A Streamlined Workflow for the Rapid Identification of Unknowns Using MS Structure ID

Brent G. Pautler, Richard Lee & Joe DiMartino
Advanced Chemistry Development, Inc.
Toronto, ON, Canada
www.acdlabs.com

Introduction

When performing any type of screening assay, one of the most significant analysis challenges lies in the identification of unknown components. While high resolution and high mass accuracy enables greater precision when predicting elemental composition, these estimates lack structural information. When faced with a true unknown, even the fragment ion analysis has limitations and can only highlight particular functional groups at best.

In this work, we present a workflow that utilizes the available analytical information, such as predicted elemental formula, fragment ions and the inclusion/exclusion of known fragments to refine a mass-based local library search (PubChem & ChemSpider). Combining these parameters provides an accumulative filter that reduces search results to a manageable number, or best case scenario, a single structure. The workflow highlighted in this paper uses an unknown pharmaceutical sample as a test case but the software may be applied to samples from any number of sources such as food & beverages, formulated products, metabolomics, and environmental analyses.

Experimental

The pharmaceutical sample was analyzed by HPLC-FTMS (Thermo Fisher Scientific Orbitrap equipped with an electrospray ionization source). The datasets were then loaded into ACD/Spectrus and processed using MS Structure ID. Local versions of the PubChem and ChemSpider databases were used for the mass search. (Note that the ChemSpider library has been curated to include only those chemicals not found within PubChem.) Additional confirmation of the results was then performed.

Identification of Unknown Components

The data was processed in ACD/MS Structure ID, using the IntelliXtract (IX) 2.0 algorithm to extract all chromatographic components; these are displayed in Figure 1. The component at t_R=5.102 minutes and $[M+H]^+$ 272.2017 m/z is selected for identification.

Figure 1: *The components of the pharmaceutical sample, extracted by IntelliXtract 2.0.*

This selected component has MS2 data associated with its $[M+H]^+$ peak which includes 272.2017 m/z parent mass. It is used as the source for the elemental composition determination and to query the PubChem & ChemSpider databases. Based on the resulting isotope pattern, it is expected that only C, H, N, and/or O atoms are present in the molecular formula. The Formulae Generator suggests $C_{18}H_{25}NO$ as the best fit based on the accurate mass.

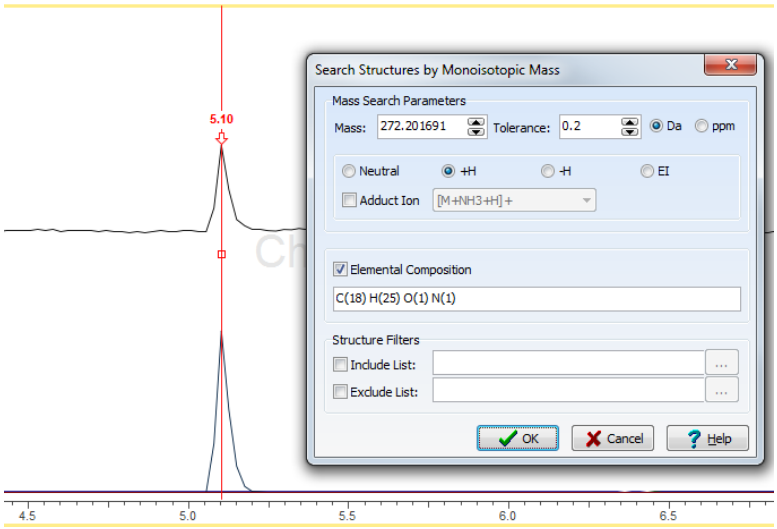

Figure 2: *Selecting the m/z value and Elemental Composition with which to query the databases*

Searching for the selected m/z value (272.2017) with the specified elemental composition in the local databases (containing approximately 96 million compounds), and a tight tolerance of 2ppm yields an initial set of 6692 structures. After removing duplicate stereoisomers becomes 5752 unique structures (**Figure 3**).

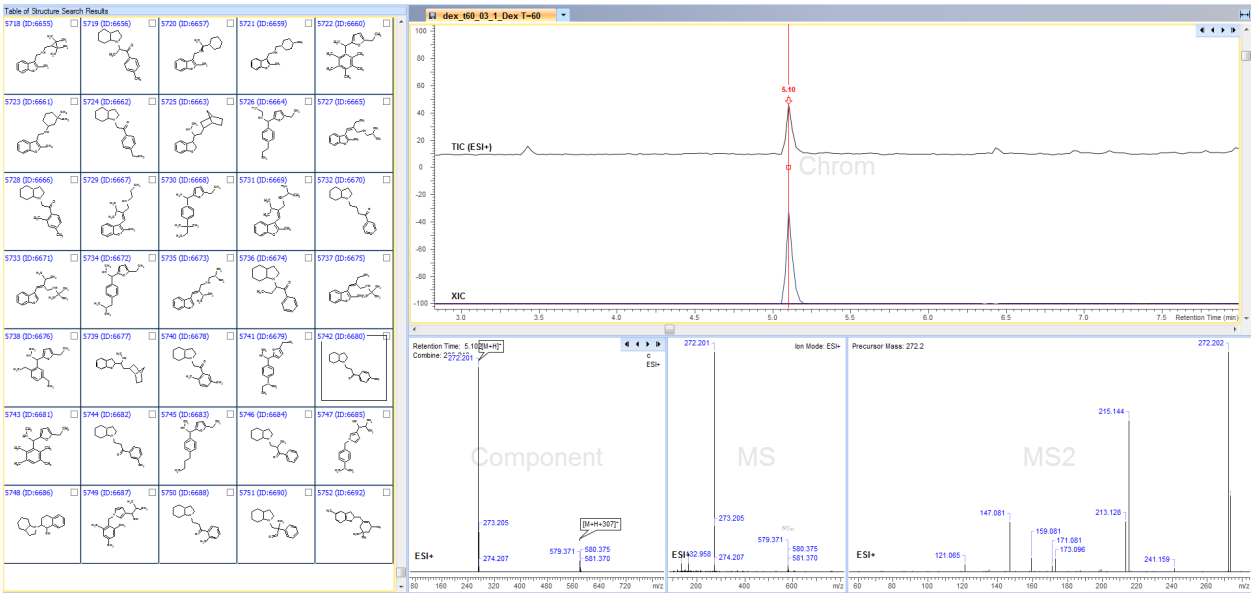

Figure 3: *The results table indicates 5752 structures have been found for the given target mass.*

Include/Exclude List from Spectrus DB

As this initial list of possible structures is still quite large, a more manageable list can be generated by additional filtering. By adding structure fragments to include and/or exclude lists within Spectrus DB (**Figure 4**), the list of possible structures can be drastically reduced.

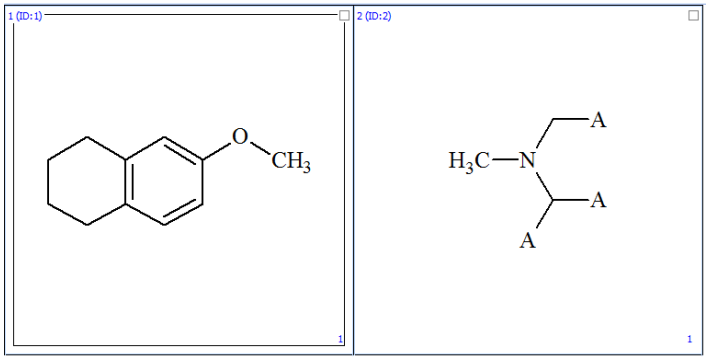

Figure 4: *Additional structure fragments added as an Include List. These fragments were selected through close examination of the MS2 data along with complementary analytical data acquisition.*

After removing duplicate structures from the additional fragment information search, the list is reduced to a much more manageable 22 candidates (**Figure 5**).

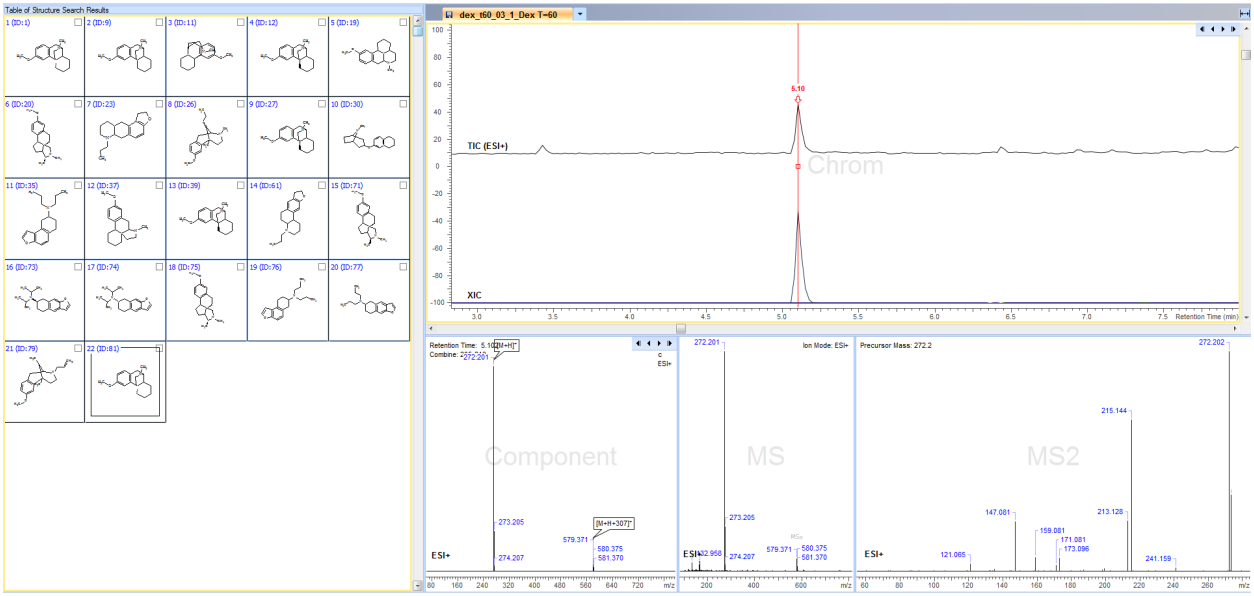

Figure 5: *The list of potential structure candidates has been reduced from 5752 to 22 after the additional known structure fragments were added as an Include List.*

Ranking Structures Returned by the Databases

The additional functionalities available in ACD/MS Structure ID can be used to further rank structures based on an "Assignment Score" from the AutoAssignment Tool. AutoAssignment uses a fragment-prediction engine to fragment the parent structure and the generated fragments are then matched to the MS2 spectrum presenting a score / data match value (**Figure 6**).

After running AutoAssignment, 8 structures possess a Match (dMS) value of 0.75 or higher, with only 2 having values above 0.80.

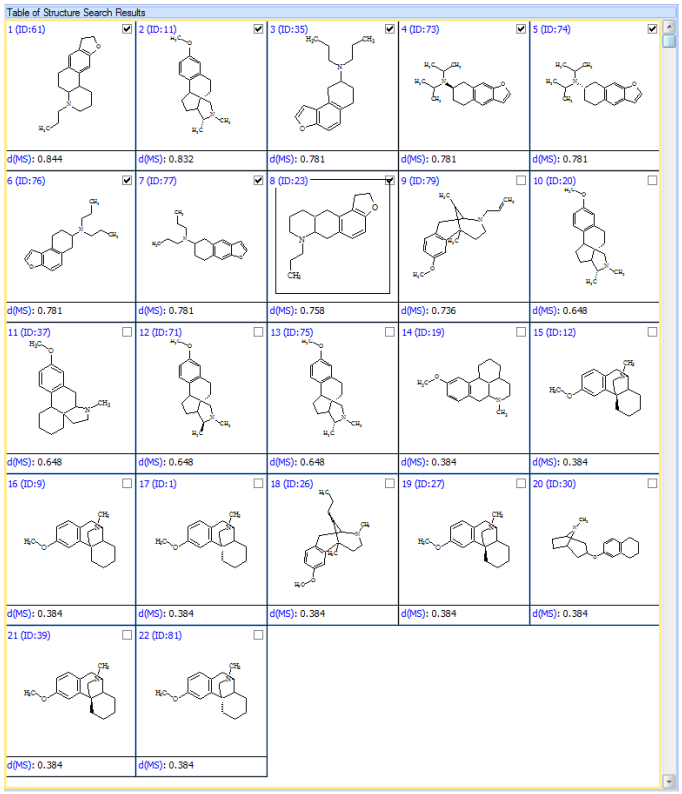

Figure 6: *The potential structure candidates after AutoAssignment d(MS) ranking.*

The top structure candidates can be examined more closely by extracting the MS2 spectra in a separate window and performing additional AutoAssignment / Fragmentation analysis by increasing the number of fragments in the generator on each structure to help deduce the structure that best structure matches the data (**Figure 7**).

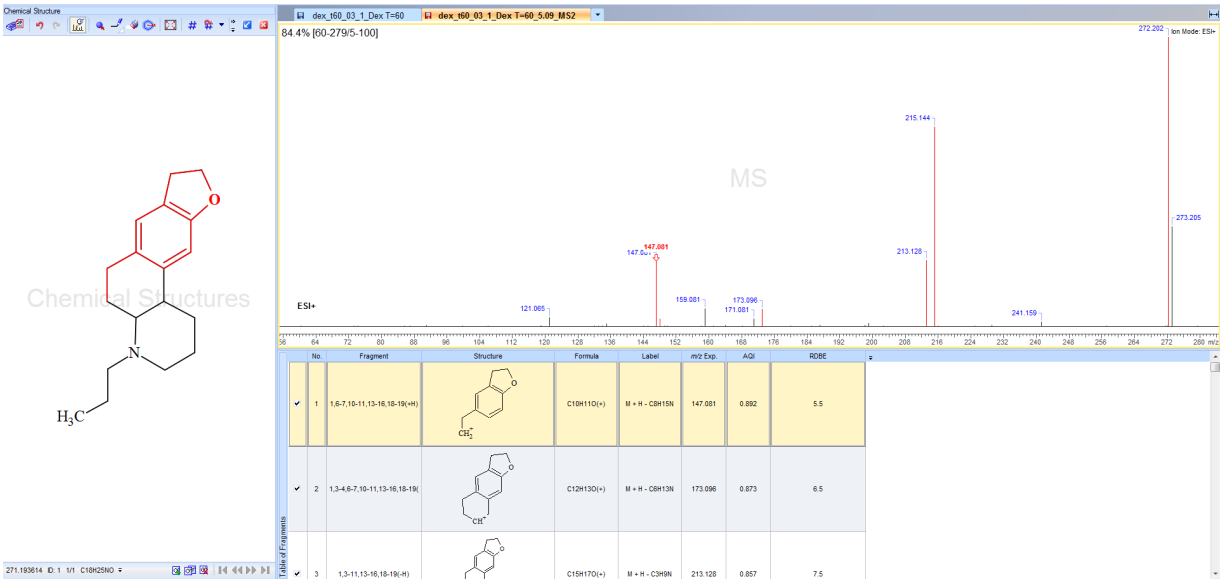

Figure 7: *Further Fragmentation Analysis of the top candidate in ACD/MS Structure ID.*

Conclusions

This example demonstrates how ACD/MS Structure ID is able to quickly search a wide range of potential structures, curate a reasonable and relevant list of candidates, and identify the most likely structure for a chromatographic peak, all in an extremely efficient workflow.

To learn more about ACD/Labs MS Structure ID, please visit **www.acdlabs.com/msstructureid/**