

The Case for CASE: Computer-Assisted Structure Elucidation

Introduction

Computer-Assisted Structure Elucidation (CASE) has come to be a broadly accepted method for the derivation of novel or difficult chemical structures, especially those of natural products or drug metabolites and impurities. For completely error-free structure determination, one must first evaluate all isomers that truly match connectivity criteria defined by experiment – but it is impossible to do this manually without bias.

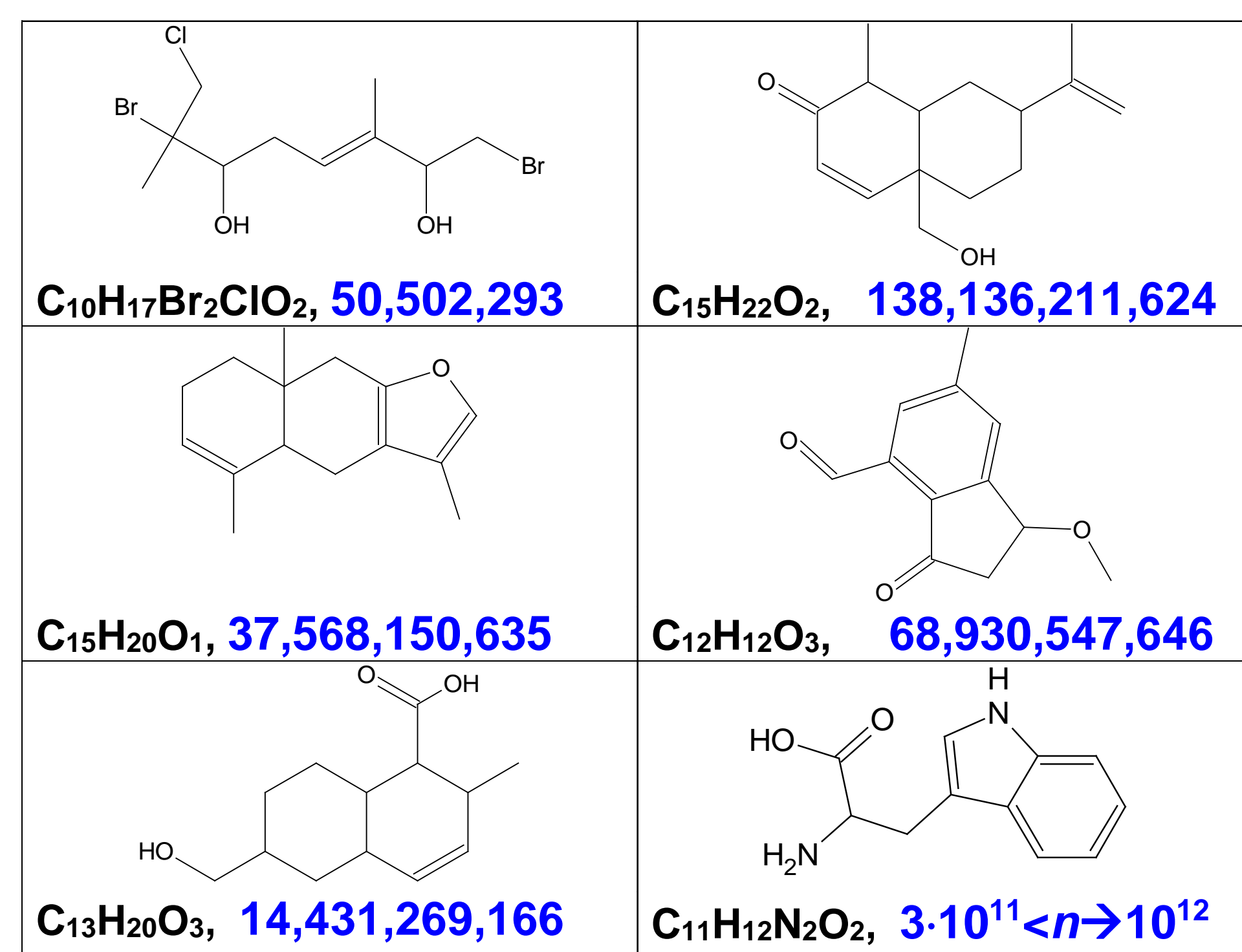


Figure 1: The large number of possible isomers for some relatively small organic molecules.

Modern NMR techniques give ever-increasing amounts of valuable data to substantiate structural evaluations, including lower limits of detection, and more information on connectivities. However, errors in published structures are rampant [1] giving weight to the argument for computer-assisted evaluation of structures. [2]

CASE Methodology

CASE follows a simple set of steps to comprehensively evaluate structures against available data:

1. Interpret experimental data to extract knowledge:
 - Molecular Formula
 - Integrals
 - Chemical shifts
 - Multiplicities
 - Connectivities
 - Known fragments
 - Known exclusions
2. Search structure space to derive all possible structures
3. Rank-order based on set criteria
 - Predicted chemical shift

Data Interpretation

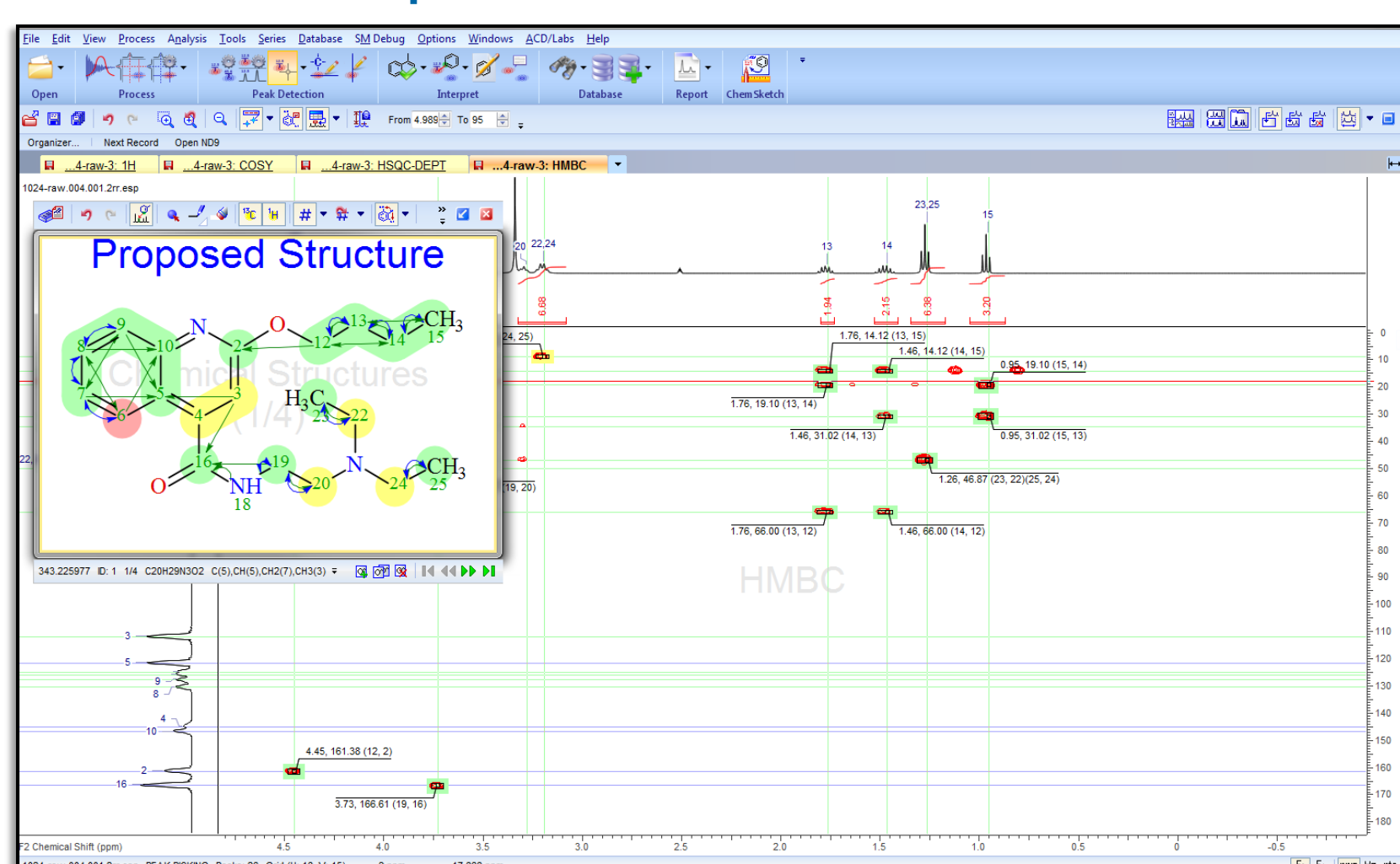


Figure 2: In ACD/Labs NMR software, peak assignments are synchronized throughout a set of related NMR experiments with NMR Sync. During interpretation, the software displays correlations for assigned spectra and structures, highlighting those which are likely to be erroneous.

References

1. Nicolaou, KC; Snyder, SA. *Angew. Chem. Int. Ed. Engl.*, **2005**, Feb 4:44(7); 1012-44.
2. Elyashberg, ME; Williams, AJ; Blinov, KA. *Nat. Prod. Rep.*, **2010**, 27(9): 1296-1328.
3. Moser, A; Elyashberg, ME; Williams, AJ; Blinov, KA; DiMartino, JC. *J. Cheminf.*, **2012** 4:5.
4. Ravu, RR, et al. *J. Nat Prod.*, **2015**, 78(4): 924-8.
5. http://www.acdlabs.com/comm/elucidation/2015_05.php

Structure Generation

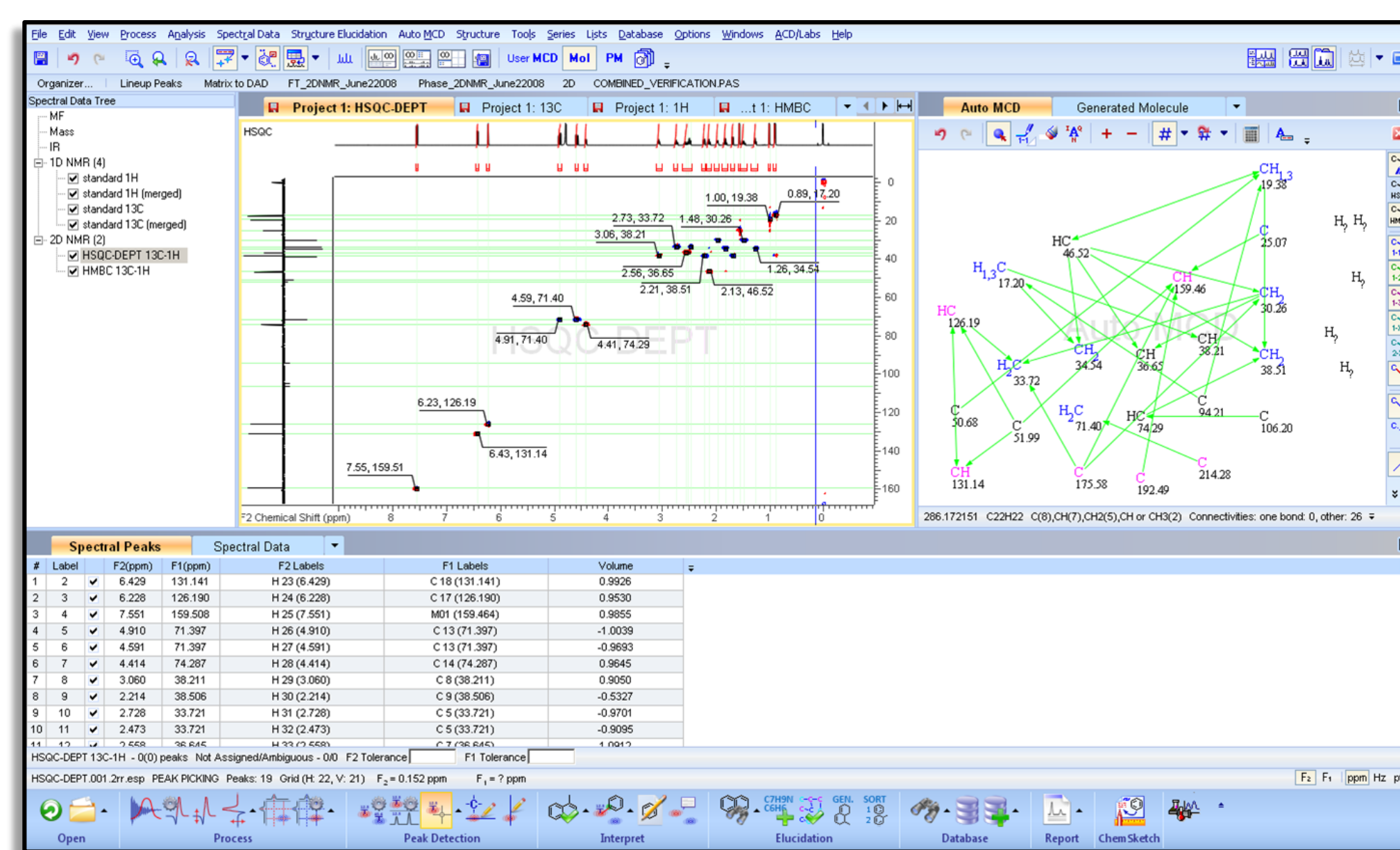


Figure 3: A Molecular Connectivity Diagram (MCD) is created automatically during data interpretation. This can be used both for de novo elucidation, or to perform an alternative check on proposed structures.

Ranking the Best Candidates

Using the advanced algorithms of ACD/Structure Elucidator Suite, the entirety of the chemical space can be queried for possible structures which agree with the experimental data. The software efficiently evaluates all structural possibilities, and then ranks the best candidates for review based on the differences between the experimental and predicted chemical shifts.

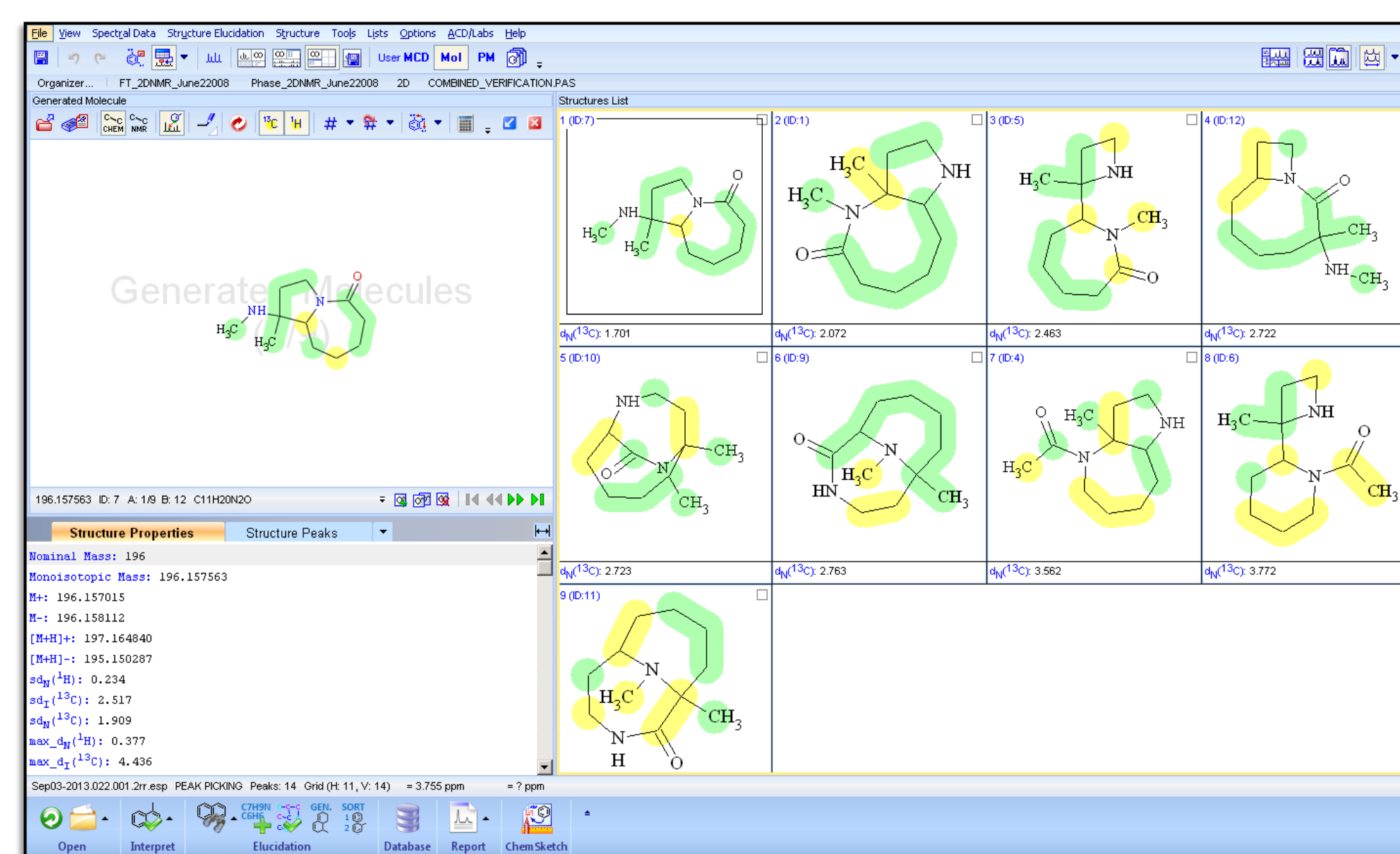



Figure 4: A list of potential structures, ranked by their associated chemical shift deviations. Structural assignments are color-coded based on their agreement with the experimental data.

Rapid Identification of Structures

With the right databases, known structures can also be quickly identified from previous work. The following DBs are available using ACD/Labs software:

- ACD/HNMR and CNMR DB
 - 320,000 assigned chemical structures
- ACD/Structure Elucidator DB
 - 425,000 structures
 - 2,300,000 structural fragments
- ChemSpider DB
 - 22,000,000 predicted compounds with ¹³C, ¹H, ¹⁹F, ³¹P, and ¹⁵N NMR
- Dictionary of Natural Products
 - Over 220,000 compounds
- Marinit
 - > 21,000 compounds (marine sources)
- AntiBase
 - > 35,000 compounds (microorganisms and higher fungi)



Advanced Chemistry Development, Inc.
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com

The Structure Elucidation Challenge

Between 2003 and 2011, 112 different datasets were submitted from various industries as challenges to the ACD/Structure Elucidator Suite system. [3] The structures determined ranged from 100 to more than 1000 Da (Figure 5) and the software had excellent success determining the correct structures (Figure 6).

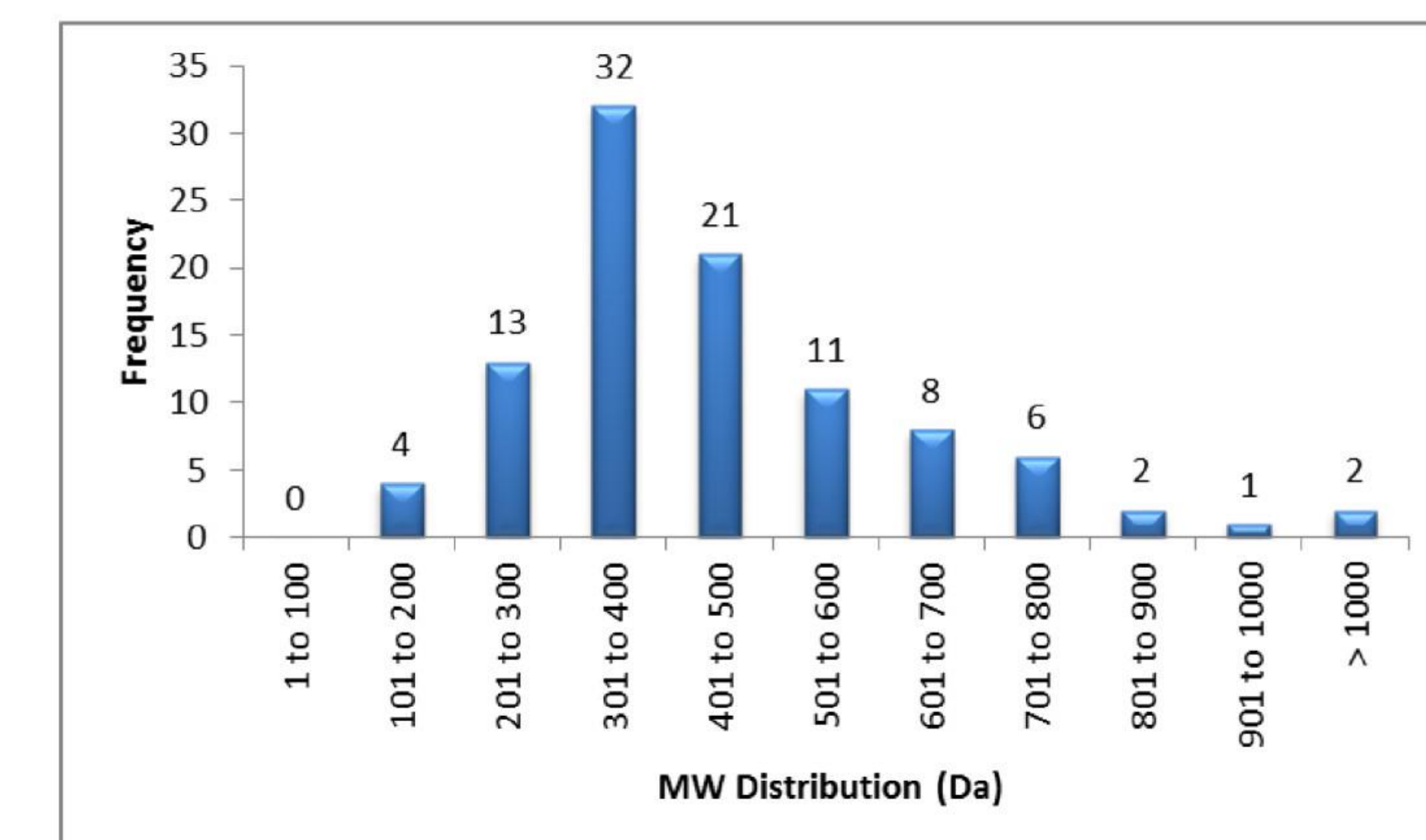


Figure 5: The distribution of molecular weight in structural challenges submitted to ACD/Labs.

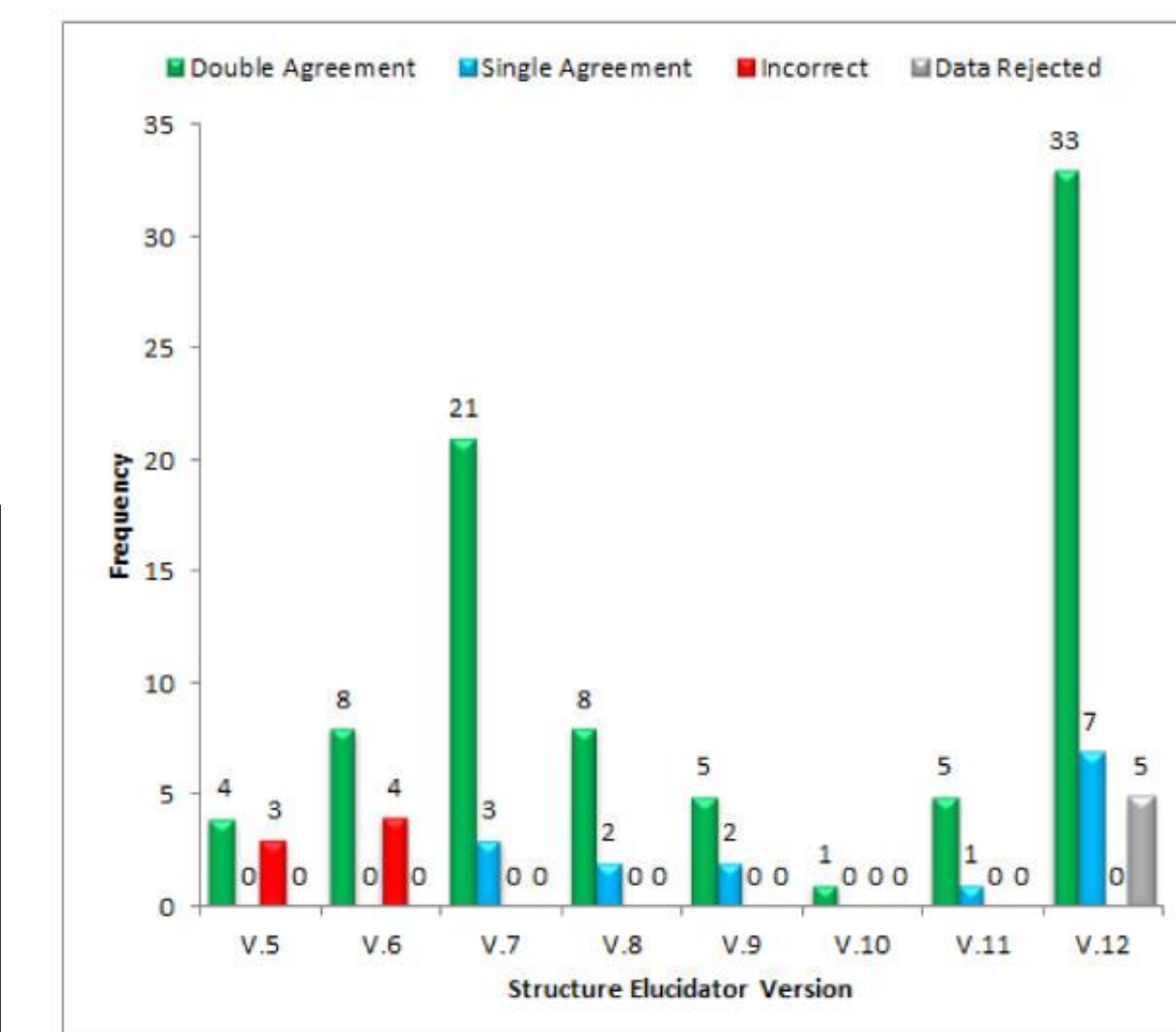


Figure 6: The outcomes of structural evaluations performed by ACD/Structure Elucidator Suite from 2003-2011.

In one recent example, Structure Elucidator Suite was able to determine the structure of a large, symmetrical dimer known as Bacillus A [4] (containing 86 skeletal units) even with the presence of ambiguous connectivities. [5]

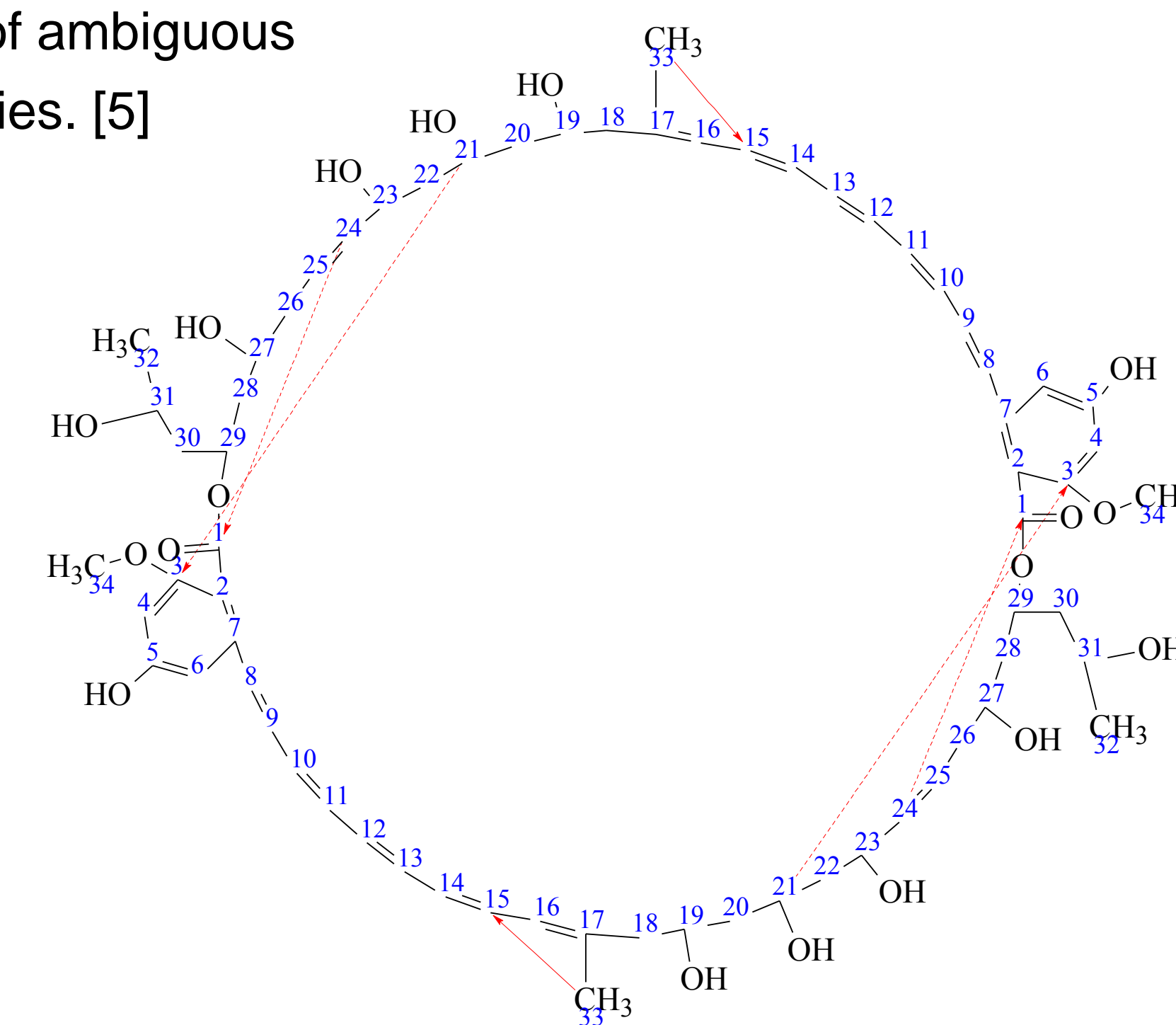


Figure 5: The elucidated structure of Bacillus A

Conclusions

Modern CASE systems such as Structure Elucidator Suite provide the necessary capability accurately elucidate a novel chemical structure for complex molecules based on readily available NMR data sets. This allows organizations to avoid expensive, labor-intensive, and time-consuming synthetic efforts. The routine use of CASE tools should also prevent the introduction of many improper structures into the literature, or into compound registries.



Reprints:
conferences@acdlabs.com