Automated Molecular Formula Determination in Open Access LC/MS in Drug Discovery.

Vladimír Čápka[1], Sharon Tentarelli[1], Hongliang (Leo) Xu[2]

[1]AstraZeneca R&D Boston, 35 Gatehouse Dr., Waltham, MA 02451, [2]Cerno Bioscience, 14 Commerce Dr., Danbury, CT 06810

AstraZeneca

Overview

•Molecular formula determination based on mass and spectral isotope pattern can play an important role in compound identification.

•Sample analysis using a single quadrupole MS instrument and MassWorks™ software with isotopic pattern calibration can yield molecular formula determination.

•User-friendly automated software processing has been developed to allow for integration into open-access systems in the drug discovery environment.

•Capabilities and limitations of open-access molecular formula determination are compared.

Introduction

•In drug discovery applications, molecular formula determination based on accurate mass and spectral isotope pattern is typically measured using high resolution and performed by skilled analytical chemists.

•Use of high resolution instrumentation such as TOF in open access is problematic due to required user expertise and operation complexity.

•A typical open-access environment contains single quadrupole MS instruments operated by synthesis chemists with moderate MS expertise.

•MassWorks™ software with isotopic pattern calibration can be used for molecular formula determination on a single quadruple instrument.

•Interfacing MassWorks™ software with MassLynx™ software can automate much of the calibration and sample analysis process, minimizing the need for end-user expertise.

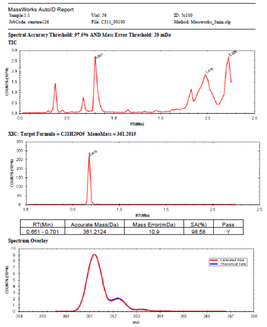
•Effect of user input of target ion, element list, and target molecular formula was investigated.

•28 model compounds (MW 292-652 Da) with typical elemental composition in drug discovery were used.

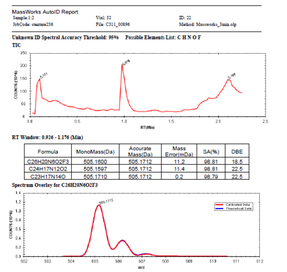
Results

Three Typical Workflows

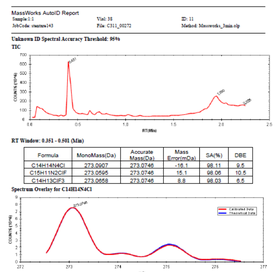
- Chemist desires confirmation of the presence of a known compound.
- Chemist inputs expected molecular formula during sample login.
- Report shows confirmation, or failure and proposed alternate formulas.



- Chemist has limited molecular knowledge about an unknown sample.
- Chemist inputs target ion of interest and/or elements likely to be present during sample login.
- Report shows proposed formulas based on ion and/or element list.



- Chemist has a completely unknown sample
- Chemist inputs no molecular information during sample login.
- Report shows proposed formulas for predominant ions within spectra based on default element list



Formula Identification Using a Variety of Different Workflows

Table I. Effectiveness of sample identification for 28 test compounds based on different user input during sample login..

Input →	Formula		No input			Target ion			Elements			Elements and target ion		
Formula	Mass Error (mDa)	SA (%)	Rank	Mass Error (mDa)	SA (%)	Rank	Mass Error (mDa)	SA (%)	Rank	Mass Error (mDa)	SA (%)	Rank	Mass Error (mDa)	SA (%)
$C_{26}H_{28}N_4O_2$	8.3	98.8	NF			NF			1	-10.6	98.44	2	0	98.7
$C_{30}H_{27}ClN_2O$	-7.4	98.7	TNA			NF			TNA			2	-4.5	99.0
$C_{15}H_{10}N_4OS$	5.1	99.0	1	-11.3	99.2	1	-11.3	99.2	2	9.3	99.29	3	-9.3	98.6
$C_{17}H_{18}FNO_4S$	5.2	98.3	TNA			NF			TNA			2	0.5	98.7
$C_{19}H_{23}NO_2$	8.5	99.3	TNA			1	1.7	99.3	TNA			1	5.3	98.2
$C_{18}H_{24}ClN_3O_3$	2.7	98.0	3	-14.0	98.1	2	-2.7	98.3	2	-4.0	98.59	1	-6.5	99.0
$C_{13}H_{16}N_4O_2S$	6.7	98.2	1	-11.5	99.1	NF			3	2.6	98.96	1	2.3	98.7
$C_{15}H_{17}Cl_2NO_2$	9	98.7	1	-7.2	99.1	2	-5.2	99.4	1	7.6	99.22	1	-2.4	99.2
$C_{20}H_{23}ClN_6O$	7.1	98.3	NF			3	-15.7	99.0	2	0.8	98.78	2	3.2	98.8
$C_{19}H_{22}ClN_3O$	15	94.2	TNA			NF			TNA			NF		
$C_{22}H_{26}N_2O_2$	2.4	99.1	TNA			3	-5.6	99.5	TNA			2	-4.9	98.9
$C_{23}H_{28}N_2O_3$	-4.5	99.4	NF			2	-7.6	99.4	2	-3.1	99.11	2	-1.4	99.2
$C_{25}H_{24}N_4O_6$	3.4	98.5	NF			1	-11.6	97.4	NF			2	-3.9	99.1
$C_{27}H_{28}N_4O_3$	-13.7	98.9	1	-3.2	98.9	NF			1	-7.4	99.01	1	-11.1	99.1
$C_{26}H_{26}N_4O_3$	-13.5	98.6	NF			NF			1	-10.2	98.98	1	0.2	99.2
$C_{27}H_{29}N_5O_3$	-12.9	98.7	TNA			NF			TNA			2	-1.7	98.5
$C_{28}H_{31}N_5O_3$	-13.9	98.1	TNA			NF			TNA			3	-14.1	97.5
$C_{27}H_{22}F_2N_4O_4$	-12.6	98.0	TNA			1	-9.5	98.4	TNA			2	-5.4	99.2
$C_{26}H_{19}F_3N_6O_2$	11.8	97.8	NF			NF			1	11.2	98.81	NF		
$C_{26}H_{27}N_5O_2$	-2.9	98.1	NF			NF			2	-5.8	99.17	2	-7.4	98.8
$C_{26}H_{21}ClN_8O_2S$	-12.1	97.8	NF			1	1.0	98.2	3	10.2	98.39	NF		
$C_{27}H_{23}FN_6O_4$	-13.2	98.6	TNA			NF			TNA			1	-1.4	99.3
$C_{30}H_{33}N_5O_3$	-17	97.9	NF			NF			NF			2	-15.8	95.0
$C_{25}H_{20}ClN_5O_4$	-12.6	99.0	NF			1	-12.5	98.8	NF			2	-8.3	99.0
$C_{24}H_{19}ClN_4O_2$	-14.1	98.1	NF			NF			1	-12.4	98.79	1	-2.6	99.0
$C_{29}H_{23}F_3N_8O_5S$	1.5	97.8	NF			NF			NF			NF		
$C_{28}H_{32}ClN_3O_6$	-11.4	99.0	TNA			NF			TNA			2	1.4	99.0
$C_{29}H_{34}ClN_3O_6$	-2.9	99.0	TNA			NF			TNA			2	8.0	99.3
Total correctly identified	27		5			11			13			24		

One failed sample (SA=94.2) due to poor ionization.

Unlikely workflow; some sample information is usually known (elements, mass).

Most likely workflow; mass and elements are known from other open access analyses and reaction scheme.

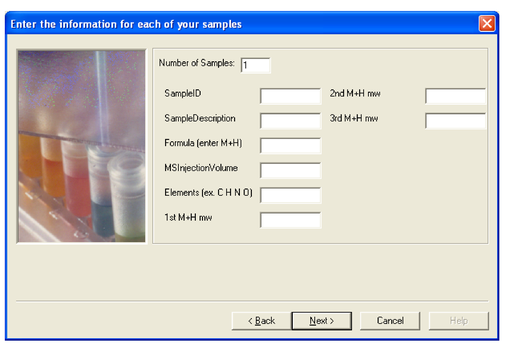
NF = Not Found; correct formula was not within top 3 formulas.
TNA = Target Not Analyzed; target ion was not the largest ion within the spectrum.

Materials and Methods

	Equipment Setup
Instrument	Waters ZQ single quadrupole Agilent 1100 HPLC
Column	Waters XBridge™ C18, 3.5 μm, 2.1 x 20 mm
Flow Rate	1.8 mL/min
Injection Volume	5 μL
Mobile Phase	A: 0.1% formic acid in H_2O B: 0.1% formic acid in CH_3CN
Gradient	5 to 95% B in 1.75 min
MS detection, ESI+	Continuum mode, 175 – 695 amu at 350 amu/s
Resolution	Unit, ~0.5 amu peak width half height
Calibrant	terbutaline ([M+H] 226.1443), 100 μg/mL aspartame ([M+H] 295.1294), 1000 μg/mL buspirone ([M+H] 386.2556), 10 μg/mL loperamide ([M+H] 477.2309), 10 μg/mL leucine ekephalin ([M+H] 556.2771), 100 μg/mL amiodarone ([M+H] 646.0315), 10 μg/mL
Test Compounds	28 compounds, mass range 260 – 652 amu 333 μM in 1:1 CH_3OH:H_2O

Automatic Calibration:
•Spectral accuracy calibration run daily off-hours at night unattended using QC function within MassLynx™ OALogin software.
•Calibration based on mass accuracy and spectral isotope pattern of each calibrant using MassWorks™ software (>98% SA, <5mDa error).
•Pass/fail calibration report emailed to system administrator.

Sample Analysis:
•Chemist selects designated method within MassLynx™ OALogin.
•Chemist inputs sample information, including (if known) target molecular formula, list of expected elements, and/or target ion. See example login page:
•Sample analyzed automatically by MassWorks™ when run is complete.
•Ions evaluated for mass accuracy and spectral isotope pattern to determine molecular formulas.
•PDF report containing MassWorks™ results emailed to chemist, in addition to any customary MassLynx™ report.



Customizable Parameters:
•Specified by lab system administrator within configuration file.
•Calibration specifications: compounds, SA, mDa error.
•Formula confirmation specification: SA, mDa error.
•Unknown compound formula specifications: default elements, and element limits (when no elements specified by user), $C_{0\text{-}x}H_{0\text{-}x}N_{0\text{-}x}O_{0\text{-}x}S_{0\text{-}2}P_{0\text{-}1}Cl_{0\text{-}3}Br_{0\text{-}3}F_{0\text{-}3}$, X based on compound mass.
- Number of sample peaks to analyze.
- Number of formulas to propose for each ion.

Conclusions

•Automated molecular formula determination is robust and easy to use in an open-access environment using a single quadrupole instrument without the need for high resolution instrumentation (TOF, Orbitrap).

•Software is highly successful (96%) at confirming known molecular formulas.

•For unknown compounds, specification of target mass greatly increases the likelihood that the correct formula is identified.

•For unknown compounds, chemist knowledge of likely elements present helps achieve optimum results.

•Automated open access formula determination has been successfully used in discovery chemistry project support.

Acknowledgments

AstraZeneca Infection and Cancer Chemistry