

Library Searching Without Libraries

Daniel L. Sweeney, Ph.D.

MathSpec, Inc.



EI Libraries Exist (Electron Ionization) for GCMS

NIST has 163000 compounds

Wiley7/NIST 317000 compounds

Pallisades 437000 compounds

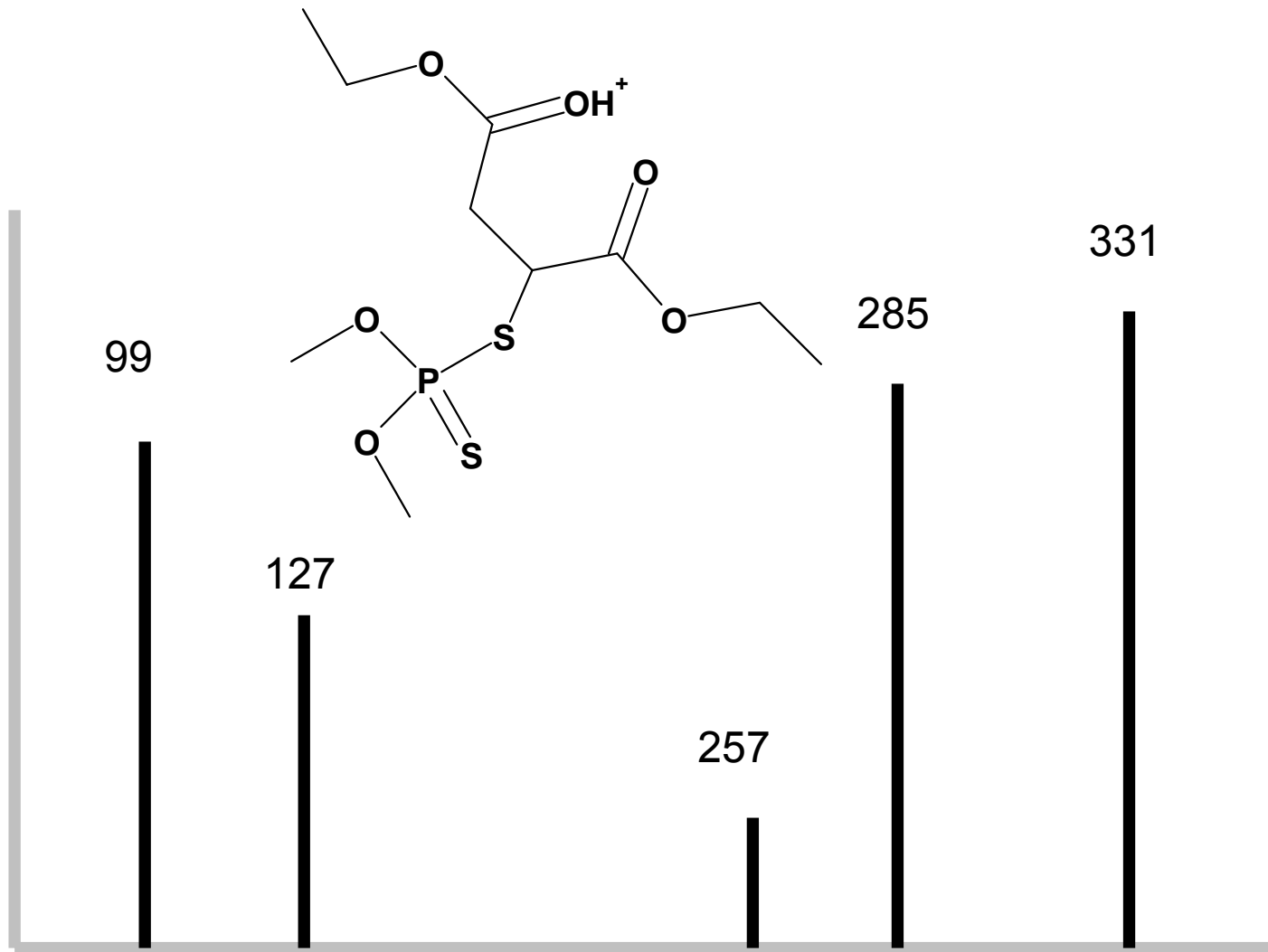
Only Small MS/MS Libraries Exist for LCMS

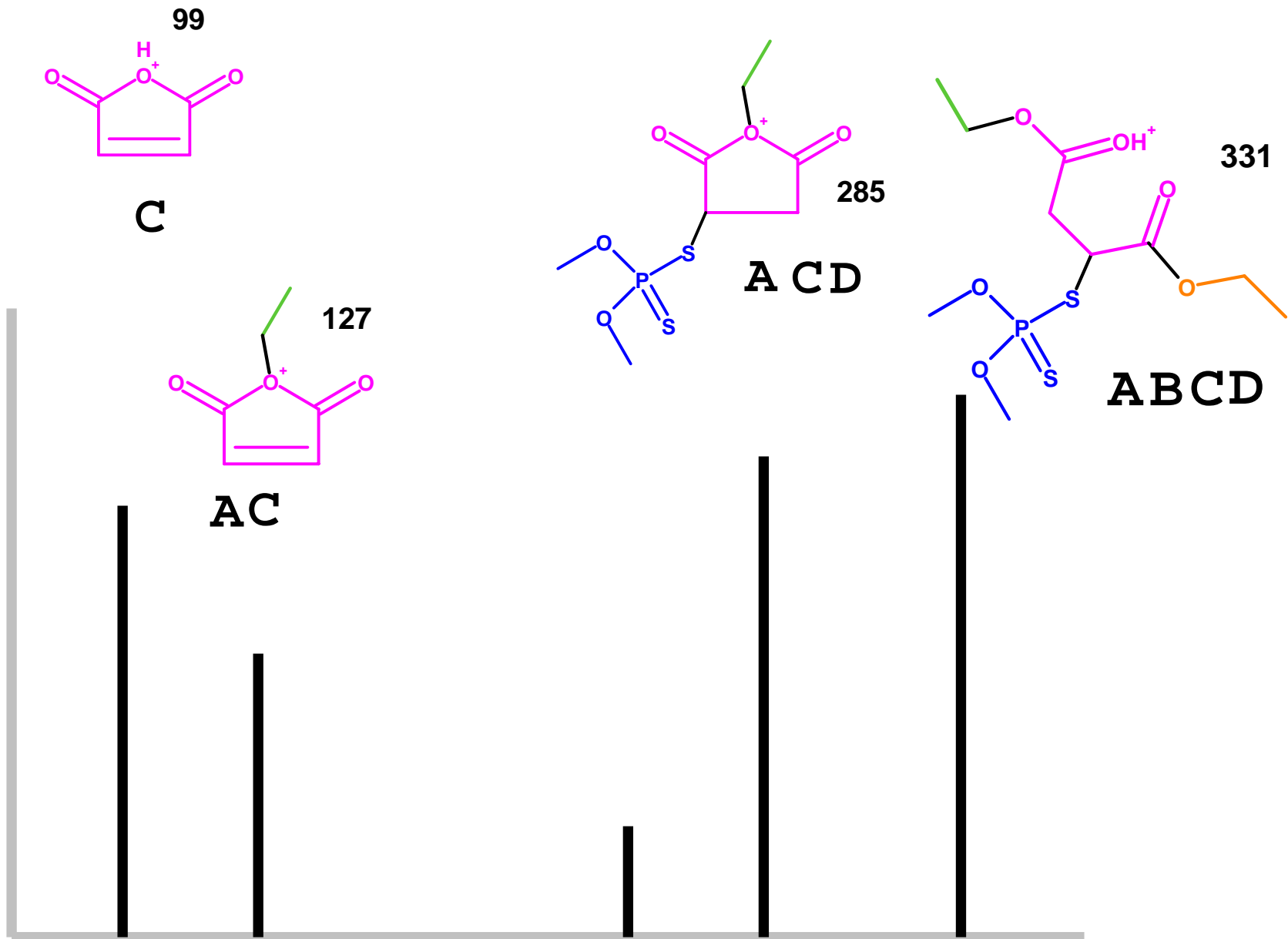
NIST has a library of about 1700 compounds

MS/MS/MS Libraries have been suggested

Applications where MS/MS mass spectral libraries are of interest

1. Environmental Samples
2. Forensics/Toxicology
3. Endogenous Metabolites
4. Natural Products
5. Binding Studies





A

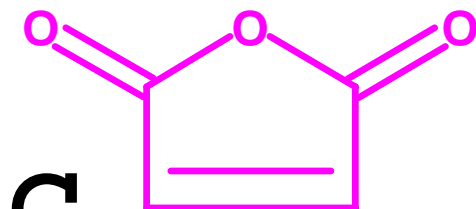


28



B

46

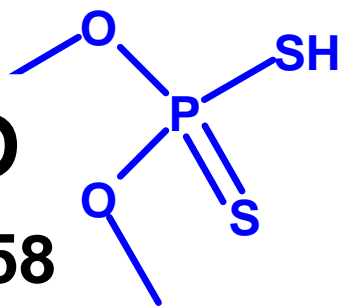


C

98

D

158



C₂H₂

28 A

C₂H₆O₁

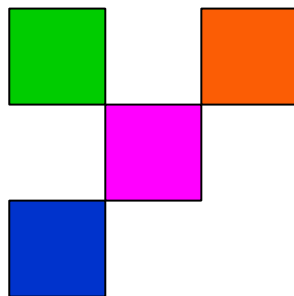
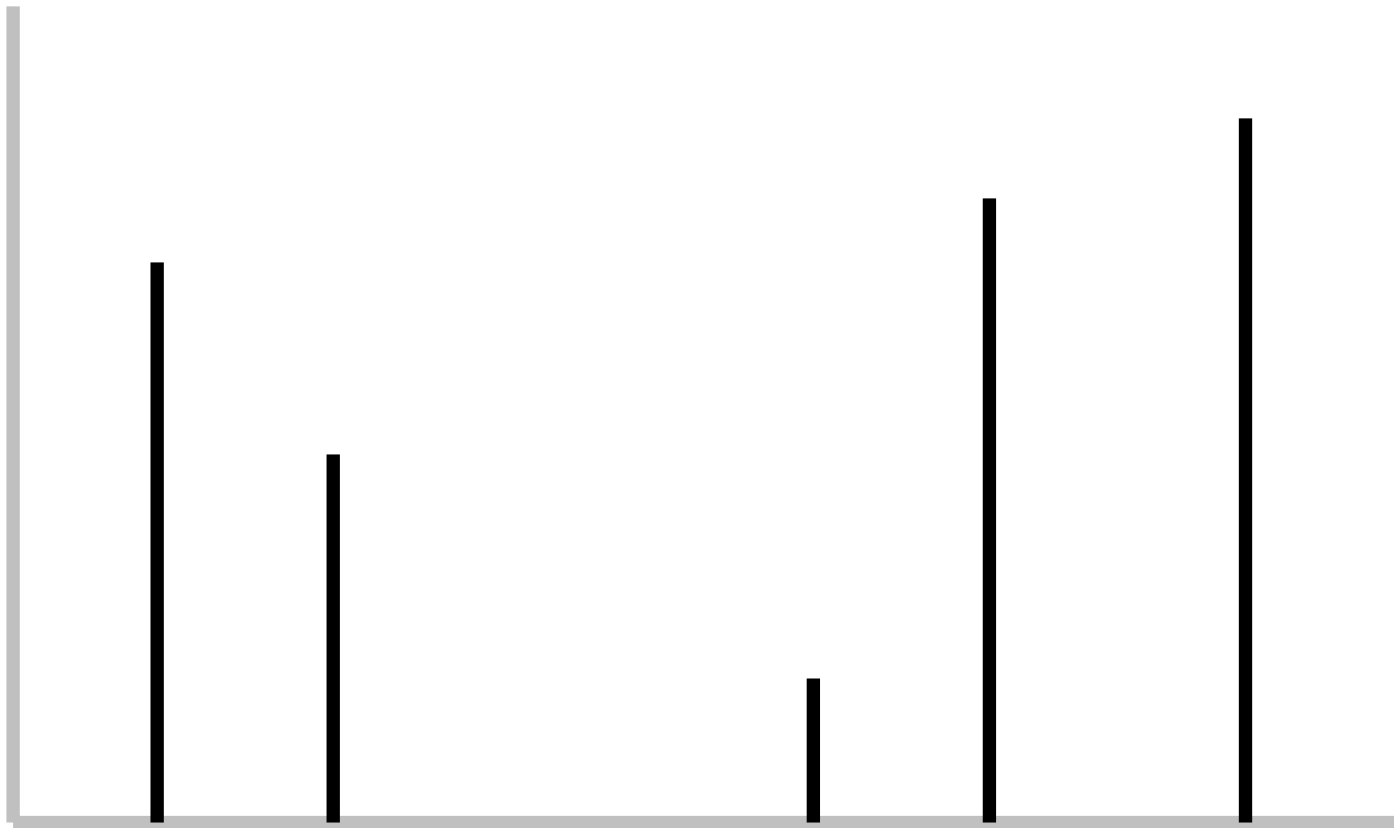
46 B

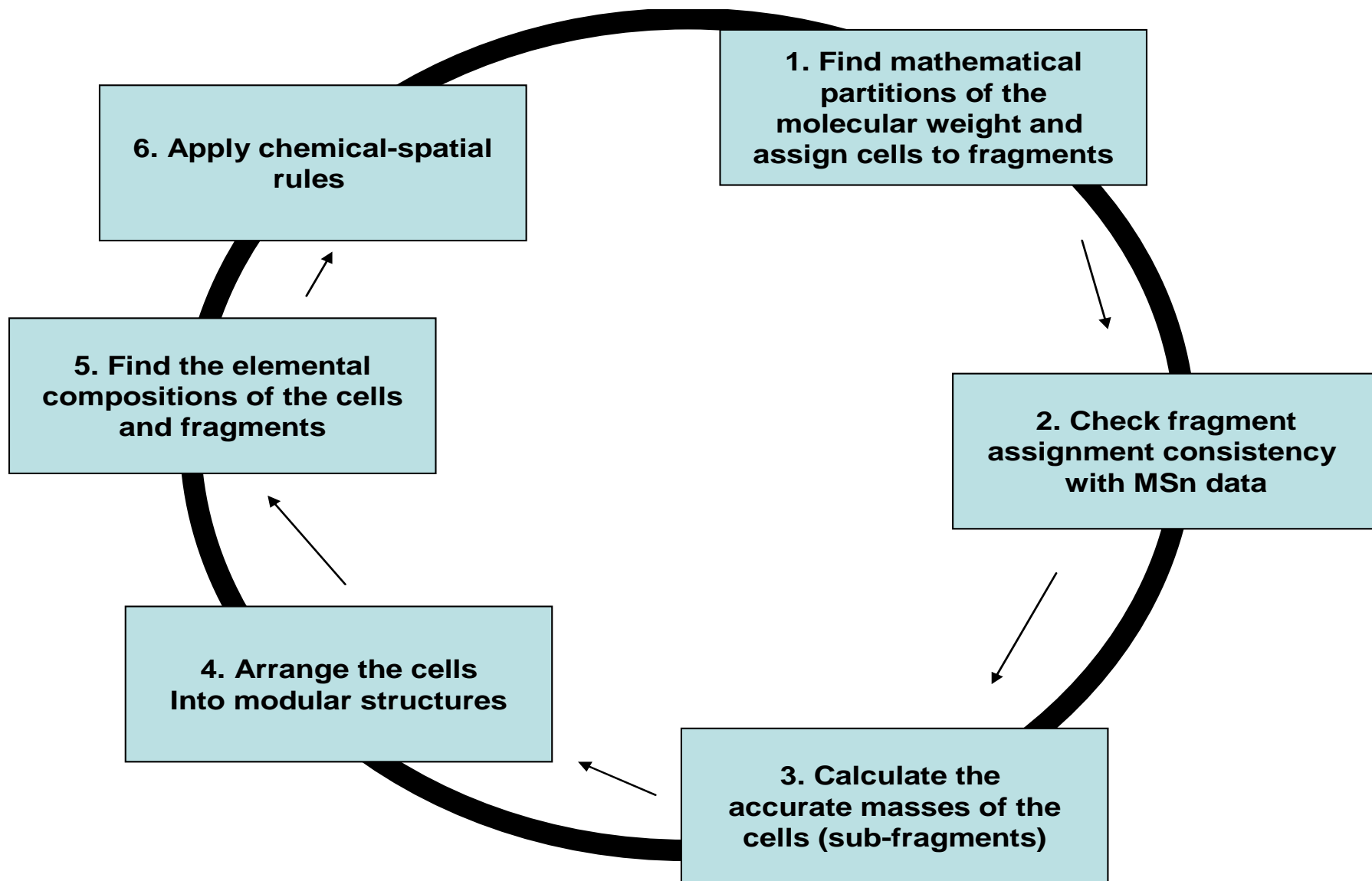
C₄H₂O₃

98 C

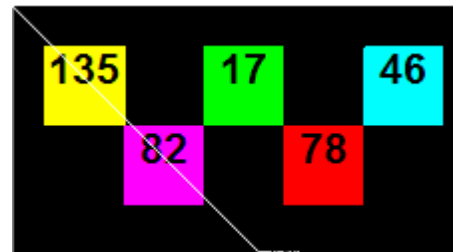
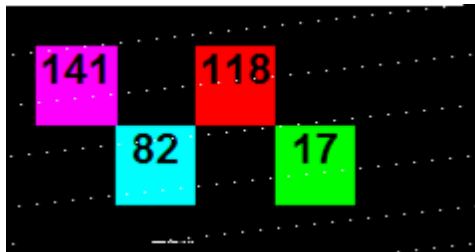
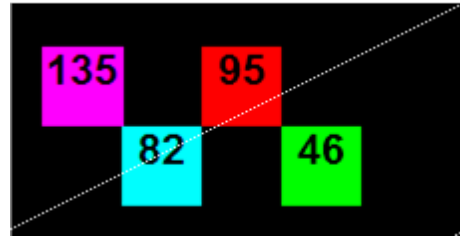
C₂H₇O₂S₂P

158 D





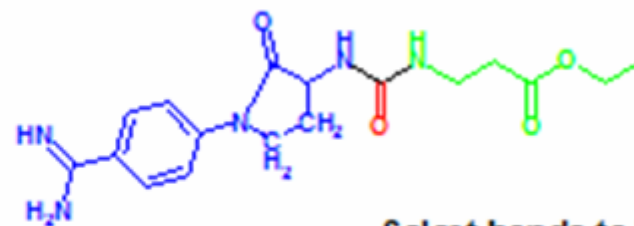
“Partition”



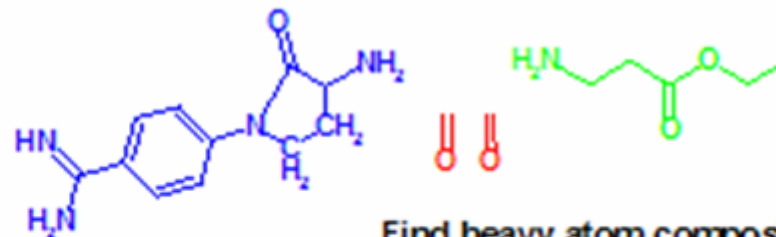
Partition saves the best ten modular structures



Find heavy atom compositions of the cells



Select bonds to break



Find heavy atom compositions of the groups



Eliminate redundant groups

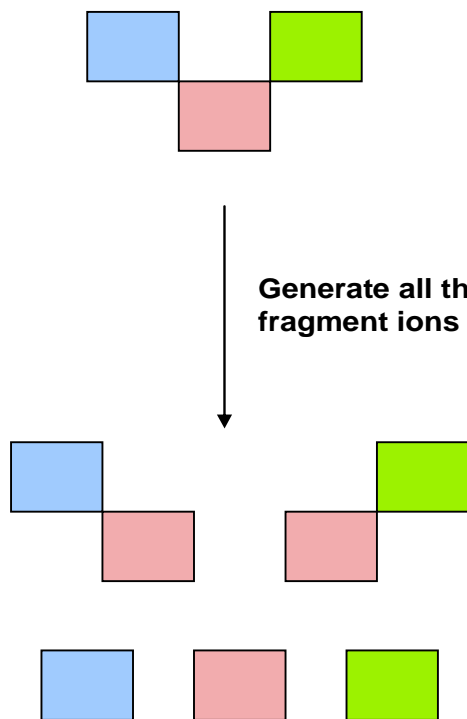


Compare heavy atom compositions

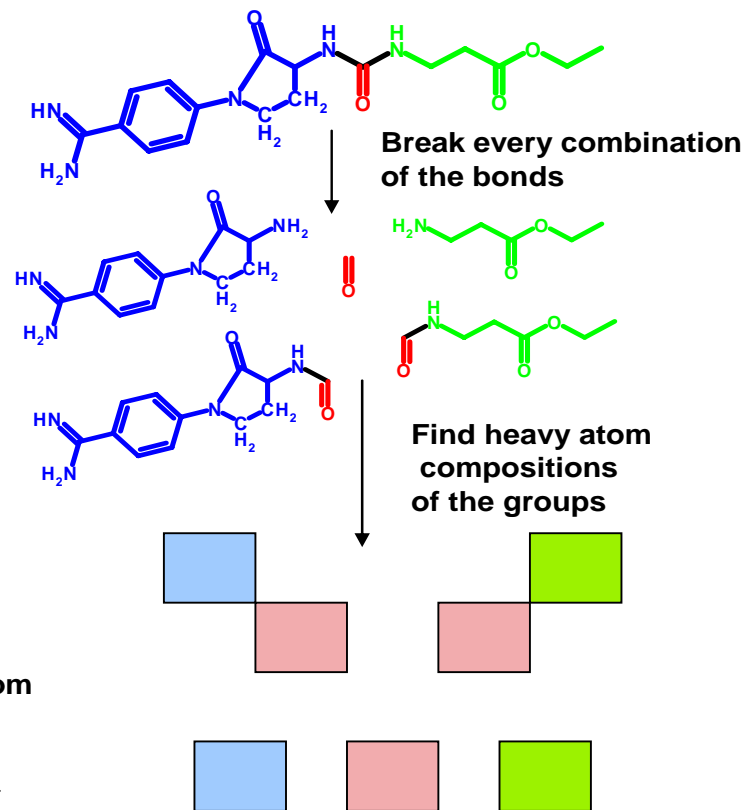


Compare heavy atom compositions of subfragments (cells)

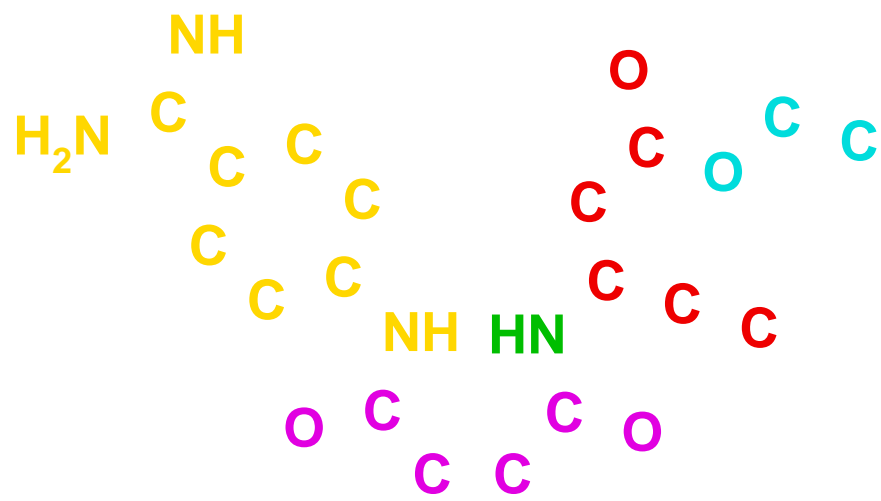
Modular structures derived by partitioning



Computerized molecular structure



Databases can be searched by comparing the set of modular structures (left) derived from the mass spectral data to molecular structures (right) in the database.

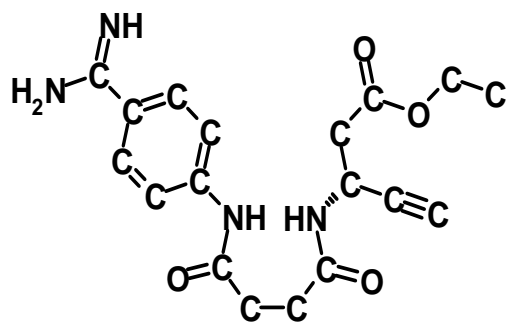


Xemilofiban: Searching PubChem

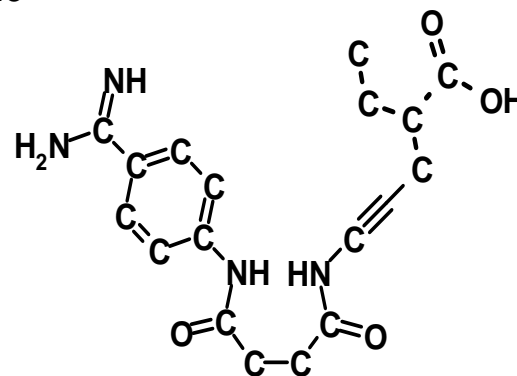
840 compounds within 7.5 ppm of exact mass
671 compounds with the same formula

Rational Numbers Search Results

Score	Compound	
74	PubChem ID# 60957	xemilofiban
74	PubChem ID# 10451025	enantiomer of xemilofiban
74	PubChem ID# 10361038	racemic xemilofiban
66	PubChem ID# 9946860	
6	PubChem ID# 8754555	
6	PubChem ID# 8145626	

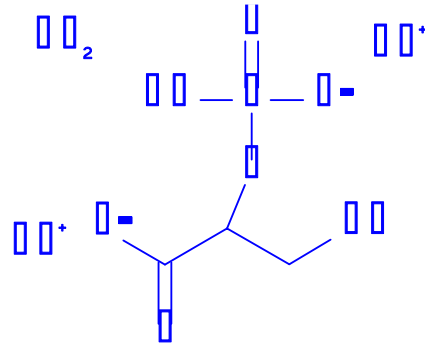


60957 (□□□ □□□□□)

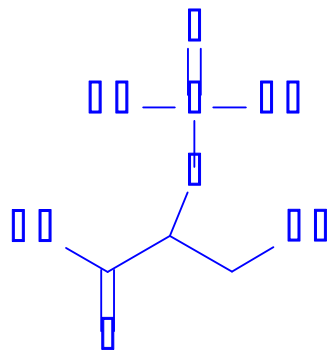


9946860

D-(+)-2-PHOSPHOGLYCERIC ACID SODIUM HYDRATE, MFCD00150613



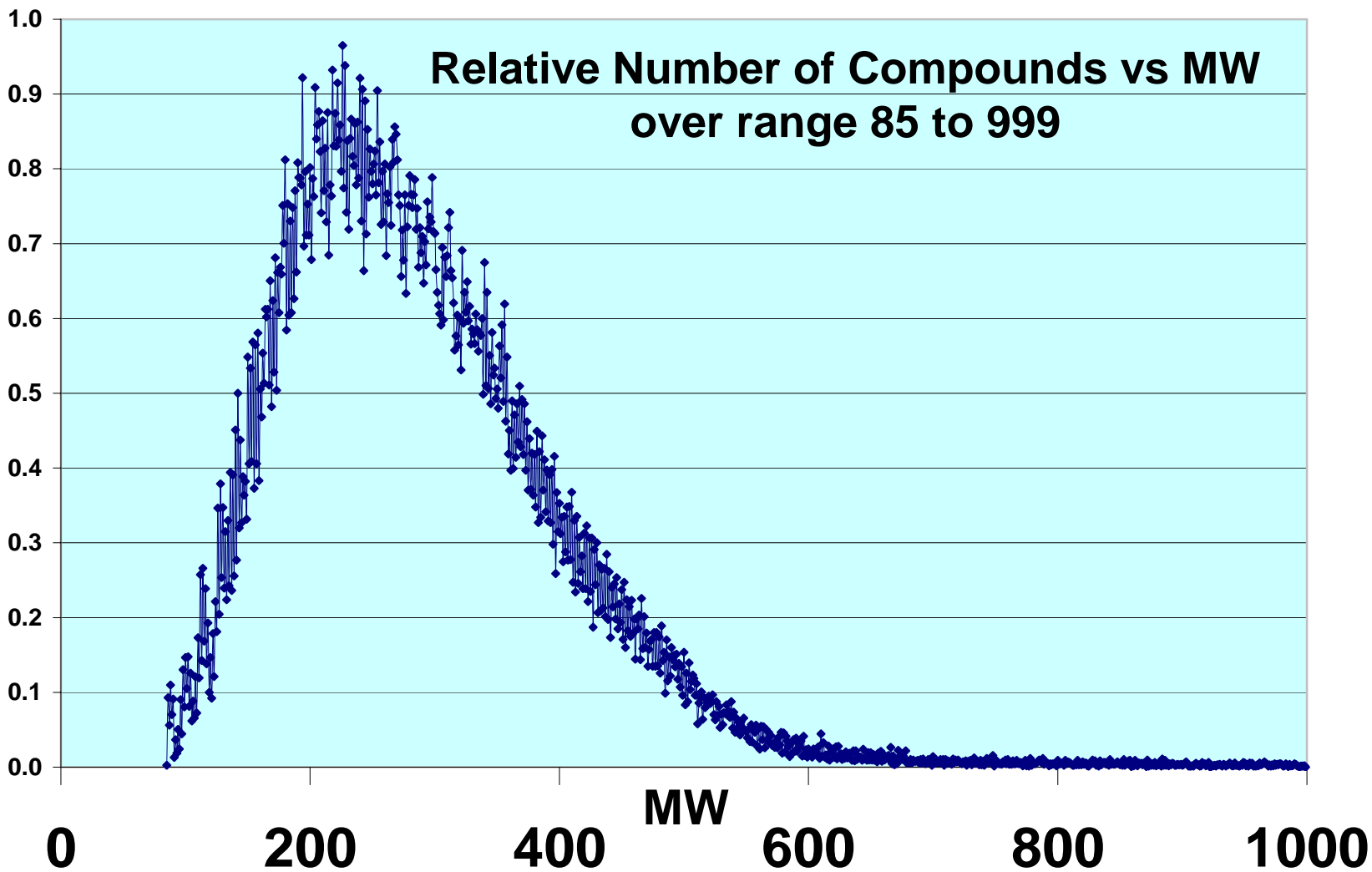
CC(O)C(=O)OP(=O)([O-])[O-].[Na+].[Na+].O.O



LCMS expected mass:
185.993

Advantages of Using an Index of the MDL® Available Chemicals Directory (ACD)

- Widely available, high quality database
- Over 300000 compounds indexed by exact mass (C,H,N,O,S, P,F,Cl, and Br) with MW between 85 and 999
- Indexed by “LCMS expected masses”: over 40000 compounds converted from salts/hydrates/quaternaries
- Search results can be imported directly into DiscoveryGate™
- Virtually all matching compounds can be conveniently purchased for subsequent comparison to unknowns



**MDL® Available Chemicals Directory : Mass distribution
of indexed compounds containing C,H,N,O,S,P,F,Cl, and Br**

Fluconazole

70.0392	12		
121.0442	2	307.1	100
127.0349	12	308.1	16.9
139.0354	3	309.1	1.5
141.0522	11		
151.0351	9		
169.0457	70		
200.0616	4		
220.0679	85		
238.0793	100		
289.1010	3		
307.1113	16		

Fluconazole “FragSearch” Results

Number of compounds in the Available Chemicals Directory with a similar molecular weight:
277

Cell Set	Formula
1	C13H12N6O1F2
2	C13H12N6O1F2
3	C13H12N6O1F2
4	C13H12N6O1F2
5	C13H12N6O1F2
6	C13H12N6O1F2
7	C13H12N6O1F2
8	C13H12N6O1F2
9	C13H12N6O1F2
10	C13H12N6O1F2
11	C13H12N6O1F2
12	C13H12N6O1F2
13	C13H12N6O1F2
14	C13H12N6O1F2
15	C13H12N6O1F2
16	C13H12N6O1F2
17	C16H11N6F1
18	C13H12N6O1F2
19	C13H12N6O1F2
20	C13H12N6O1F2
21	C13H12N6O1F2
22	C16H11N6F1

Rational Numbers Search Results

Score	Compound
126	MFCD00274549

The maximum possible score for this search (assuming all formulas above are identical):
638

DiscoveryGate® | MDL® Database Browser

[queries](#) **results** [reports](#)

[copy to report](#) [export records](#) [export selected citations](#) [page setup](#) [print](#) [save](#) [refine query](#) [lists](#) [help](#) [logout](#)

Database:

MDL® Available Chemicals Directory

Find in History: [Next](#)

[Field Index](#) **History**

- Today's Searches
 - Search #1**
- Previous Searches
- Saved Searches
- Temporary Lists

[View selected records in another database](#)

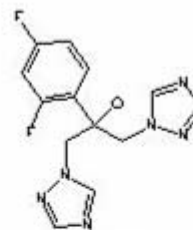
[Create List](#)

Pages: 1

[Select All](#) [Clear All](#)

Total Records: 1

MDL® Available Chemicals Directory



[Details](#)

Record#1

Pages: 1

[Select All](#) [Clear All](#)

Total Records: 1

[←](#) [→](#) [A-Z](#) [Print](#) [✖](#)

Browsing Database Results in the Grid View

Click [Details](#) to view the details of a specific substance or reaction. Scroll the data and click the data links that you want to view. To return to the result set, click [Return to Search Results](#).

DiscoveryGate® | MDL® Database Browser

queries results reports

copy to report export records export selected citations page setup print save refine query lists help logout MDL® Available Chemicals Directory

Database: MDL® Available Chem... ?

Find in History: [] Next

Field Index History

- Today's Searches
 - Search #1
- Previous Searches
- Saved Searches
- Temporary Lists

← → A-Z Pr

Browsing Database Results in the Single-Record View

Use the check box to select those results that you want to copy to a report, export, or view in another database. Then click the **Copy to Report** button, **Export Records** button, or **View selected records in another database** link. If you select a record and then want to deselect it, simply click the box again.

To **zoom** a substance, **double-click** it to open the structure editor, and then use the **Zoom tool**.

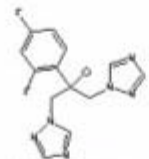
To use a structure as the basis of a new query, click **Use as Query** in the structure box.

Click **View selected records in another database** to transfer a set of records to another application. Two dialogs appear one after the other, one that asks you which records and another that asks you which application. These other

[Return to Search Results](#) [View selected records in another database](#) Record # 1 **Total Records: 1**

Also found in: [Beilstein](#) [CCR](#) [CMC](#) [DWPI](#) [MDRR](#) [MetaCore](#) [MetaDrug](#) [Metabolite](#) [OHS](#) [Patent](#) [PharmaPendium](#) [PubChem](#) [SCD](#) [Toxicity](#) [xPharm](#)
in: [MSDS](#) [Chemistry](#)

MDL® Available Chemicals Directory [Select all citations / Deselect all citations](#)



[Use as Query](#)

 Select current record

Available Data

Click on a link to add the information to this page

Set current view as default

[Substance](#) (1) [Suppliers](#) (32)

[Prices](#) (32) [Model](#) (1)

Substance [\(hide\)](#)

ACD Registry Number	108226
Availability	Large and small quantities
MDL Number	MFC00274549
CAS Registry Number	86386-73-4
Chemical Name and Synonyms :	
<ul style="list-style-type: none"> • 2,4-DIFLUORO-1-BIS(1H-1,2,4-TRIAZOL-1-YLMETHYL)BENZYL ALCOHOL • 2,4-DIFLUORO-ALPHA,ALPHA-1-BIS(1H-1,2,4-TRIAZOL-1-YLMETHYL)BENZYL ALCOHOL • A-(2,4-DIFLUOROPHENYL)-A-(1H-1,2,4-TRIAZOL-1-YLMETHYL)-1H-1,2,4-TRIAZOL-1-ETHANOL • AKOS 92108 • FLUCONAZOLE 	
Molecular Formula	C13 H12 F2 N6 O
Molecular Weight	306.275
Rule of Five	0
Computed partition coefficient (CLogP)	0.25
Molecular weight of largest fragment	306.275
Number of proton acceptors	7
Number of proton donors	1
Number of rotatable bonds (TDF)	5

[Top of Page](#)

[Return to Search Results](#) [View selected records in another database](#) Record # 1 **Total Records: 1**

Conclusions

Accurate mass fragmentation data can be used to generate modular structures.

Modular structures have enough information to compare to molecular structures.

Molecular structure databases can be indexed for LCMS and effectively searched with mass spectral fragmentation data.

Acknowledgements

**Carmen Nitsche, Mary Beth Walsh, and Yoshiko Fukui
at MDL Elsevier**

**Marty Duey, Ari Shamash and Alka Gupta at Sun
Microsystems**

References

1. Sweeney, D.L. Small molecules as mathematical partitions. *Anal. Chem.* **2003**, 75(20), 5362–73.
2. Wu, Q. Multistage accurate mass spectrometry: a “basket in a basket” approach for structure elucidation and its application to a compound from combinatorial synthesis. *Anal. Chem.* **1998**, 70, 865–72.
3. Thurman, E.M., Ferrer, I., and Zweigenbaum, J. High-Resolution and Accurate Mass analysis of Xenobiotics in Food. *Anal. Chem.* **2006**, 78, 6703-6708.
4. Watson, I.A.; Mahoui, A.; Duckworth, D.C.; Peake, D.A. A strategy for structure confirmation of drug molecules via automated matching of structures with exact mass MS/MS spectra. Proceedings of the 53rd ASMS Conference on Mass Spectrometry, June 5–9, 2005, San Antonio, TX.
5. Hill, A.; Mortishire-Smith, R. Automated assignment of high-resolution collisionally activated dissociation mass spectra using a systematic bond disconnection approach. *Rapid Commun. Mass Spectrom.* **2005**, 19, 3111–18.
6. Rourick, R.A.; Volk, K.J.; Klohr, S.E.; Spears, T.; Kerns, E.H.; Lee, M.S. Predictive strategy for the rapid structure elucidation of drug degradants. *Pharm. Biomed. Anal.* **1996**, 14, 1743–52.
7. Thompson, D.M., Richards, D.S., Fancy, SA, Perkins, G.I., Pullen, F.S., Thom, C. A comparison of accurate mass techniques for the structural elucidation of fluconazole *Rapid Commun. Mass Spectrom.* **2003**, 17, 2804.