LC/MS Unknown Identifications Using MSMS Libraries
Part VIII Identification of Unknowns with "Spectraless" Libraries

12/27/20

James Little
tvasailor@gmail.com
https://littlemsandsailing.com/
Kingsport, TN

- *Retired* Research Fellow, Eastman Chem. Co.*
- *42 years experience unknown identification*
- *Now Consultant, MS Interpretation Services*
- *Specialties: EI GC-MS, LC-MS/MS, Chemical Ionization, Accurate Mass, Derivatization, MS library management, SciFinder, Chemspider, Surfactant ID, NMR, GC-IR, organic synthesis, matrix ionization effects, etc.*

Eastman Chemical Company, Main Site, Kingsport, TN
50 Manufacturing Sites Worldwide, ~14,500 Employees

>50 Mass Specs Networked
Worldwide

* https://en.wikipedia.org/wiki/Eastman_Chemical_Company

Table of Contents

Topic **Slide #**

- Introduction of Presenter, James Little 1
- Table of Contents 2
- Series Topics 3
- Introduction to Approach 4
- Origin of Term "Known Unknowns" 5
- Overall Process Including NIST Library Searches of EI and Tandem Spectra 6
- ChemSpider Internet Searches 7
- CAS Registry Using SciFinder and STN Express.....8
- Summary Comparing ChemSpider and CAS Registry Approaches 9
- Example 1: SciFinder Search of CAS Registry for UV Additive in Polymer 10
- Example 2: SciFinder Search of CAS Registry for More Obscure Compound 13
- Example 3: ChemSpider Identification of Additive in Polymer by Monoisotopic Mass ..14
- Live Demo YouTube Presentation 16
- Presentation References (Internet Links) 17
- Acknowledgements.....18

LC/MS Unknown Identifications Using MSMS Libraries

Part I: Overview of Software and User Customized Configurations
Part II: NIST MSMS Search Software and Libraries
Part III: More Detailed Discussion of MSMS Hybrid Search
Part IV: Importing MSMS Spectra
Part V: NIST Structure Searches
Part VI: MS Interpreter Correlation of Substructure to MSMS Ions
Part VII: Using and Creating Other MSMS Libraries
Part VIII: Identification of Unknowns with "Spectraless" Libraries
Part IX: Using Agilent MassHunter with NIST MSMS Search

Part VIII Identification of Unknowns with "Spectraless" Libraries[8-19]

Introduction

Accurate mass LC-MS^n

ChemSpider The free chemical database

(174 M entries)

SCIFINDER A CAS SOLUTION

(67 M entries)

⇨ ***Candidates*** ⇨ ***list narrowed*** ⇨ ***Confirmed***

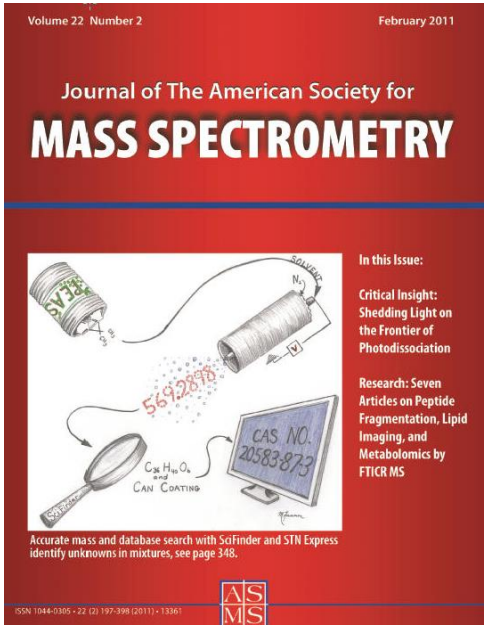
Overview:

- Need to identify compounds ***not found*** in Commercial or "Crowd-Sourced" Libraries
- Compounds can be ***readily identified*** with accurate mass LC-MS^n data and large databases
- Most useful databases: Scifinder[12-15] (CAS Registry) and ChemSpider[16-18] (Royal Society of Chemistry)
- These are ***primarily*** "Spectraless" databases
- ***Candidate*** compounds are ***found*** by searching either molecular formulae and/or accurate mass (+/- error)
- Candidates ***list narrowed*** by no. of associated references, key words, substructure, etc.
- ***Confirmed*** by interpreting MSMS spectrum, no. of exchangeable protons, UV-VIS spectrum, relative retention time, micro-synthesis of mixture, purchasing, etc.

Origin of Term "Known Unknowns"

- Compounds whose identities are ***known*** in the world, but ***unknown*** to us
- Term "known unknowns" ***stolen by us*** from Rumsfeld[7]

> ". . . there are *known knowns*; there are things we know we know. We also know there are *known unknowns*; that is to say we know there are some things we do not know. But there are also *unknown unknowns* -- the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones."
>
> Donald Rumsfeld, weapons of mass destruction in Iraq News Briefing, Feb. 12, 2002



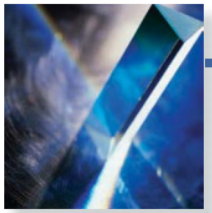
Reference[14]

Part VIII Identification of Unknowns with "Spectraless" Libraries

Overall Process Including NIST Library Searches of EI and Tandem Spectra

- The SciFinder/ChemSpider is ***part*** of our overall search
- Total approach shown in flowchart below (**1**)
- Discussed in LCGC Article[9] (**2**)

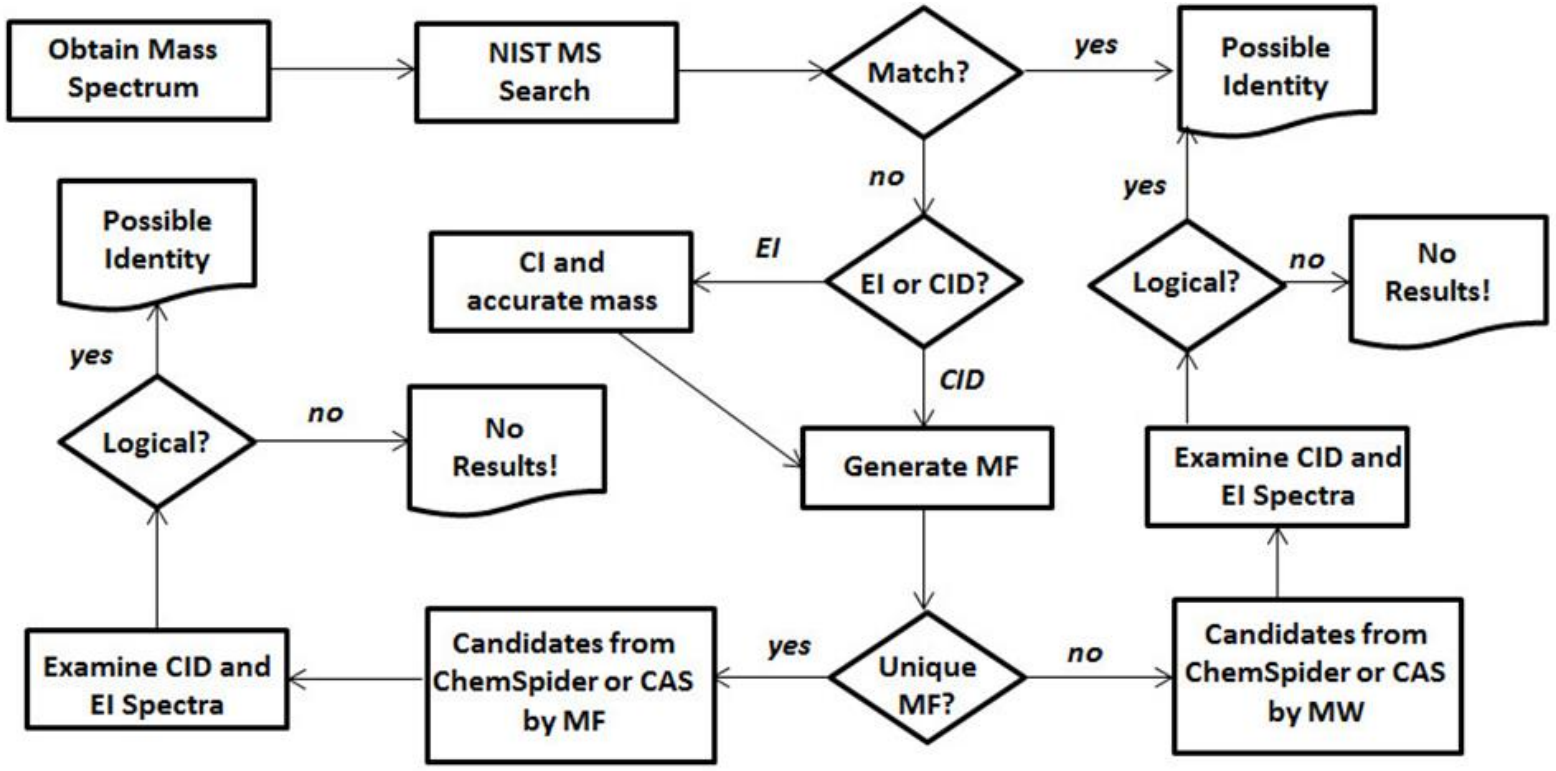
114 LCGC NORTH AMERICA VOLUME 31 NUMBER 2 FEBRUARY 2013 www.chromatograpl



MS – THE PRACTICAL ART

Identifying "Known Unknowns" in Commercial Products by Mass Spectrometry **2**

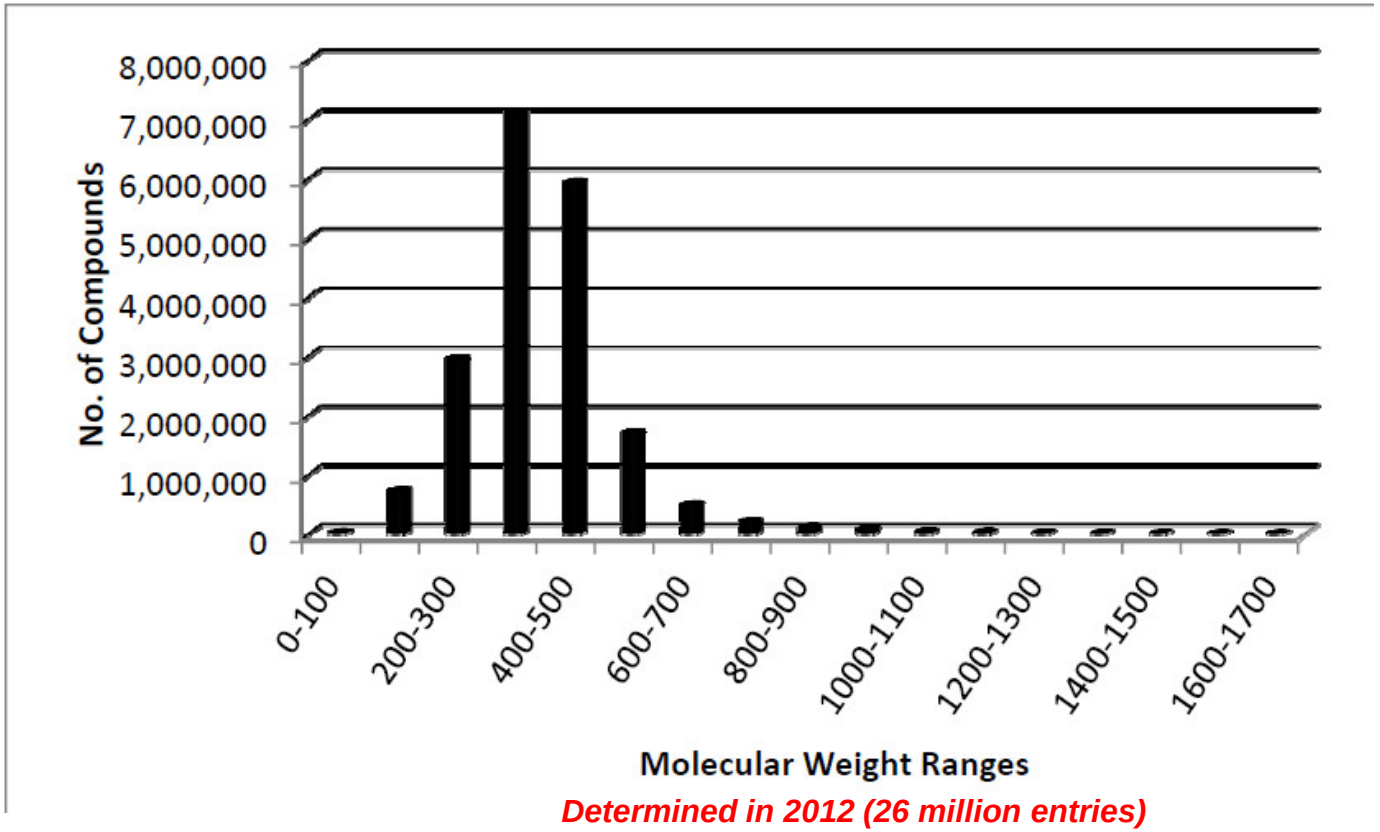
1 Simplified Flowchart for Identifying "Known Unknowns"



MF = molecular formula; *MW* = molecular weight; *CID* = collision-induced dissociation; *CI* = chemical ionization; *EI* = electron ionization

ChemSpider Internet Searches

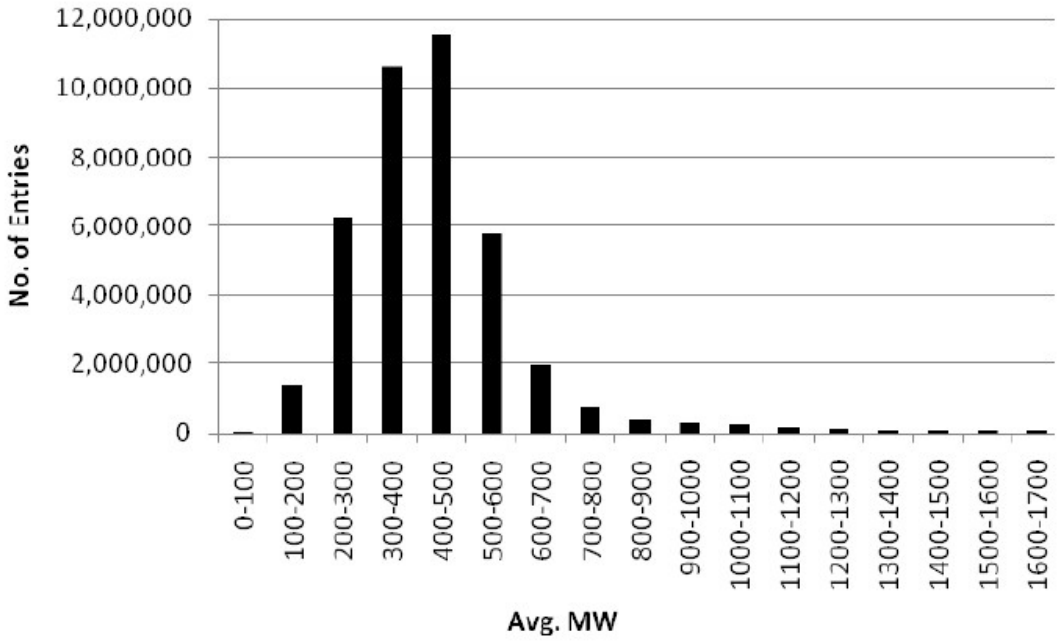
- ChemSpider has ***67 million*** structures from 100's of data sources
- ***Free*** service with internet interface offered by Royal Society of Chemistry
- ***Narrow*** candidate list by sorting in descending order by the number of associated references
- ***Best method*** is to use molecular formula at lower molecular weights
- ***Better*** *to search by monoisotopic mass at molecular* **weights>600** *because difficult to narrow to one specific* formula in that range



Determined in 2012 (26 million entries)

CAS Registry Using SciFinder and STN Express

- CAS Registry with SciFinder has 174 million unique organic and inorganic chemical substances
- Fee based service from American Chemical Society
- ***Narrow*** candidate list by number of references, substructure, key words from sample history, etc.
- ***Best method*** is to use molecular formula
- ***Does not*** search by monoisotopic mass, only molecular weight using STN Express
- STN Express is a command base interface, SciFinder much easier to use
- Error for latter is 4-5 times greater



Average MW distribution of compounds in CAS Registry (Dec. 2009) determined by s /mw command in STN Express.

Summary Comparing ChemSpider and CAS Registry Approaches

Search Approach	Pros	Cons
ChemSpider	-free via internet -good user interface -automation by instrument manufacturer using Web API (Application Program Interface) -ability to search monoisotopic mass	-smaller No. of entries (~26 M) and references -can't refine by key word
CAS Registry with SciFinder or STN Express	-larger No. of entries (~65 M) and references -refine by key word -good SciFinder user interface for MF searches on internet	-fee charged -no API available for instrument manufacturer automation -no ability to search by monoisotopic mass, ***only*** MW with ***complicated*** STN Express interface

Example 1: SciFinder Search of CAS Registry

Search CAS Registry with SciFinder by MF: UV Additive in Polymer

- Request to Identify UV additive in polymer
- Found in LC chromatogram by characteristic UV spectrum
- MF by accurate mass and isotope pattern

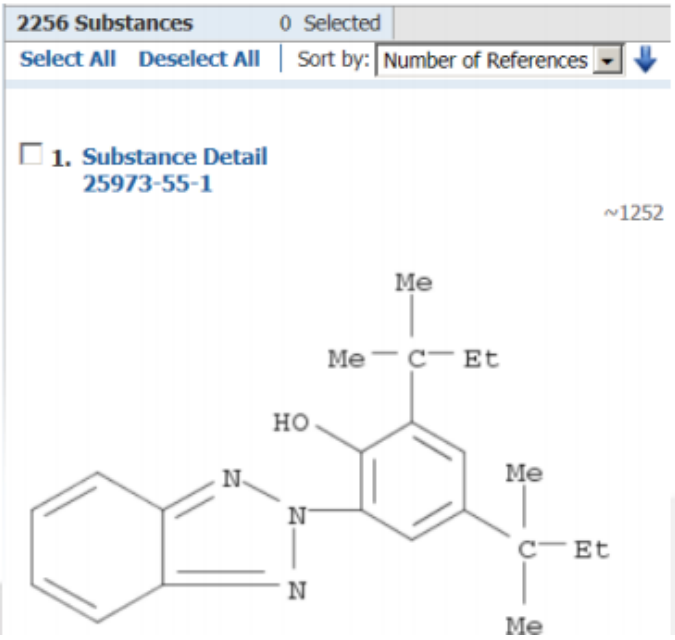
~65,000,000 → 2,256 → 2,059 → #1 1,252 Ref, #2 20, #3 19

CAS Registry

Get Substances $C_{22}H_{29}N_3O$

Sort by Number of References

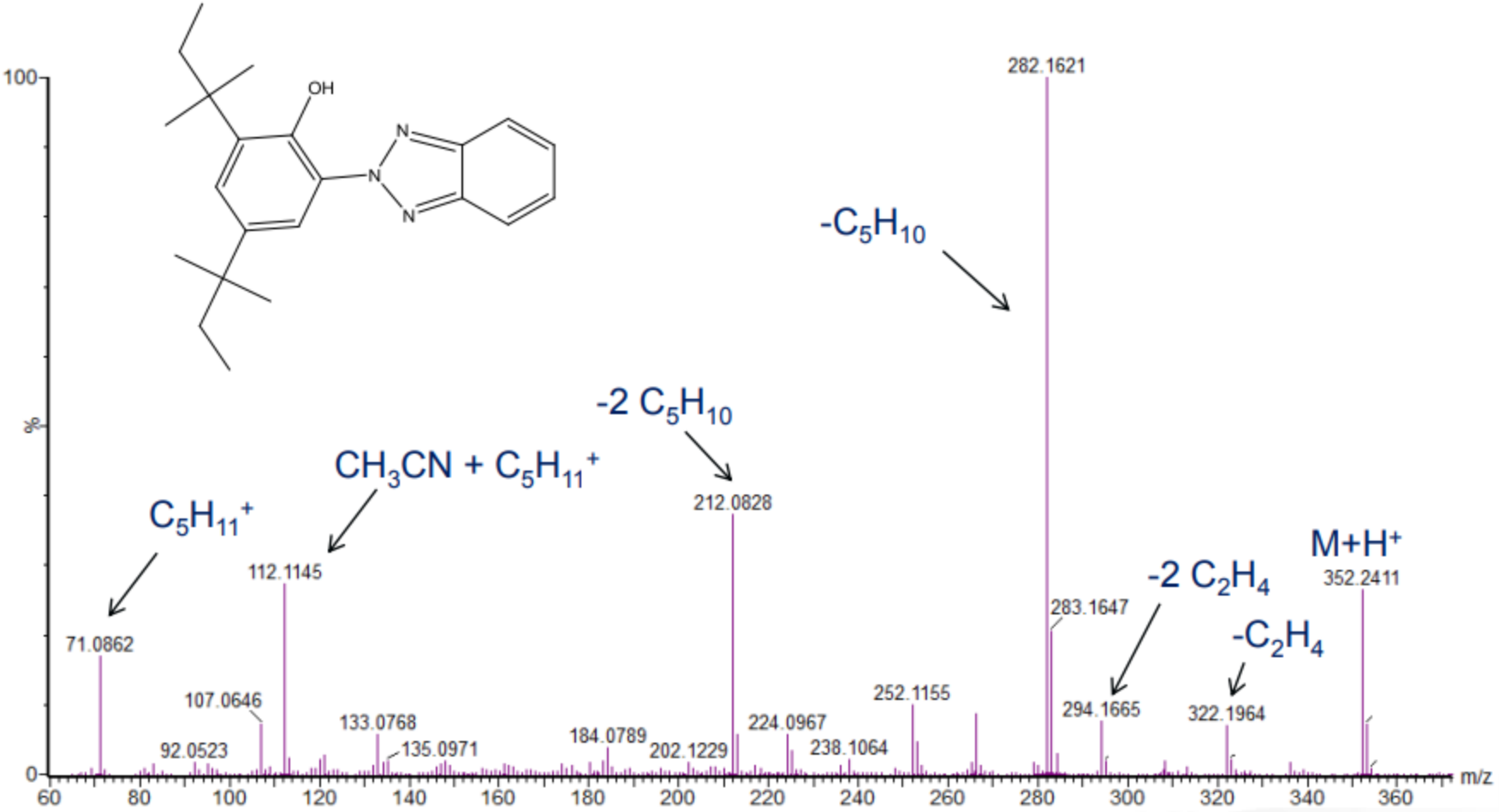
#1 Correct



Example 1: SciFinder Search of CAS Registry

Confirmation of Candidate Structure

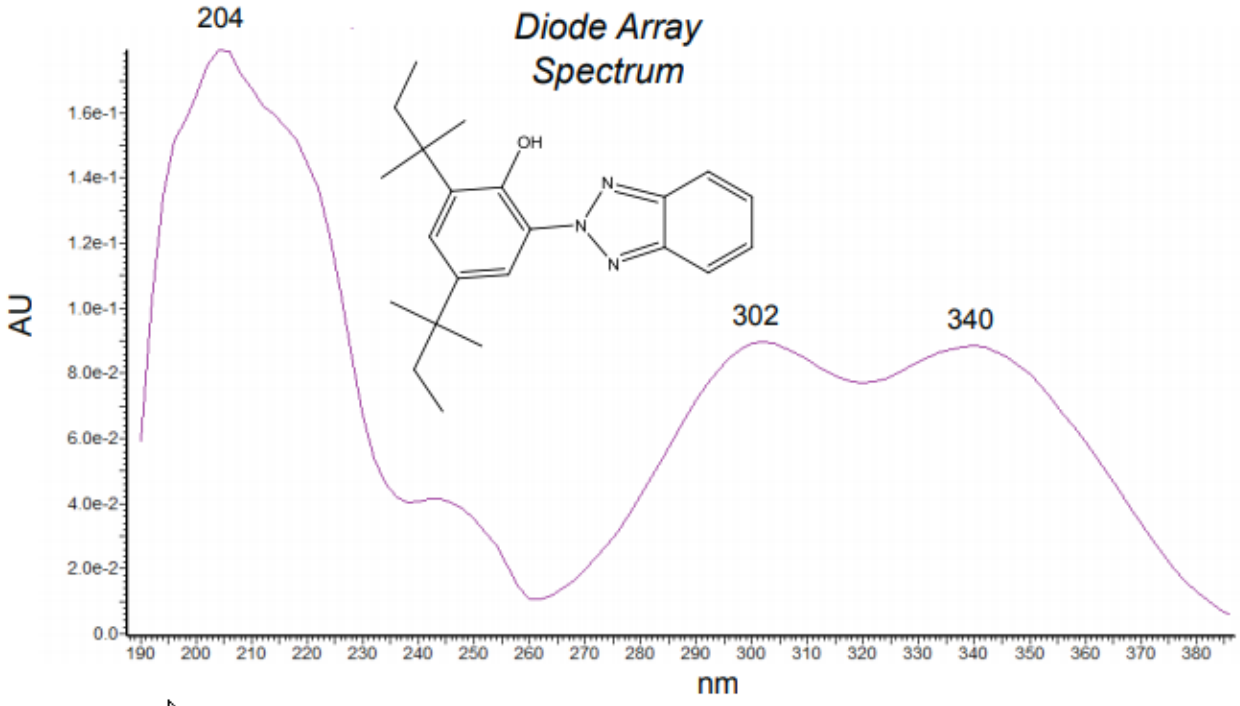
- Interpretation of MS/MS Spectrum



Example 1: SciFinder Search of CAS Registry

Confirmation of Candidate Structure (continued)

- UV spectrum characteristic of UV Absorber



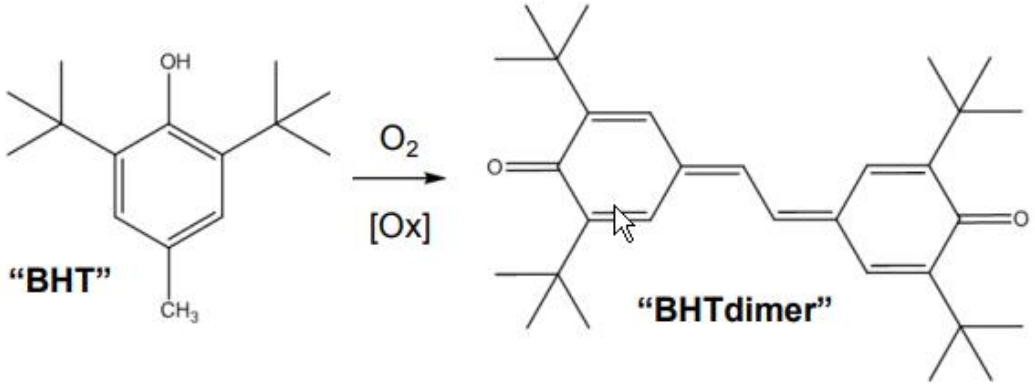
- One exchangeable proton by deuterium exchange *via infusion* in solvent mix with *minimal* D_2O
- Ultimately confirmed by purchased standard

Example 2: SciFinder Search of CAS Registry

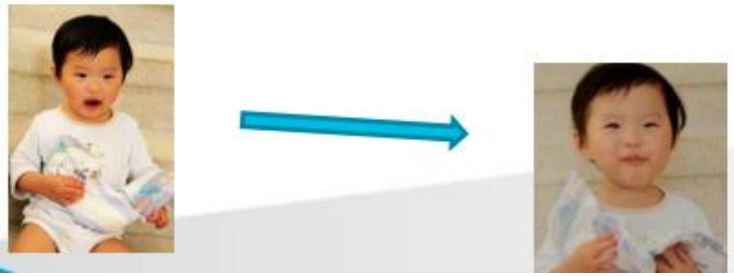
MF Search Refined with Key Word for More Obscure Compounds

- Identification of *greenish yellow* species in *baby diaper* adhesive

~65,000,000	→	264	→	482	→	1 Ref.
CAS Registry		Get Substances $C_{30}H_{42}O_2$		Get *all* references for *all* substances		Refined *all* references with "greenish yellow"



- Ironic, BHT included to stabilize, led to color from oxidation

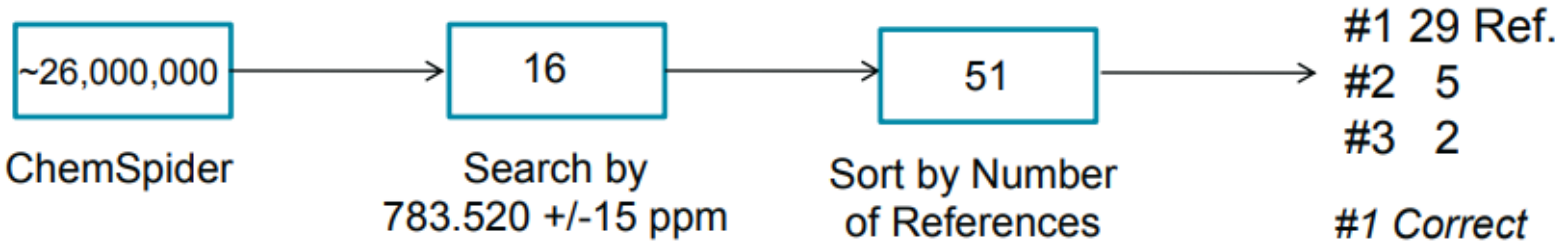


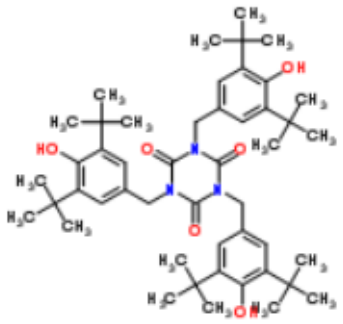
Problem Solved!

Example 3: ChemSpider

Search by Monoisotopic Mass: Antioxidant in Polymer

- Identification of additive in polymer
- Search ChemSpider by monoisotopic mass

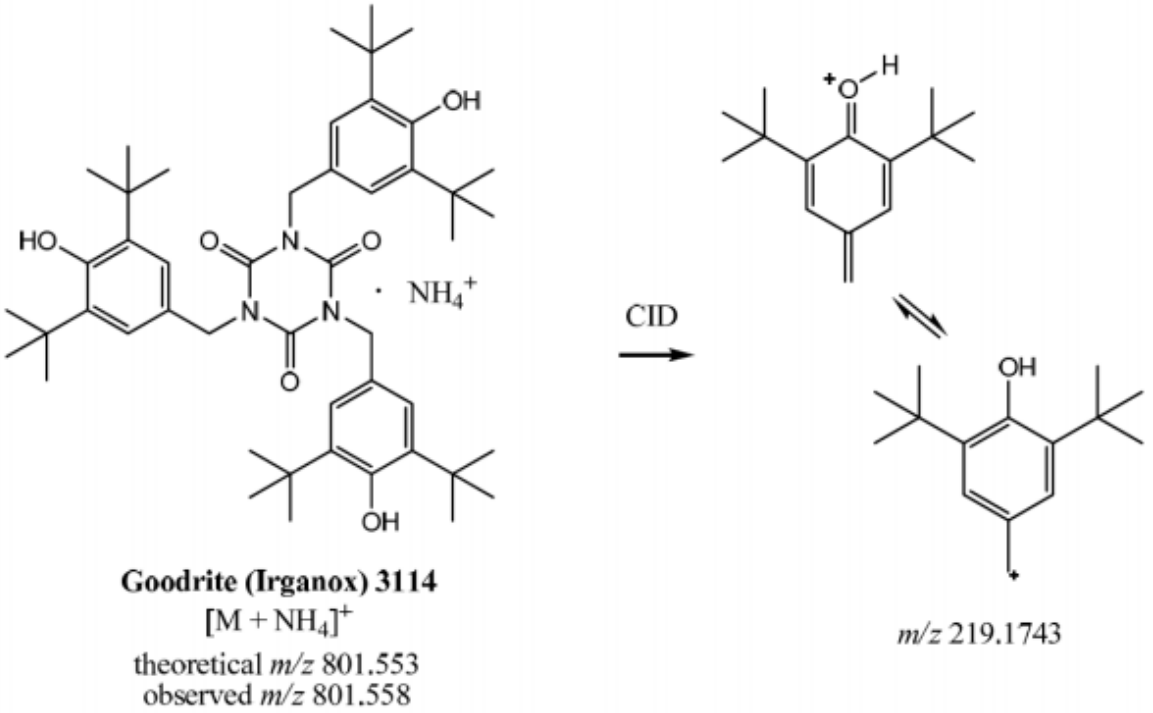


ID	Structure	Molecular Formula	Molecular Weight	# of Data Sources	# of References
84061		$C_{48}H_{69}N_3O_6$	784.078	17	29

Example 3: ChemSpider

Confirmation of Antioxidant Candidate Structure

- Initially confirmed by MS/MS fragmentation pattern



Goodrite (Irganox) 3114
$[M + NH_4]^+$
theoretical *m/z* 801.553
observed *m/z* 801.558

m/z 219.1743

- 3 exchangeable protons by infusion in solvent mix with D_2O
- Ultimately confirmed by purchased standard

Live Demo on YouTube
LC/MS Unknown Identifications Using MSMS Libraries
Part VIII Identification of Unknowns with “Spectraless” Libraries

Acknowledgements

- **Eastman:** Bill Tindall, Kent Morrill, Curt Cleven, Adam Howard, Jean Coffman, Mike Ramsey, Sen Li
- **ETSU School of Pharmacy:** Stacy Brown
- **CAS:** Anthony Machosky
- **Waters Corporation:** Jim Lekander
- **Agilent Technologies:** Mike Scott
- **Art Work for Journal Cover:** Minta Fannon
- **NIST:** Steve Stein
- **NIST Contractor:** David Sparkman