A Review Of Computer Assisted Structure Elucidation (CASE) Methodology

David C. Adams, Dimitris Argyropoulos

Advanced Chemistry Development Inc., 8 King Street East, Suite 107, Toronto, Ontario, Canada

Introduction

Computer-Assisted Structure Elucidation (CASE) applications are widely used today to characterize chemical structures from both natural and synthetic products. In the past decade, multiple reviews have been published to illustrate that ACD/Structure Elucidator Suite [1-4] is capable of identifying the chemical structure of very complex molecules with the aid of NMR and MS information.

For the past 5 years ACD/Labs has performed and posted an "Elucidation of the Month" based on published experimental data (http://www.acdlabs.com/comm/elucidation). Here we review the variety of almost 60 structures solved as part of this program and show details about the types of structures solved, their sizes (number of heavy atoms), the proton content, the elucidation time and the confidence level on the validity of the proposed structure.

Elucidation Strategy

ACD/Structure Elucidator needs as minimum input an HSQC and an HMBC spectrum. It can operate with only an HSQC and a 1D-^{13}C spectrum under some very special conditions but the recommended minimum is a 1D-^{1}H, 1D-^{13}C, an HSQC and an HMBC spectrum. Spectra of other nuclei can also be included (^{15}N for example) as well as other types of heteronuclear correlation spectra (H2BC, ADEQUATE, INADEQUATE etc.) or multiple HMBC's. A Molecular Formula or an accurate mass spectrum from which a MF can be calculated is also needed. Furthermore, information from other techniques (FTIR, Raman) can aid in the elucidation.

The general strategy is to first enter the spectral information (spectra and picked peaks) which will be automatically combined to create a Molecular Connectivity Diagram (MCD). The MCD can be manually edited if needed (e.g. to define C=O groups). Once the MF is entered the data can be checked for consistency and, if they pass, structure generation can start. Not all ^{13}C or ^{1}H signals need to be found in spectra nor do all correlations need to be unambiguous. The program can tolerate ambiguity but this will increase the elucidation time.

If there are no or very few and irrelevant structures generated then there are a number of options to try. The first one would be to enable "Fuzzy" generation, i.e. allow the software to extend connectivities. By default the number of bonds corresponding to a particular cross peak of a 2D spectrum is calculated based broadly on the peak intensity. "Fuzzy" generation allows this to be extended accommodating for, e.g. 4 or 5 bond correlations in an HMBC.

If "fuzzy" generation is not successful the next thing to try would be to allow bonds between heteroatoms (e.g. $-NO_2$ groups) and even further to allow bonds between heteroatoms of the same type (e.g. -N=N- or -O-O- groups). The example of Virosaine from February 2015 nicely illustrates all of these strategies.

The Molecules Solved

All the molecules solved were natural products. Out of them there were 7 alkaloids, 14 terpenes, 4 peptides, 5 metabolites and a few other classes of compounds. These are shown in Table 1.

The vast majority of the molecules had a molecular weight of 250-450 Da (>60%) while almost 20% had more than 600 (Figure 1).

More than 60% of the structures were solved in less than 1 minute and more than 90% were solved in less than 1 hour (Figure 2).

Name	M.W.	M.F.
a-Botryocsnthin	1033.68	$C_{74}H_{112}O_2$
Phoriospongin A	1136.72	$C_{52}H_{82}N_{11}O_{15}Cl$
Taslamide B	979.17	$C_{50}H_{74}N_8O_{12}$
Psychotripine	514.66	$C_{33}H_{34}N_6$
Lycojaponicumin	291.34	$C_{16}H_{21}NO_4$
Daphmacromine A	431.56	$C_{25}H_{37}NO_5$
Lasionectrin	344.36	$C_{19}H_{20}O_6$
Aetheramide	718.87	$C_{41}H_{54}N_2O_9$
Geranylphenazinediol	348.44	$C_{22}H_{24}N_2O_2$
Asperjinone	380.39	$C_{22}H_{20}O_6$
Indol Alkaloid	371.48	$C_{18}H_{17}N_3O_2S_2$
Lycojaponicumin D	277.36	$C_{16}H_{23}NO_3$
Protuboxepin A	377.44	$C_{22}H_{23}N_3O_3$
Geralcin A	312.36	$C_{15}H_{24}N_2O_5$
Polypropionat	320.46	$C_{20}H_{32}O_3$
Asidia SAAF	638.74	$C_{27}H_{44}O_{10}S_2Na_2$
Jatrophalactam	331.45	$C_{20}H_{29}NO_3$
Aquatolide	246.3	$C_{15}H_{18}O_3$
Phosphoiodyn A	437.25	$C_{16}H_{25}O_3NP_I$
Ascidia sydneiensis SAAF	638.74	$C_{27}H_{44}O_{10}S_2Na_2$
Epohelmin A	293.44	$C_{18}H_{31}NO_2$
Schizocommunin	289.29	$C_{17}H_{11}N_3O_2$
Barmumycina	277.31	$C_{15}H_{19}NO_4$
TAEMC161 (Viridol)	354.35	$C_{20}H_{18}O_6$
Mandelalide A	624.76	$C_{33}H_{52}O_{11}$
Gymnopalyne	218.64	$C_{12}H_7O_2Cl$
Trigoflavidol A	612.62	$C_{35}H_{32}O_{10}$
Acremolin	231.25	$C_{11}H_{13}N_5O$
Strynuxline A	408.45	$C_{23}H_{24}N_2O_5$
Puberunine	437.53	$C_{23}H_{35}NO_7$
Schiglautone A	502.68	$C_{30}H_{46}O_6$
Virosaine	235.24	$C_{12}H_{13}NO_4$
Laevinoid A	342.38	$C_{20}H_{22}O_5$
Sinensilactam	403.38	$C_{20}H_{21}NO_8$
Bacillusin A	1197.44	$C_{68}H_{92}O_{18}$
Garcimulin A	602.8	$C_{38}H_{50}O_6$
Heterodimer P. Kaurabassana	556.52	$C_{31}H_{24}O_{10}$
Arboridinine	294.39	$C_{19}H_{22}N_2O$
Spirochensilide A	482.65	$C_{30}H_{42}O_5$
Tronoharine	336.43	$C_{21}H_{24}N_2O_2$
Callyspongiolide	628.59	$C_{33}H_{42}BrNO_6$
Flueggether A	438.51	$C_{25}H_{30}N_2O_5$
Ruthmycin	386.39	$C_{21}H_{22}O_7$
Teotihuacanin	356.37	$C_{20}H_{20}O_6$
Sarglaperoxide A	384.46	$C_{23}H_{28}O_5$
Mannolide A	344.4	$C_{20}H_{24}O_5$
Astellifadiene	340.58	$C_{25}H_{40}$
Aspterpenacid A	446.62	$C_{27}H_{42}O_5$
Pepluacetal	360.49	$C_{22}H_{32}O_4$
Ciliatonoid A	424.53	$C_{26}H_{32}O_5$
Euphorikanin A	314.42	$C_{20}H_{26}O_3$
Alistonitrine A	367.44	$C_{21}H_{25}N_3O_3$
4-Bromobenzoic-biscognienyne A	445.35	$C_{23}H_{25}O_4Br$
Sophaline C	338.44	$C_{21}H_{26}N_2O_2$
Phomopsterone	486.64	$C_{29}H_{42}O_6$
Delicoferone	552.48	$C_{28}H_{24}O_{12}$
Waspergillamide A	432.43	$C_{20}H_{24}N_4O_7$
Spiroschincarin A	586.63	$C_{31}H_{38}O_{11}$

Table 1: Names of the molecules whose structure was solved, their molecular weights and corresponding molecular formulas.

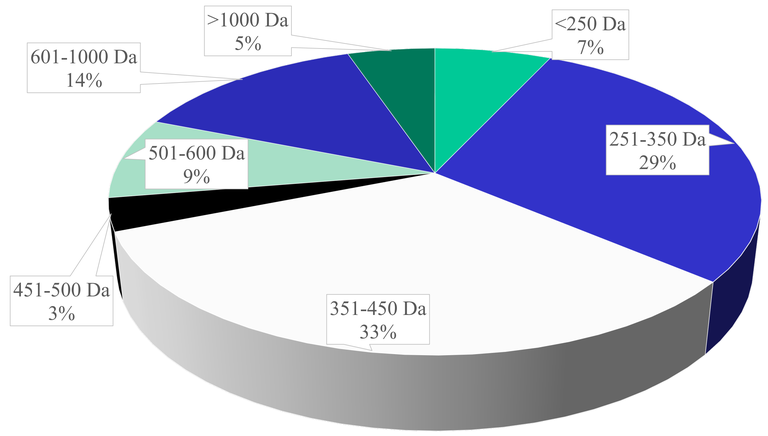

Figure 1: Distribution of the molecular weights of the structures solved.

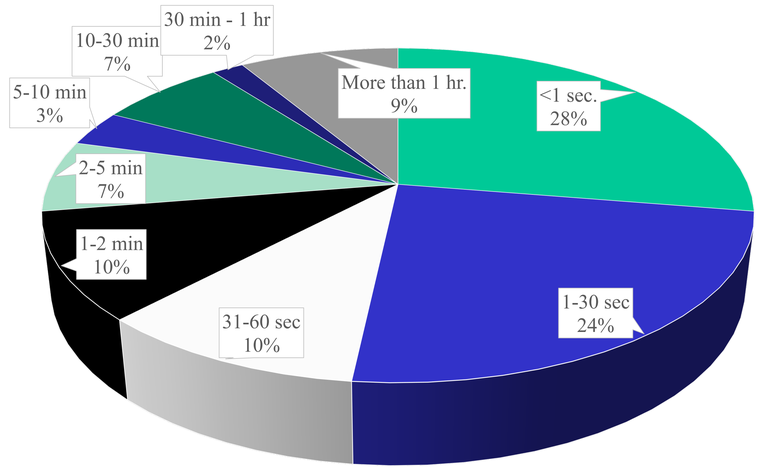

Figure 2: Elucidation time distribution for the solved structures.

The number of heavy atoms (C+N) in the structures varied from 13 to 86, almost in accordance with the molecular weights. 12 of the structures solved (20%) fall under Crews Rule[5], i.e., the ratio of the number of protons to the number of heavy atoms is less than 1, which asserts that such molecules may be difficult or impossible to characterize. Furthermore, 26 of the structures solved had ratio values between 1 and 1.3, i.e. very low. Nevertheless this did not seem to have an effect on the elucidation time or the confidence in the solution; some structures with low ratios were solved faster than structures with higher ones.

Most of the structures were solved with a very high confidence level. This means that the "Best Structure" had a significantly smaller mean difference in chemical shift between the experimental and theoretical ^{13}C spectra than the other calculated structures. That is, the top calculated structure was the clear favorite. In the very few cases where the "Best Structure" had comparable accuracy to lesser ranking results, the correct structure was deciphered by examining the ^{1}H NMR spectra or additional 2D NMR data.

Conclusions

The ACD/Labs Structure Elucidator CASE system has proven itself time and again on various challenges presented. The large majority of structures are solved in a relatively short time (< 1 hour) with a very high degree of confidence.

Please contact us if you would like to participate in our Structure Elucidation Challenge where we will solve your structure using NMR data.

References

1. Elyashberg, M. E.; Williams, A. J.; Martin, G. E. *Prog. NMR Spectr*. **2008,** *53*, 1-104.
2. Steinbeck, C. Nat. Prod. Rep. **2004,** 21, 512-518.
3. Naman, C.B.; Li, J.; Moser, A.; Hendrycks, J. M.; Benatrehina, P. A.; Chai, H.; Yuan, C.; Keller, W. J.; Kinghorn, A.D., *Organic Letters*, **2015,** *17*, 2988-2991
4. Lorenc, C.; Sauri, J.; Moser, A.; Buevich, A. V.; Williams, A. J.; Williamson, R. T.; Martin, G. E.; Peczuh, M. W.; *ChemistryOpen*, **2015**, *4,* 577-580.
5. White, K. N.; Amagata, T.; Oliver, A. G.; Tenney, K.; Wenzel, P. J.; Crews, P. *J. Org. Chem*. **2008**, *73*, 8719-8722.

www.acdlabs.com/structureelucidator