presence or absence of neighboring heteroatom (“fb”, forbidden, in the given case). The ambiguous connectivities (dotted lines) accounted for by two overlapped ^1H resonances at 7.17 ppm. The ^{13}C chemical shifts of sp^2 -hybridized carbons connected by a chemical bond with a nitrogen atom are known [18] to vary broadly. As the molecule contains seven nitrogen atoms no edits were applied to the MCD to prevent mistaken atom labeling.

No contradictions were detected in the 2D NMR data as a result of MCD checking; therefore Strict Structure Generation was initiated to give the results: $k = 295 \rightarrow 154 \rightarrow 142$, $t_g = 6$ s. ^{13}C NMR spectrum prediction for all members of the output file revealed that the best structure was characterized by deviation values of 5–7 ppm which indicates an erroneous solution. It was assumed that some contradictions not detected by the program existed in the COSY and/or HMBC data. In this case employing FSG can help to find the correct solution. The results of FSG at $m = 1$ and $a = 16$: $k = 13,729 \rightarrow 7,121 \rightarrow 5,483$, $t_g = 9$ min 08 s, 1 of 31 connectivities was extended, 31 of 31 connectivity combinations were processed. The three top similar structures of the ranked output file are presented in Fig. 5.96.

Figure 5.96 shows that the best structure is identical to the structure of Aplidiopsamine A and the ^{13}C chemical shift assignment performed by the authors [36] is the same as that found automatically by the program (see structure 5.67). The arrow denotes a nonstandard HMBC connectivity revealed during FSG.

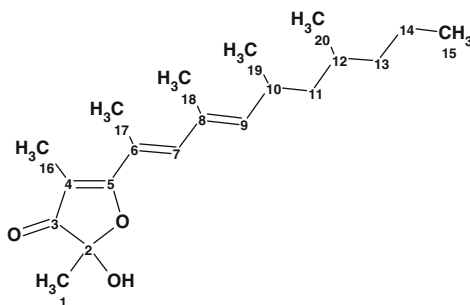


5.67

It is worth noting that structures #1 and #2 in Fig. 5.2 differ only by the position of the NH_2 group in the molecule, which markedly influences the calculated average deviation values.

5.32 Polypropionate

Marine mollusks of the genus *Siphonaria*, commonly known as false limpets, are shelled, air-breathing herbivores that are believed to have a marine ancestry. When disturbed, siphonariid limpets secrete a sticky white mucus from their lateral pedal glands. The mucus released by *Siphonaria* species is rich in polypropionate secondary metabolites. Colonies of *S. oculus* occur along the southern and eastern coast of South Africa and southern Mozambique. The *S. oculus* specimens were steeped in acetone, and the acetone extract was subjected to extensive chromatography to afford three new polypropionate metabolites [38]. Spectroscopic data of Polypropionate **5.68** are used here to describe employing CASE for the structure elucidation of this compound.



5.68

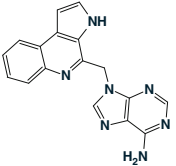
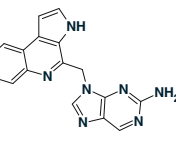
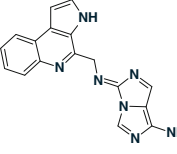
		
$d_A(^{13}\text{C})$: 1.550 $d_I(^{13}\text{C})$: 1.554 $d_N(^{13}\text{C})$: 1.755 $\text{max}_d d_A(^{13}\text{C})$: 3.710	$d_A(^{13}\text{C})$: 2.564 $d_I(^{13}\text{C})$: 2.268 $d_N(^{13}\text{C})$: 2.480 $\text{max}_d d_A(^{13}\text{C})$: 8.860	$d_A(^{13}\text{C})$: 3.962 $d_I(^{13}\text{C})$: 3.952 $d_N(^{13}\text{C})$: 4.348 $\text{max}_d d_A(^{13}\text{C})$: 15.860

Fig. 5.96 Aplidiopsamine A: Three top similar structures of the ranked output file

Polypropionate was isolated as a colorless oil, with a molecular formula established as $\text{C}_{20}\text{H}_{32}\text{O}_3$ from HRESIMS data (m/z 320.2350 [M^+] (calculated for $\text{C}_{20}\text{H}_{32}\text{O}_3$, 320.2351)), which implied five degrees of unsaturation. The characteristic bands at 3,339 and $1,686\text{ cm}^{-1}$ in the IR spectrum of **5.68** indicated the presence of both hydroxyl and carbonyl functionalities, respectively.

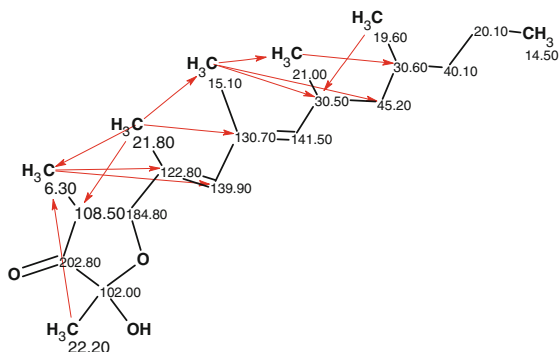
^1H , ^{13}C , HSQC, and HMBC NMR data tabulated in the [38] are presented in Table 5.32.

The MCD created from the NMR data is shown in Fig. 5.97.

MCD overview The two light blue atoms C 102.00 and C 108.5 can exist in the molecule either in sp^2 -hybridized or sp^3 states if the formation of the O–C–O substructure is possible. No MCD edits were made. The number of hydrogen atoms attached to the neighboring carbons were input in accordance with the ^1H multiplicities shown in column M(J) of Table 5.32.

MCD checking revealed the presence of at least 4 HMBC NSCs. FSG was therefore initiated with the options $m = 4\text{--}20$, $a = 16$, “Stop Generation when Structures Generated.” The options indicate that the expected number of nonstandard connectivities may be from 4 to 20 and any lengths for the NSCs are allowed. ^{13}C chemical shift prediction was carried out during structure generation to reject structures characterized by average ^{13}C chemical shift deviations $d > 4$ ppm. Results: $k = 6 \rightarrow 2 \rightarrow 1$, $t_g = 56$ min, 12 (!) from 43 connectivities have been extended and 18,440,500 from 15,338,678,264 (only 0.1 %) possible connectivity combinations were used.

The single output structure ($d_A = 1.12$, $d_N = 1.10$, $d_I = 1.45$ ppm) coincided with the structure of Polypropionate **5.68**. ^{13}C chemical shift assignment and all 12 nonstandard connectivities are displayed on structure **5.69**:



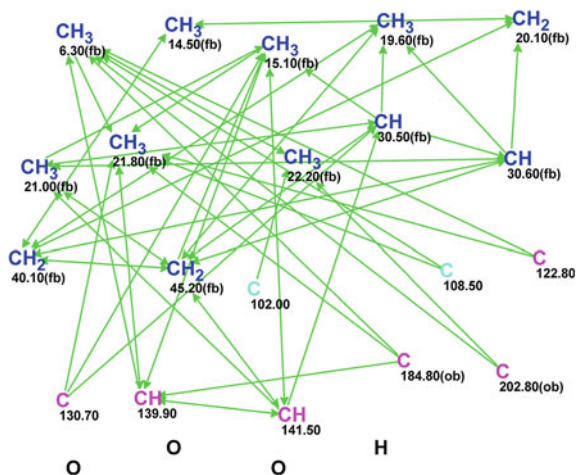
5.69

Table 5.32 Polypropionate: Spectroscopic NMR data

Label	δC	$\delta\text{C}_{\text{calc}}$	CH_n	δH	M(J)	C HMBC
C1	22.2	21.93	CH_3	1.55	S	C2, C3, C16
C2	102	102.15	C	–	–	–
C3	202.8	203.97	C	–	–	–
C4	108.5	108.42	C	–	–	–
C5	184.8	182.99	C	–	–	–
C6	122.8	125.8	C	–	–	–
C7	139.9	141.24	CH	6.26	d(1.2) ^a	C17, C18, C9, C5
C8	130.7	131.24	C	–	–	–
C9	141.5	144.02	CH	5.2	d(9.8)	C18, C7, C11
C10	30.5	32.19	CH	2.49	u	C9, C19, C11, C8
C11	45.2	45.21	CH_2	1.05	u	C13, C20, C19
C11	45.2	45.21	CH_2	1.21	u	C12, C10, C9
C12	30.6	30.54	CH	1.28	u	C13, C11, C10
C13	40.1	39.46	CH_2	1.06	u	C20
C13	40.1	39.46	CH_2	1.19	u	C15, C11, C12
C14	20.1	20.19	CH_2	1.29	u	C13, C12, C15
C14	20.1	20.19	CH_2	1.23	u	–
C15	14.5	14.17	CH_3	0.85	t(7.2)	C14, C13
C16	6.3	6.99	CH_3	1.52	S	C6, C4, C5, C3, C7
C17	21.8	15.19	CH_3	1.98	S	C5, C16, C7, C6, C4, C8, C18
C18	15.1	15.63	CH_3	1.69	S	C8, C7, C11, C19, C10, C9
C19	21	21.54	CH_3	0.87	d(6.5)	C12, C10, C11, C9

^a Small coupling constants can be interpreted as an indication of the absence of hydrogens attached to the nearest neighbor skeleton atoms

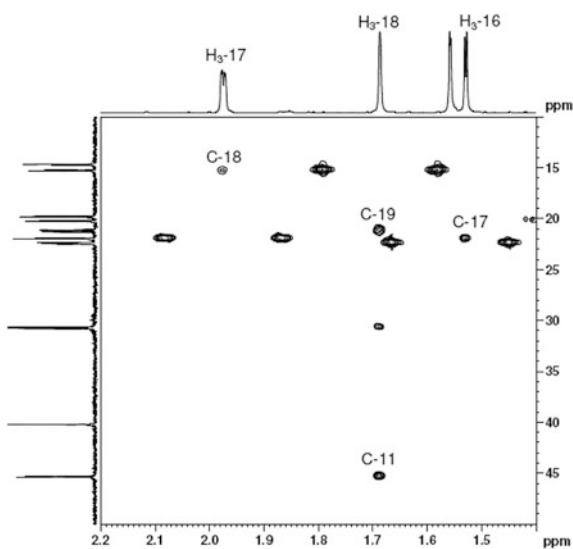
Fig. 5.97 Polypropionate:
The molecular connectivity diagram



Some interesting observations can be made when analyzing the solution obtained:

- The number of NSCs in the HMBC data turned out to be 12, and four of the nonstandard connectivities were of 4 C–C chemical bonds length (${}^5J_{\text{CH}}$) as shown in Fig. 5.98 taken from the Supporting Information associated with article [38]. Note that the peaks corresponding to these correlations are of intensity comparable with that common for correlations of a standard length (${}^{2-3}J_{\text{CH}}$).
- Only 0.1 % of the total number of all possible connectivity combinations was used in the process of structure generation as a result of an algorithm that is capable of selecting the most perspective connectivity combinations. Nevertheless the processor time consumed during the structure generation was about

Fig. 5.98 Polypropionate: A region (F1 = δ 1.4–2.2 ppm; F2 = δ 10–50 ppm) of the HMBC spectrum (CDCl₃, 600 MHz) of structure 5.68 illustrating the following five bond HMBC correlations: H17 to C18, H18 to C11 and C19, and H16 to C17



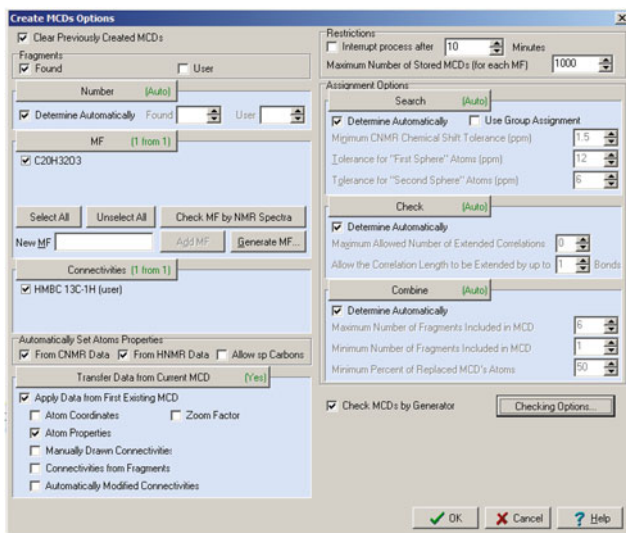


Fig. 5.99 Polypropionate: Options of MCD creation

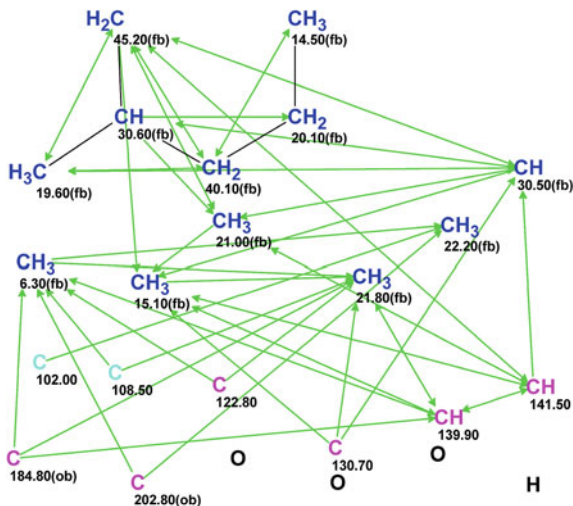
an hour. An approximation suggest that trying all ca. 15 billion possible connectivity combinations would need about 40 days of processor time, i.e., the solution of the problem would become practically impossible.

- iii. In spite of the great number of NSCs with different lengths an unambiguous solution to the problem was found automatically.
- iv. The most advanced mode of FSG, where no restriction is imposed on the lengths of nonstandard connectivities ($a = 16$), allowed us to solve the problem under the condition where the HMBC data contained 8 NSCs of ${}^4J_{\text{CH}}$ type and 4 NSCs of ${}^5J_{\text{CH}}$ type. Note that the real number of NSCs and their lengths became known only after the problem has been solved.

The problem was solved but FSG consumed about an hour of processor time. We could expect that application of the Fragment Library could help to reduce this time. For this goal a fragment search of the ${}^{13}\text{C}$ chemical shifts was performed on the Fragment Library which was completed with the detection of 73 fragments whose subspectra are “projected” onto the experimental ${}^{13}\text{C}$ NMR spectrum.

The command **Structure Elucidation/Create MCD Using Fragments** was then initiated with the Options presented in Fig. 5.99. All options influencing the MCD creation were set in such a manner to execute all procedures automatically. As a result only one MCD was created in which one found fragment was “implanted” (Fig. 5.100).

Fig. 5.100 Polypropionate:
The MCD containing one
Found Fragment

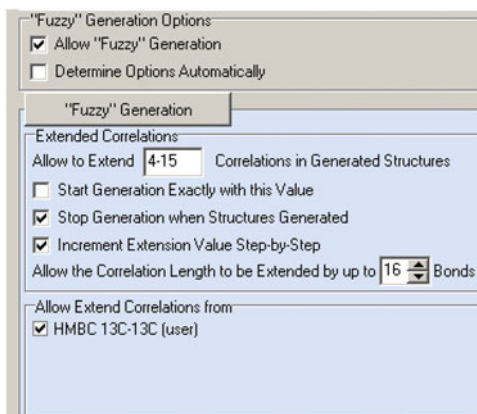


FSG was initiated again with the options shown in Fig. 5.101.

The following results were obtained: $k = 6 \rightarrow 6 \rightarrow 3$, $t_g = 52$ s, 12 from 33 connectivities have been extended during generation, 500,900 from 354,817,320 (0.1 %) possible connectivity combinations have been used during generation. The ranked output file is presented in Fig. 5.102.

As Fig. 5.102 shows the correct structure was ranked first and the competing structures can be rejected with a high reliability due to the large difference between deviations. At the same time the compute time reduced from 56 min down to 52 s, i.e., more than 60 times.

Fig. 5.101 Polypropionate:
Options for FSG from the
MCD created using Found
Fragments



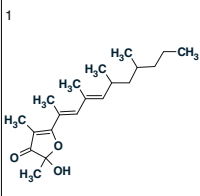
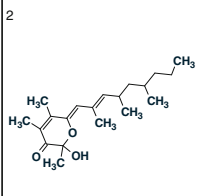
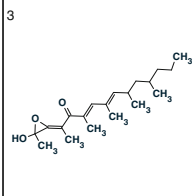
		
$d_A(^{13}\text{C}): 1.116$ $d_N(^{13}\text{C}): 1.094$ $d_I(^{13}\text{C}): 1.450$ $\text{max_}d_A(^{13}\text{C}): 6.610$	$d_A(^{13}\text{C}): 5.649$ $d_N(^{13}\text{C}): 5.809$ $d_I(^{13}\text{C}): 6.159$ $\text{max_}d_A(^{13}\text{C}): 33.740$	$d_A(^{13}\text{C}): 6.379$ $d_N(^{13}\text{C}): 5.917$ $d_I(^{13}\text{C}): 5.566$ $\text{max_}d_A(^{13}\text{C}): 57.870$

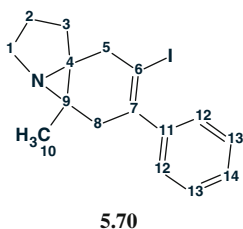
Fig. 5.102 Polypropionate: The ranked structural output file obtained using Found Fragments

5.33 Aziridine

Kummerlöwe and co-workers [39] investigated one of the products obtained by reacting an azide-containing 1,5-enyne in the presence of electrophilic iodine sources. Initially the researchers tried to elucidate the structure of this new compound using classical methods commonly employed in such cases. High resolution mass spectrometry unambiguously provided the molecular formula for the unknown: $\text{C}_{16}\text{H}_{18}\text{NI}$, $m/z = 351.0486$ [351.0484 calculated for $\text{C}_{16}\text{H}_{18}\text{NI}$ (M^+)]. The following spectroscopy data were initially acquired at the first stage of the investigation: IR spectrum, 1D ^1H and ^{13}C spectra in combination with two-dimensional COSY, HSQC, ^1H - ^{13}C HMBC, and ^1H - ^{15}N HMBC experiments. Eleven fragments were identified from the data: a phenyl group, a methyl group, five methylene groups (three forming an isolated chain), a tertiary nitrogen atom, an iodine atom, and four quaternary carbon atoms. The ^1H - ^{13}C HMBC spectrum revealed 63 long-range correlations and the ^1H - ^{15}N HMBC spectrum exposed seven cross peaks thereby correlating almost every fragment with every other fragment and indicating a very compact structure. Because it was difficult to deduce the structure from these data, a 2D 1,1-ADEQUATE spectrum was also recorded on a Bruker Avance 900 MHz spectrometer equipped with a 5 mm cryogenically cooled TXI probe head optimized for proton detection. The 1,1-ADEQUATE data did identify the adjacent quaternary carbons unequivocally. While this was useful information this additional data did not help to elucidate the structure.

Since classical NMR analysis failed the authors decided to make an attempt to solve the problem in an unconventional way by using residual dipolar couplings (RDCs) [40]. In accordance with the methodology associated with RDC they assumed that as long as sufficient anisotropic parameters can be measured, and a large enough set of structural models can be constructed, it should be possible to identify the correct chemical structure.

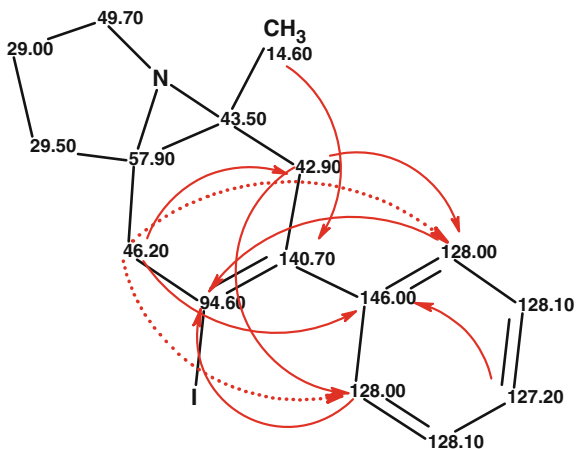
In order to measure the RDCs the compound was aligned in a stretched polystyrene/chloroform gel. The corresponding scalar couplings were measured in a chloroform solution sample. Fourteen proposed structures were tested using the experimental data. Analysis of the RDC data suggested that the aziridine structure **5.70** (7-iodo-10-methyl-8-phenyldihydro-3*H*-2,1,2-(pent[2]ene[1, 5, 5]triy)pyrrolidine) is the correct one.



To confirm structure **5.70** suggested by the RDC data almost 100 mg of the reaction product was synthesized and a 2D INADEQUATE spectrum was acquired using 3 days of spectrometer time. The structure **5.70** elucidated using the RDC data was unambiguously confirmed by the INADEQUATE data. In addition, labeling the starting material of the reaction with ^{15}N -azide and measuring ^{13}C - ^{15}N couplings for the ^{15}N -labeled compound was performed. Both additional experiments clearly supported structure **5.70**.

Retrospective data analysis showed that the ^1H - ^{13}C HMBC spectra contained nine NSCs, (those having $^nJ_{\text{HC}}$, $n > 3$). This is not surprising considering that the molecule is a highly rigid system. Structure Elucidator interprets the combinations of the available ^1H - ^1H and ^1H - ^{13}C correlations to derive carbon-carbon connectivities to produce nine nonstandard C to C connectivities (see Fig. 5.103), which was the main issue preventing structure elucidation using a traditional approach. The initial system of “axioms” used for the structure elucidation from the HMBC

Fig. 5.103 Aziridine: Structure **5.70** with arrows showing the nonstandard HMBC connectivities. The connectivities 46.2–128.0 marked by the dotted lines correspond to $^5J_{\text{CH}}$



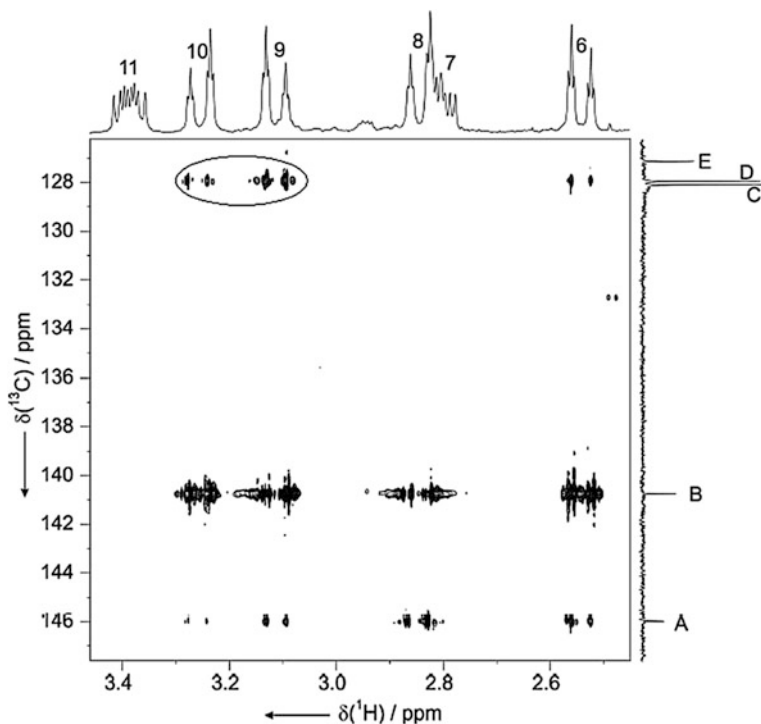


Fig. 5.104 Aziridine: A selected region of the ^1H - ^{13}C HMBC spectrum showing the unusually intense signals between protons 9 (3.12 ppm) and 10 (3.26 ppm) of the methylene group and the carbon atoms D (128.0 ppm) in the *ortho* positions of the phenyl-group (marked by an *ellipse*)

data became extremely contradictory due to the presence of nonstandard connectivities. Moreover, two unexpected *intense* $^5J_{\text{CH}}$ cross peaks correlating two protons with the *ortho*-carbons of the phenyl group (see Fig. 5.103) were identified in the ^1H - ^{13}C HMBC spectrum. The corresponding part of the HMBC spectrum is presented in Fig. 5.104 taken from the Supporting Information of Ref. [39].

We suggest that this can be explained as a result of the hindered rotation of the phenyl group due to the large volume of the iodine atom.

At the same time the authors [39] found that structure **5.70** was almost certainly excluded from the potential set of structures because the ^{13}C chemical shifts predicted by ChemDraw [41] differed significantly from the experimental data.

As mentioned, the highly complex nature of the 2D NMR data prompted the authors [39] to conclude that the problem could not be solved using a classical approach. In making this decision they only considered the NMR data in isolation from algorithmic-assisted CASE approach.

The experimental data presented in the work [39] (Table 5.33) were therefore analyzed by us using Structure Elucidator [42].

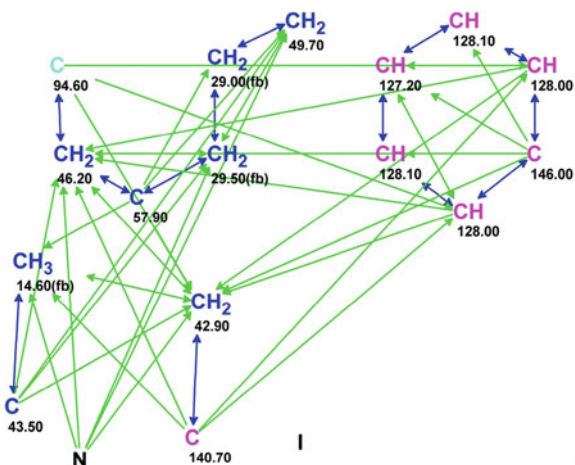
Table 5.33 Aziridine: Spectroscopic NMR data

Label	δC	δC_{calc}	XH_n	δH	M(J)	C HMBC	^{15}N HMBC
C1	49.7	44.87	CH ₂	2.81	u	C4, C9, C2, C3	–
C1	49.7	44.87	CH ₂	3.39	u	C2, C9, C3	N1
C2	29	21.32	CH ₂	1.74	u	C3, C1, C4	–
C2	29	21.32	CH ₂	2.12	u	C1, C3, C4	–
C3	29.5	29.17	CH ₂	2.07	u	C4, C2, C5, C1, C9	N1
C3	29.5	29.17	CH ₂	1.89	u	C4, C9, C5, C1, C2	–
C4	57.9	51.12	C	–	–	–	–
C5	46.2	44.29	CH ₂	3.26	u	C11, C7, C4, C9, C3, C6, C12	N1
C5	46.2	44.29	CH ₂	3.12	u	C6, C12, C11, C3, C7, C4, C8, C9	N1
C6	94.6	93.77	C	–	–	–	–
C7	140.7	144.4	C	–	–	–	–
C8	42.9	34.37	CH ₂	2.54	u	C7, C11, C6, C4, C10, C12, C5, C9	N1
C8	42.9	34.37	CH ₂	2.86	u	C7, C11, C6, C9, C4	N1
C9	43.5	39.99	C	–	–	–	–
C10	14.6	20.06	CH ₃	1.17	u	C8, C9, C4, C7	N1
C11	146	141.96	C	–	–	–	–
C12	128	126.4	CH	7.18	u	C14, C6, C7, C13	–
C13	128.1	127.84	CH	7.35	u	C11, C12	–
C14	127.2	127.4	CH	7.28	u	C12, C11	–
N1	100 ^a	–	N	–	–	–	–

^a Fictitious ^{15}N chemical shift

The molecular formula, 1D ^{13}C and 1H , HSQC, 1H - ^{13}C HMBC, 1H - ^{15}N HMBC and 1,1-ADEQUATE spectra were input into the program and MCD was created (Fig. 5.105).

Fig. 5.105 Aziridine:
The molecular connectivity diagram. 1,1-ADEQUATE connectivities drawn on MCD by hand are denoted by *blue bold lines*



FSG was run with the following result: only one correct structure, **5.70**, was generated in 0.7 s. The application of a CASE approach therefore allowed us to instantly and unambiguously find the single correct structure from the HMBC and 1,1-ADEQUATE data. It has now been shown a number of times that 1,1-ADEQUATE data in conjunction with other 2D NMR data is a very valuable data combination as input for CASE programs.

References

1. Thongbai B, Surup F, Mohr K, Kuhnert E, Hyde KD, Stadler M (2013) *Gymnopalynes A* and *B*, chloropropynyl-isocoumarin antibiotics from cultures of the basidiomycete *Gymnopus* sp. *J Natl Prod* 76(11):2141–2144. doi:[10.1021/np400609f](https://doi.org/10.1021/np400609f)
2. Miao FP, Liang XR, Yin XL, Wang G, Ji NY (2012) Absolute configurations of unique harziane diterpenes from *Trichoderma* species. *Org Lett* 14(15):3815–3817. doi:[10.1021/ol3014717](https://doi.org/10.1021/ol3014717)
3. Sikorska J, Hau AM, Anklin C, Parker-Nance S, Davies-Coleman MT, Ishmael JE, McPhail KL (2012) Mandelalides A–D, cytotoxic macrolides from a new *Lissoclinum* species of South African tunicate. *J Org Chem* 77(14):6066–6075. doi:[10.1021/jo3008622](https://doi.org/10.1021/jo3008622)
4. Mu ZQ, Gao H, Huang ZY, Feng XL, Yao XS (2012) Puberunine and puberudine, two new C18-diterpenoid alkaloids from *Aconitum barbatum* var. *puberulum*. *Org Lett* 14(11):2758–2761. doi:[10.1021/ol3008217](https://doi.org/10.1021/ol3008217)
5. Zou J, Du X, Pang G, Shi YM, Wang WG, Zhan R, Kong LM, Li XN, Li Y, Pu JX, Sun HD (2012) Ternifolide A, a new diterpenoid possessing a rare macrolide motif from *Isodon ternifolius*. *Org Lett* 14(12):3210–3213. doi:[10.1021/ol3013205](https://doi.org/10.1021/ol3013205)
6. Wu J, Tokuyama S, Nagai K, Yasuda N, Noguchi K, Matsumoto T, Hirai H, Kawagishi H (2012) Strophasterols A to D with an unprecedented steroid skeleton: from the mushroom *Stropharia rugosoannulata*. *Angew Chem Int Ed Engl* 51(43):10820–10822. doi:[10.1002/anie.201205351](https://doi.org/10.1002/anie.201205351)
7. Cai JY, Zhang Y, Luo SH, Chen DZ, Tang GH, Yuan CM, Di YT, Li SH, Hao XJ, He HP (2012) Aphanamixoid A, a potent defensive limonoid, with a new carbon skeleton from *Aphanamixis polystachya*. *Org Lett* 14(10):2524–2527. doi:[10.1021/ol3008149](https://doi.org/10.1021/ol3008149)
8. Meng FY, Sun JX, Li X, Yu HY, Li SM, Ruan HL (2011) Schiglautone A, a new tricyclic triterpenoid with a unique 6/7/9-fused skeleton from the stems of *Schisandra glaucescens*. *Org Lett* 13(6):1502–1505. doi:[10.1021/ol200188n](https://doi.org/10.1021/ol200188n)
9. Wang LN, Zhang JZ, Li X, Wang XN, Xie CF, Zhou JC, Lou HX (2012) Pallambins A and B, unprecedented hexacyclic 19-nor-secolabdane diterpenoids from the Chinese liverwort *Pallavicinia ambigua*. *Org Lett* 14(4):1102–1105. doi:[10.1021/ol3000124](https://doi.org/10.1021/ol3000124)
10. Ishigami ST, Goto Y, Inoue N, Kawazu S, Matsumoto Y, Imahara Y, Tarumi M, Nakai H, Fusetani N, Nakao Y (2012) Cristaxenicin A, an antiprotozoal xenicane diterpenoid from the deep sea gorgonian *Acanthoprimnoa cristata*. *J Org Chem* 77(23):10962–10966. doi:[10.1021/jo302109g](https://doi.org/10.1021/jo302109g)
11. Raju R, Gromyko O, Fedorenko V, Luzhetskyy A, Plaza A, Muller R (2012) Juniperolide A: a new polyketide isolated from a terrestrial actinomycete, *Streptomyces* sp. *Org Lett* 14(23):5860–5863. doi:[10.1021/ol302766z](https://doi.org/10.1021/ol302766z)
12. Vansteelandt M, Blanchet E, Egorov M, Petit F, Toupet L, Bondon A, Monteau F, Le Bizec B, Thomas OP, Pouchus YF, Le Bot R, Grovel O (2013) Ligerin, an antiproliferative chlorinated sesquiterpenoid from a marine-derived *Penicillium* strain. *J Nat Prod* 76(2):297–301. doi:[10.1021/np3007364](https://doi.org/10.1021/np3007364)

13. Kim H, Chin J, Choi H, Baek K, Lee TG, Park SE, Wang W, Hahn D, Yang I, Lee J, Mun B, Ekins M, Nam SJ, Kang H (2013) Phosphoiodyns A and B, unique phosphorus-containing iodinated polyacetylenes from a Korean sponge *Placospongia* sp. *Org Lett* 15(1):100–103. doi:[10.1021/ol3031318](https://doi.org/10.1021/ol3031318)
14. Kim H, Chin J, Choi H, Baek K, Lee TG, Park SE, Wang W, Hahn D, Yang I, Lee J, Mun B, Ekins M, Nam SJ, Kang H (2013) Phosphoiodyns A and B, unique phosphorus-containing iodinated polyacetylenes from a Korean sponge *Placospongia* sp (Correction). *Org Lett* 15(21):5614–5615. doi:[10.1021/ol402803k](https://doi.org/10.1021/ol402803k)
15. Whitson EL, Sun H, Thomas CL, Henrich CJ, Sayers TJ, McMahon JB, Griesinger C, McKee TC (2012) Synergistic TRAIL sensitizers from *Barleria alluaudii* and *Diospyros maritima*. *J Nat Prod* 75(3):394–399. doi:[10.1021/np200805z](https://doi.org/10.1021/np200805z)
16. Plaza A, Garcia R, Bifulco G, Martinez JP, Huttel S, Sasse F, Meyerhans A, Stadler M, Muller R (2012) Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus “Aetherobacter”. *Org Lett* 14(11):2854–2857. doi:[10.1021/ol3011002](https://doi.org/10.1021/ol3011002)
17. Ohlendorf B, Schulz D, Erhard A, Nagel K, Imhoff JF (2012) Geranylphenazinediol, an acetylcholinesterase inhibitor produced by a *Streptomyces* species. *J Nat Prod* 75(7):1400–1404. doi:[10.1021/np2009626](https://doi.org/10.1021/np2009626)
18. Pretsch E, Bühlmann P, Affolter C (2000) Structure determination of organic compounds - tables of spectral data. Springer, Berlin
19. Fu Y, Zhang Y, He H, Hou L, Di Y, Li S, Luo X, Hao X (2012) Strynuxlines A and B, alkaloids with an unprecedented carbon skeleton from *Strychnos nux-vomica*. *J Nat Prod* 75(11):1987–1990. doi:[10.1021/np300339r](https://doi.org/10.1021/np300339r)
20. Okanya PW, Mohr KI, Gerth K, Steinmetz H, Huch V, Jansen R, Muller R (2012) Hyaladione, an S-methyl cyclohexadiene-dione from *Hyalangium minutum*. *J Nat Prod* 75(4):768–770. doi:[10.1021/np200776v](https://doi.org/10.1021/np200776v)
21. Elyashberg ME, Williams AJ, Blinov KA (2012) Contemporary computer-assisted approaches to molecular structure elucidation. *New Developmnts in NMR*, vol 1. RSC Publishing, Cambridge
22. Pettit GR, Tan R, Bao GH, Melody N, Doubek DL, Gao S, Chapuis JC, Williams L (2012) Antineoplastic agents. 587. Isolation and structure of 3-epipancratistatin from *Narcissus* cv. Ice Follies. *J Nat Prod* 75(4):771–773. doi:[10.1021/np200862y](https://doi.org/10.1021/np200862y)
23. Hu Y, Legako AG, Espindola AP, MacMillan JB (2012) Erythrolic acids A-E, meroterpenoids from a marine-derived *Erythrobacter* sp. *J Org Chem* 77(7):3401–3407. doi:[10.1021/jo300197z](https://doi.org/10.1021/jo300197z)
24. Ma C, Li Y, Niu S, Zhang H, Liu X, Che Y (2011) N-hydroxypyridones, phenylhydrazones, and a quinoxalinone from *Isaria farinosa*. *J Nat Prod* 74(1):32–37. doi:[10.1021/np100568w](https://doi.org/10.1021/np100568w)
25. Jin Z, Mashuta MS, Stolowich NJ, Vaisberg AJ, Stivers NS, Bates PJ, Lewis WH, Hammond GB (2012) Physangulidines A, B, and C: three new antiproliferative withanolides from *Physalis angulata* L. *Org Lett* 14(5):1230–1233. doi:[10.1021/ol203498a](https://doi.org/10.1021/ol203498a)
26. Lee SU, Asami Y, Lee D, Jang JH, Ahn JS, Oh H (2011) Protuboxepins A and B and protubonines A and B from the marine-derived fungus *Aspergillus* sp. SF-5044. *J Nat Prod* 74(5):1284–1287. doi:[10.1021/np100880b](https://doi.org/10.1021/np100880b)
27. Wang XC, Zheng ZP, Gan XW, Hu LH (2009) Jatrophalactam, a novel diterpenoid lactam isolated from *Jatropha curcas*. *Org Lett* 11(23):5522–5524. doi:[10.1021/ol902349f](https://doi.org/10.1021/ol902349f)
28. Wright AD, Schupp PJ, Schror JP, Engemann A, Rohde S, Kelman D, de Voogd N, Carroll A, Motti CA (2012) Twilight zone sponges from Guam yield theonellin isocyanate and psammaphysins I and J. *J Nat Prod* 75(3):502–506. doi:[10.1021/np200939d](https://doi.org/10.1021/np200939d)
29. El Aouad N, Perez-Moreno G, Sanchez P, Cantizani J, Ortiz-Lopez FJ, Martin J, Gonzalez-Menendez V, Ruiz-Perez LM, Gonzalez-Pacanowska D, Vicente F, Bills G, Reyes F (2012) Lasionectrin, a naphthopyrone from a *Lasionectria* sp. *J Nat Prod* 75(6):1228–1230. doi:[10.1021/np3002942](https://doi.org/10.1021/np3002942)
30. Cimmino A, Andolfi A, Zonno MC, Troise C, Santini A, Tuzi A, Vurro M, Ash G, Evidente A (2012) Phomentrioloxin: a phytotoxic pentasubstituted geranylcylohexentriol produced by

- Phomopsis* sp., a potential mycoherbicide for *Carthamus lanatus* biocontrol. J Nat Prod 75 (6):1130–1137. doi:[10.1021/np300200j](https://doi.org/10.1021/np300200j)
31. Nakanishi K, Solomon PH (1977) Infrared absorption spectroscopy. Holden Day, Oakland
 32. Scott AI (1964) Interpretation of ultraviolet spectra of natural products. Pergamon Press Ltd, Oxford
 33. Feng Y, Wang L, Niu S, Li L, Si Y, Liu X, Che Y (2012) Naphthalenones from a *Perenniporia* sp. inhabiting the larva of a phytophagous weevil, *Euops chinensis*. J Nat Prod 75 (7):1339–1345. doi:[10.1021/np300263u](https://doi.org/10.1021/np300263u)
 34. Zhang W, Liu Z, Li S, Yang T, Zhang Q, Ma L, Tian X, Zhang H, Huang C, Zhang S, Ju J, Shen Y, Zhang C (2012) Spiroindimicins A–D: new bisindole alkaloids from a deep-sea-derived actinomycete. Org Lett 14(13):3364–3367. doi:[10.1021/ol301343n](https://doi.org/10.1021/ol301343n)
 35. Benididir MA, Martin M-T, Tran Huu Dau M-E, Grellier P, Rasoanaivo P, Guéritte F, Litaudon M (2012) Goniomedines A and B: unprecedented bisindole alkaloids formed through the fusion of two indole moieties via a dihydropyran unit. Org Lett 14(16):4162–4165. doi:[10.1021/ol301832t](https://doi.org/10.1021/ol301832t)
 36. Carroll AR, Duffy S, Avery VM (2010) Aplidiopsamine A, an antiplasmodial alkaloid from the temperate Australian ascidian, *Aplidiopsis confluata*. J Org Chem 75(23):8291–8294. doi:[10.1021/jo101695v](https://doi.org/10.1021/jo101695v)
 37. Crews P, Rodríguez J, Jaspars M (2010) Organic structure analysis. Oxford University Press, Oxford
 38. Bromley CL, Popplewell WL, Pinchuck SC, Hodgson AN, Davies-Coleman MT (2012) Polypropionates from the South African marine mollusk *Siphonaria oculus*. J Nat Prod 75 (3):497–501. doi:[10.1021/np2009384](https://doi.org/10.1021/np2009384)
 39. Kummerlowe G, Crone B, Kretschmer M, Kirsch SF, Luy B (2011) Residual dipolar couplings as a powerful tool for constitutional analysis: the unexpected formation of tricyclic compounds. Angew Chem Int Ed Engl 50(11):2643–2645. doi:[10.1002/anie.201007305](https://doi.org/10.1002/anie.201007305)
 40. Kummerlowe G, Schmidt S, Luy B (2010) Cross-fitting of residual dipolar couplings. Open Spectroscopy J 4:16–27
 41. Cambridge Soft Corporation, CS Chem Draw PRO
 42. Elyashberg ME, Blinov KA, Molodtsov SG, Williams AJ (2012) Elucidating “undecipherable” chemical structures using computer assisted structure elucidation approaches. Magn Reson Chem 50:22–27. doi:[10.1002/mrc.2849](https://doi.org/10.1002/mrc.2849)

Glossary

Average Deviation The sum of deviations of the absolute values calculated for each atom divided by the number of atoms for which chemical shifts were calculated.

AutoMCD An MCD which is automatically created as the Table of Spectral Data is filled in. The AutoMCD displays the initial (not edited) data. It can also be edited by the user in a manner intrinsic to the same manner used for editing the User MCD.

Atom Properties Atom hybridization (sp , sp^2 , sp^3), valence, possibility of having a heteroatom as a neighbor—*ob* (obligatory) and *fb* (forbidden), total number of hydrogen atoms attached to neighbor atoms $n(\text{H})$, charge on atom.

Ambiguous Connectivities Connectivities whose lengths can not be specified due to signal overlapping in the NMR spectra.

ACD/ChemSketch A chemical structure editor developed by ACD/Labs. The program is adapted to operating together with Structure Elucidator and activated simultaneously when the program is started. All subcommands of the **File>Create Report** submenu available in the Structure Elucidator window delivers reports, tables, structural lists, stick spectra, etc. which appear in the ChemSketch window.

Complex Deviation A match factor that takes into account both ^{13}C and ^1H average chemical shift deviations. It is calculated for deviations obtained when chemical shift prediction is carried out by using a neural networks based algorithm, $d_{\text{complex}} = d_N(^{13}\text{C}) + 10d_N(^1\text{H})$.

Characteristic Spectral Feature Spectral features which fall into definite intervals characteristic of the specified atoms and fragments.

Exact Search In Search for a given chemical structure against a structural file. For instance, search for Proposed Structure in the file of generated structures.

Fuzzy Structure Generation Structure generation from 2D NMR data under the condition that the data set contains an unknown number of nonstandard correlations (connectivities) of unknown lengths.

List of Structures Visualization of a structural file in the form of a list where n ($n \geq 2$) structures along with their associated properties can be simultaneously observed in the program window. This form of structure visualization is very convenient for the analysis of ranked structural files.

Molecular Connectivity Diagram (MCD) A dialog window where the initial spectrum-structural information is presented visually and can be edited by the user. The MCD displays all atoms along with their properties and a full set of connectivities produced from the 2D NMR data.

Marking Accuracy of Spectrum Prediction This procedure supplies all atoms of a given structure with colored circles whose color allows for a visual estimate of the accuracy of chemical shift prediction for each atom.

Merged Chemical Shifts A set of ^{13}C or ^1H chemical shifts extracted from the available 1D and 2D NMR spectra.

Nonstandard Correlations Correlations corresponding to $^nJ_{\text{HH}}$ in COSY and $^nJ_{\text{CH}}$ in HMBC, when $n > 3$.

Nonstandard Connectivities Connectivities *between skeletal atoms* having a length of more than one chemical bond in COSY and more than two chemical bonds in HMBC.

Project A file containing all information related to a given problem. If the user has n alternative structural hypotheses he may create n User MCDs and then input each hypothesis in the corresponding MCD. The results of problem solving with n MCDs can be saved either as n different projects or as a joined file.

Project from Structure A project created from a structure. The User MCD associated with this project contains *all* theoretically possible connectivities in the 2D NMR spectra which were selected by the user and are intrinsic to the theoretically possible connectivities in the 2D NMR spectra which were selected by the user. Structure generation from this MCD allows for the determination of all structures that satisfy the produced connectivities.

Proposed Structure (PM) A structure hypothesized by the chemist. It is placed in the **PM** window of Structure Elucidator. An **Exact Search** of the Proposed Structure (**PM** window menu) in the Ranked Output File allows the user to determine the position of the Proposed Structure in the Ranked File.

Ranked Output File An output structure file which was ranked in ascending order of calculated average deviations between experimental and predicted chemical shifts. It can also be ranked by other parameters chosen by the user.

Standard 2D NMR Correlations Correlations corresponding to $^3J_{\text{HH}}$ in COSY and $^{2-3}J_{\text{CH}}$ in HMBC.

Standard 2D NMR Connectivities Connectivities *between skeletal atoms* having a length of one chemical bond in COSY and 1–2 chemical bonds in the HMBC.

Spectral Filter A set of tables containing CH_n ($n = 0-3$) atoms in different environments as well as molecular fragments along with the corresponding intervals of characteristic spectral features (^{13}C and ^1H NMR chemical shifts). The filter is used for elimination of generated structures containing atoms and fragments which contradict the experimental chemical shifts observed in the ^{13}C and ^1H NMR spectra.

Structural Filter A set of libraries containing fragments which must be either obligatory present or obligatory absent into each generated structure. The filter is used for elimination of structures uncommon in organic chemistry and for storing on the disc only those structures that satisfy the user requirements.

Structure Generator A program which generates all isomers corresponding to the given molecular formula and imposed structural constraints.

Strict Structure Generation Structure generation from 2D NMR data under the condition that all HMBC and COSY correlations (connectivities) are of standard length.

Spectrum Parameters Parameters of the NMR spectrometer used for spectrum acquisition, as well those parameters assigned by the user for the structure elucidation process. The *tolerance* for F1 and F2 axes was set to 0.005 ppm for all problems described in this book. An increase of the tolerance leads to the appearance of ambiguous connectivities or an increase in their number, which results in enlarging the size of the output file and the time associated with structure generation.

Substructure Search In Search for a given fragment (substructure) in a structural file or structure. For instance, search for a Found Fragment in the file of generated structures.

Similarity Search In This procedure carries out a search of structures similar to the given one. It is completed by ranking the structural file in descending order of the similarity coefficient value.

User MCD Provides an MCD which can be edited by the scientist to verify the different sets of axioms mapping to the initial information. For alternative sets of axioms, a set of MCDs can be created.

User GoodList List of obligatory fragments set by the user. The list is used by the program during structure generation.

User BadList List of forbidden fragments set by the user. The list is used by the program during structure generation.