Isotope modeling routines applied to empirical formula prediction

Simon Ashton[1] ; Richard Gallagher[2] ; Neil Loftus[1] ; John Warrander[1] ; Ichiro Hirano[3] ; Shinichi Yamaguchi[3] ; Norio Mukai[3]
[1]Shimadzu ISS, Manchester, UK ; [2]AstraZeneca, Alderley Park, Macclesfield, UK ; [3] Shimadzu Corporation, Kyoto, Japan

Overview

Predicting a candidate list based on MS and MS^n data takes into account a number of variables, including mass accuracy and mass resolution of the experimentally derived pseudo-molecular peak and related fragment ion data generated using MS^n spectra (or complement ion data) together with conventional chemical rules (such as nitrogen, electron configuration, DBE range, H/C ratio).

As the hybrid quadrupole ion trap TOF system generates MS and MS^n spectra with a low mass cut off the prediction tool set can use the fragment ion data to reduce the list of potential candidates and remove invalid formula.

Although the main variables are mass accuracy and isotopic distribution for the MS spectra, the results also highlight the importance of fragment ion data in filtering candidate's to remove invalid formulae and helping to identify the correct empirical formula with a higher degree of confidence without the need for using sub-ppm mass accuracy data.

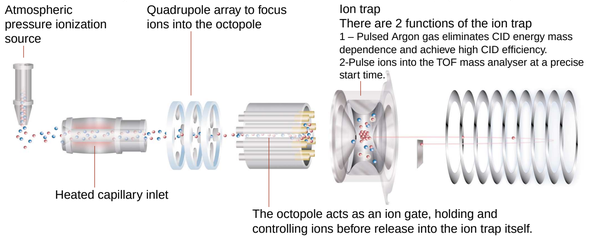
This paper will show the importance of advanced modeling tools for fitting experimental and theoretical mass accuracy and isotope distribution data applied to empirical formula prediction using a diverse chemical library as a test system.

Introduction

High mass accuracy mass spectrometry is achievable on a number of technology platforms including time of flight mass analyzers, hybrid ion trap and quadrupole time TOF's to FT-MS. Such technologies are playing an increasing role in pharmaceutical drug characterization and metabolite identification to help identify or confirm empirical formulae assignments. Recently, software tools have been applied to modeling theoretical isotopic distributions and mass accuracy data with experimentally derived data together with sets of chemical or statistical rules to help increase the probability of assigning the correct empirical formula. In this paper we describe the development of an empirical formula prediction tool that take into account MS and MS^n spectra with high mass accuracy and mass resolution.

Methods

To evaluate and develop the prediction tool set a small library of pharmaceutical compounds was evaluated on a hybrid quadrupole ion trap time of flight mass analyser (LCMS-IT-TOF, Shimadzu Corporation, Kyoto, Japan).



The software was designed to support 3 stages in data processing.

Stage 1 : MS spectrum data
Considers the molecular ion (or adduct) and mass accuracy

Parameters for MS spectrum data
To increase the likelihood of identifying the correct empirical formula conventional filters are applied, including DBE Range, Electron configuration, HC ratio, Nitrogen rule, Active elements, Adducts, Mass tolerance, Chemical rules settings.

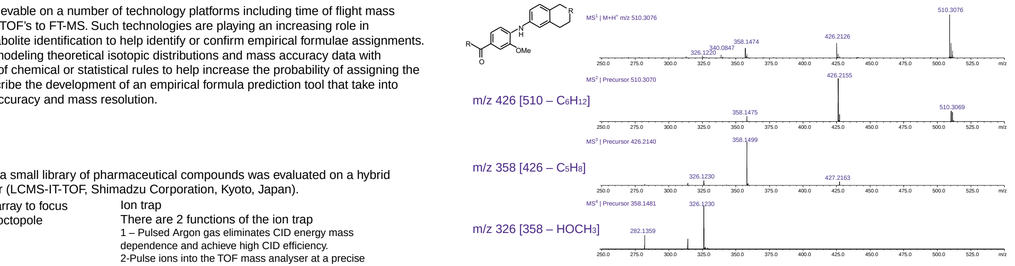
Stage 2 : MS^n spectrum data
This stage considers MS^n fragment ions generated for the molecular ion of interest and calculates the theoretical complement ion for each fragment ion.

Parameters for MS^n spectrum data
Filters out invalid formulae using the fragment ion data.
This technique is applied recursively through the various stages of MS^n data allowing elemental composition of MS^4 fragment ions filter the MS^3 precursors' formulae generated using elemental composition; MS^3 ions are used to filter MS^2 ions and MS^2 ions are used to filter the MS^1 molecular ion.

Stage 3 : Isotope fitting of possible formulae
This stage generates the theoretical isotopic distribution profile spectra for each of the possible formulae generated for a molecular ion using accurate mass and elemental composition.

Parameters for Isotope distribution
Least squared fitting routines are applied to pattern matching between experimental derived data and theoretical distributions.
A likelihood scoring algorithm takes into account the closeness of fit to the theoretical isotope data, the variation in mass accuracy and the candidate list filtered using MS^n fragment data. The log transformation ranking score lists more likely candidates with a higher score in the range 0-100.

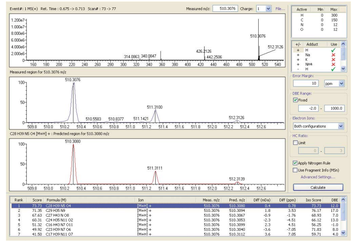
Results



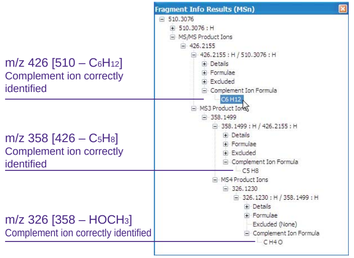
MS^n mass spectra of a new chemical entity.

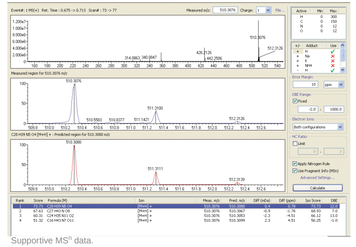
MS and MS^n data were used to predict/confirm a likely empirical formulae using mass accuracy data on the parent and fragment ions.

MS^1 data alone.
Mass accuracy and mass resolution are key parameters in identifying the correct empirical formula. To increase the reliability of the candidate list the experimentally derived MS data is also compared to the theoretical isotope pattern to provide an Iso score. In this case the empirical formulae of the chemical entity was $C_{28}H_{39}N_5O_4$. (The software tool also predicted the same empirical formulae as the first hit).

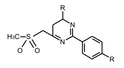


Supportive MS^n data
Using the high mass accuracy of the fragment ions it is possible to exclude possible candidates and identify the correct formulae using the fragment ion information.



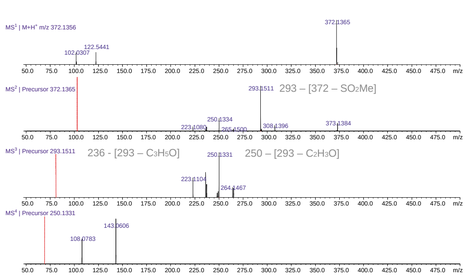


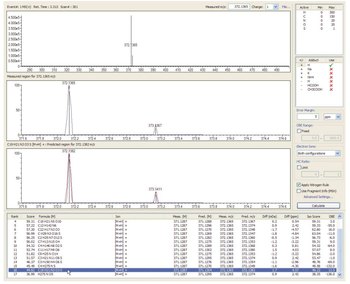
Results



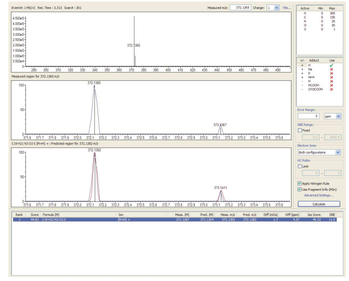
MS^n mass spectra of a new chemical entity empirical formulae $C_{19}H_{21}O_3N_3S$

MS^3 data was used to generate ions at relatively low m/z values to help increase the likelihood of the first hit using MS^n fragment ions.





Using only MS^1 data the true candidate was identified as hit number 16



Using MS^n data the true candidate was identified as hit number 1

Discussion and Conclusions

- Tools to predict empirical formulae from accurate mass data typically consider chemical rules and isotope fitting routines. In this paper we describe the application of MS^n data to help reduce candidate lists and to increase the likelihood of predicting the correct formulae.
- There are several factors that clearly affect the probability of identifying the true candidate, accurate mass being the most critical.
- However, by providing tools that consider the fragment ion data in addtion to pseudo-molecular ion infromation it has proved useful in identifying the correct candidate mass with a higher probability than using simply MS data alone.