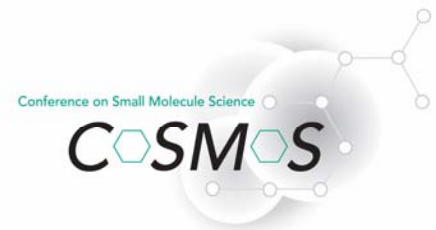


# Extracting Accurate Mass Information for Compound Identification from Complex Sample Matrices

Doug McIntyre  
Agilent Technologies



# The Power of Accurate Mass

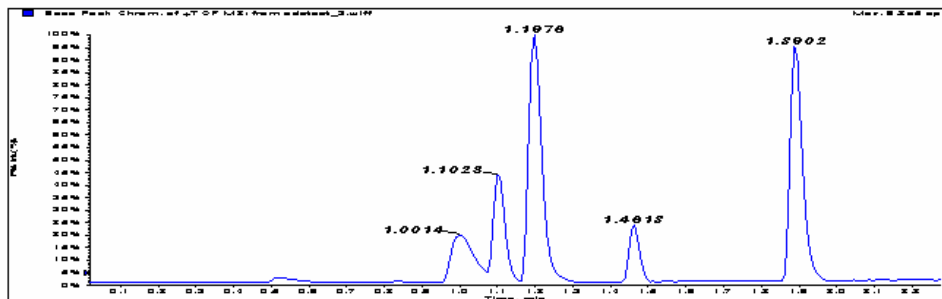
- Current API-TOF systems can measure mass with  $<3$  ppm mass error with a resolving power of 4000 to 15,000 depending on  $m/z$  value
- Since observed  $m/z$  values are indicative of molecular weight
  - This is sufficient to propose a manageable number of empirical formulas
  - Can be used to confirm presence of a suspected compound
  - Can perform compound database searching to identify target compounds

# Approaches to Compound Identification

- Forward Search
  - Integrate a chromatographic signal
  - Obtain spectrum from each peak
  - Use major  $m/z$  values to propose formulas or search database
- Reverse Search
  - Input formula, calculate MW, consider adducts, charge state
  - Extract appropriate ion chromatograms
  - Obtain spectrum from top of peak
  - Determine mass error

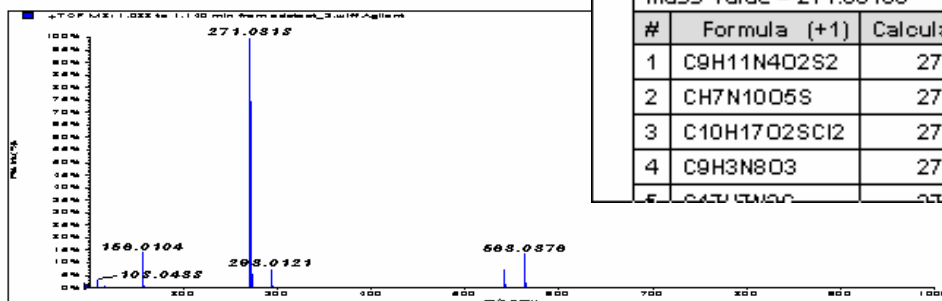
# Forward Search - Empirical Formula Generation

Sample#: adctest Sample Location: Sample Id: Operator:  
 Data File Name: C:\PE\_Sciex\_Data\Projects\adctest\Data\adctest\_3.wiff Acq Time: March 30 2005, 03:10:09 PM  
 Method: C:\Program Files\Agilent\TOF Software\demomethods\demo\_efg\_anm\mass\_list.xml



Period#: 1 Experiment#: 1

Peak#	Experiment#	Time	Area	Most Abundant Masses/scan
1	1	1.10	8.32861 E5	271.03183
2	1	1.20	2.17636 E6	279.09187
3	1	1.46	4.02119 E5	285.02045
4	1	1.89	1.96131 E6	311.081



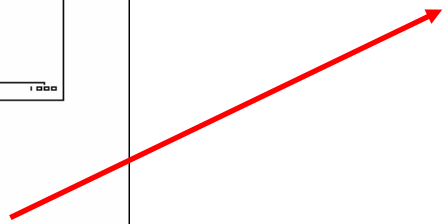
Peak#: 1 Experiment#: 1 Retention Time: 1.10 min

Mass Value = 271.03183

#	Formula (+1)	Calculated m/z	Error (mDa)	* Error (ppm)	DBE	Isotope Match	Intensity Match
1	C9H11N4O2S2	271.03180	0.03	0.1	6.5	994	236409
2	CH7N10O5S	271.03161	0.22	0.8	3.5	999	236409
3	C10H17O2SCl2	271.03208	-0.25	-0.9	1.5	714	236409
4	C9H3N8O3	271.03226	-0.43	-1.6	12.5	995	236409

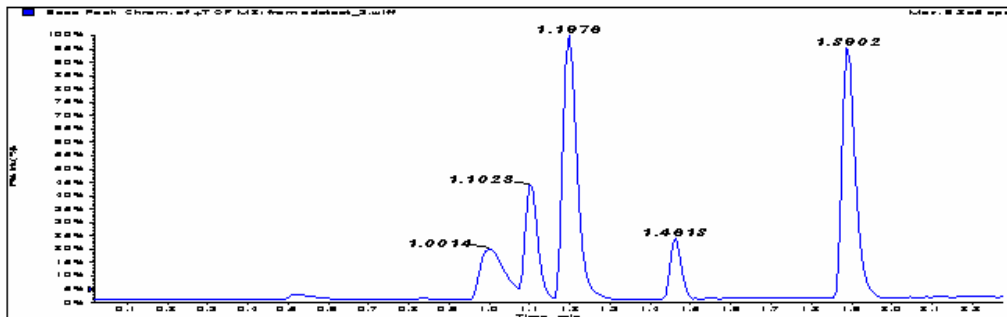
Mass Value = 271.03183

#	Formula (+1)	Calculated m/z	Error (mDa)	* Error (ppm)	DBE	Isotope Match	Intensity Match
1	C9H11N4O2S2	271.03180	0.03	0.1	6.5	994	236409
2	CH7N10O5S	271.03161	0.22	0.8	3.5	999	236409
3	C10H17O2SCl2	271.03208	-0.25	-0.9	1.5	714	236409
4	C9H3N8O3	271.03226	-0.43	-1.6	12.5	995	236409



# Forward Search - Database Search

Sample#: adctest Sample Location: Sample Id: Operator:  
 Data File Name: C:\PE Sciex\_Data\Projects\adctest\Data\adctest\_3.wiff Acq Time: March 30 2005\_03:10:09 PM  
 Method: C:\Program Files\Agilent\TDF Software\demomethods\demo\_db\_anm\mass\_list.xml

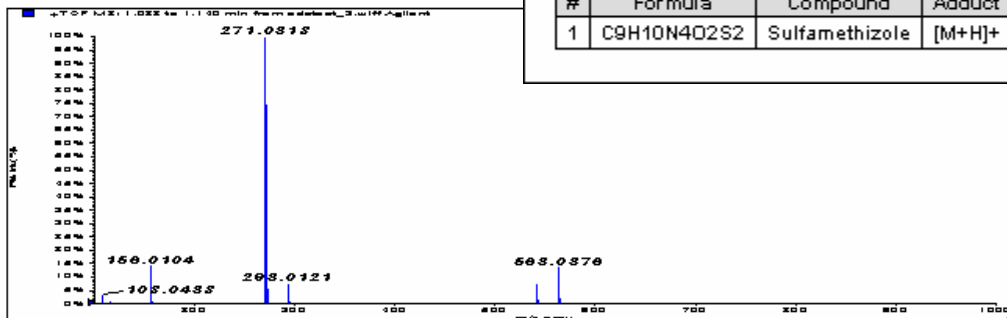


Period#: 1 Experiment#: 1

Peak#	Experiment#	Time	Area	Most Abundant Masses/scan
1	1	1.10	8.32861 E5	271.03183
2	1	1.20	2.17636 E6	279.09187
3	1	1.46	4.02119 E5	
4	1	1.89	1.96131 E6	

Mass Value = 271.03183

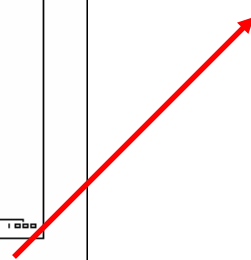
#	Formula	Compound	Adduct	Mass	Error (mDa)	* Error (ppm)	Ret. Time Error	Description
1	C9H10N4O2S2	Sulfamethizole	[M+H] <sup>+</sup>	271.03178	0.05	0.2	0.00	Sulfa mix #4



Peak#: 1 Experiment#: 1 Retention Time: 1.10 min

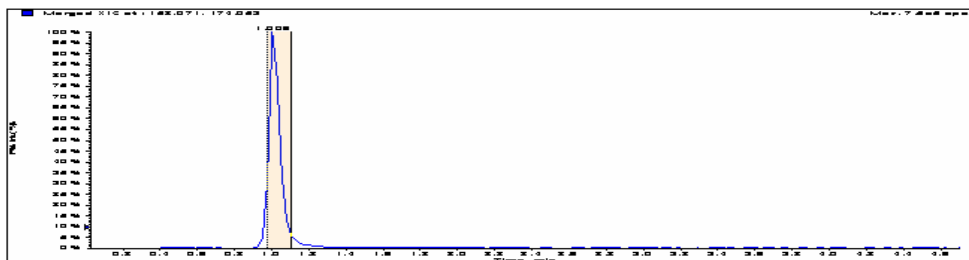
Mass Value = 271.03183

#	Formula	Compound	Adduct	Mass	Error (mDa)	* Error (ppm)	Ret. Time Error	Description
1	C9H10N4O2S2	Sulfamethizole	[M+H] <sup>+</sup>	271.03178	0.05	0.2	0.00	Sulfa mix #4

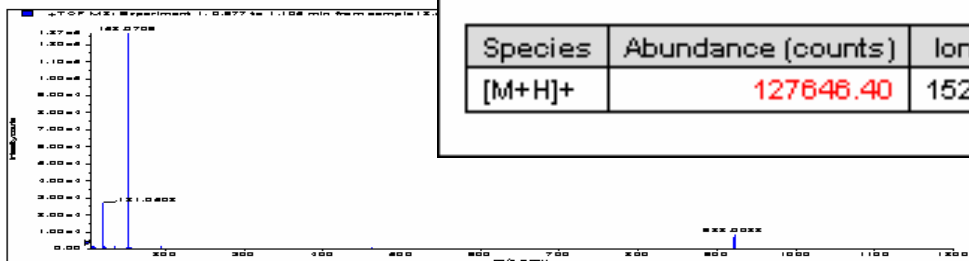


# Reverse Search – Formula Confirmation

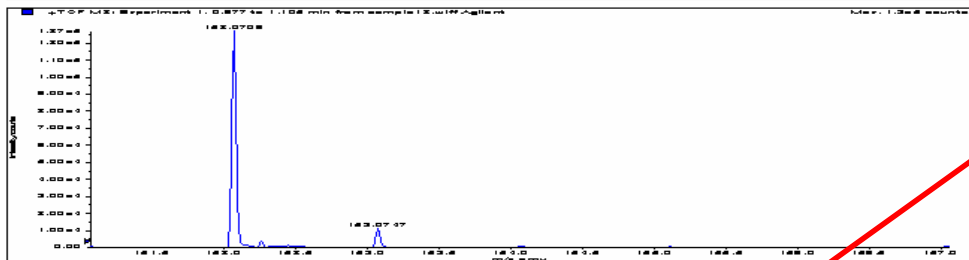
Sample Name: 4-Acetamidophenol Sample Location: P1-C4 Sample Id: Operator:  
 Data File Name: d:\pe\_sciex\_data\Projects\warp\_2\sample18.wiff Acq Time: July 31 2005\_04:38:23 PM  
 Method: D:\TOF\_Data\dmethods\MM\_etc.am\efc.xml



Merged XIC, Period#: 1 Experiment#: 1

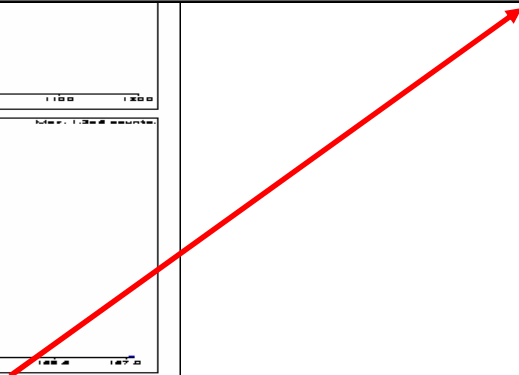


Species	Abundance (counts)	Ion Mass	Measured Mass	Error (mDa)	Error (ppm)
[M+H] <sup>+</sup>	127646.40	152.07060	152.07087	0.26800	1.76



Formula	Compound name	Mass	Peak RT (min)	Peak area	Description
C8H9NO2	--	151.06333	1.01	3.23530 E6	--

Species	Abundance (counts)	Ion Mass	Measured Mass	Error (mDa)	Error (ppm)	Ret. Time Error (min)
[M+H] <sup>+</sup>	127646.40	152.07060	152.07087	0.26800	1.76	--



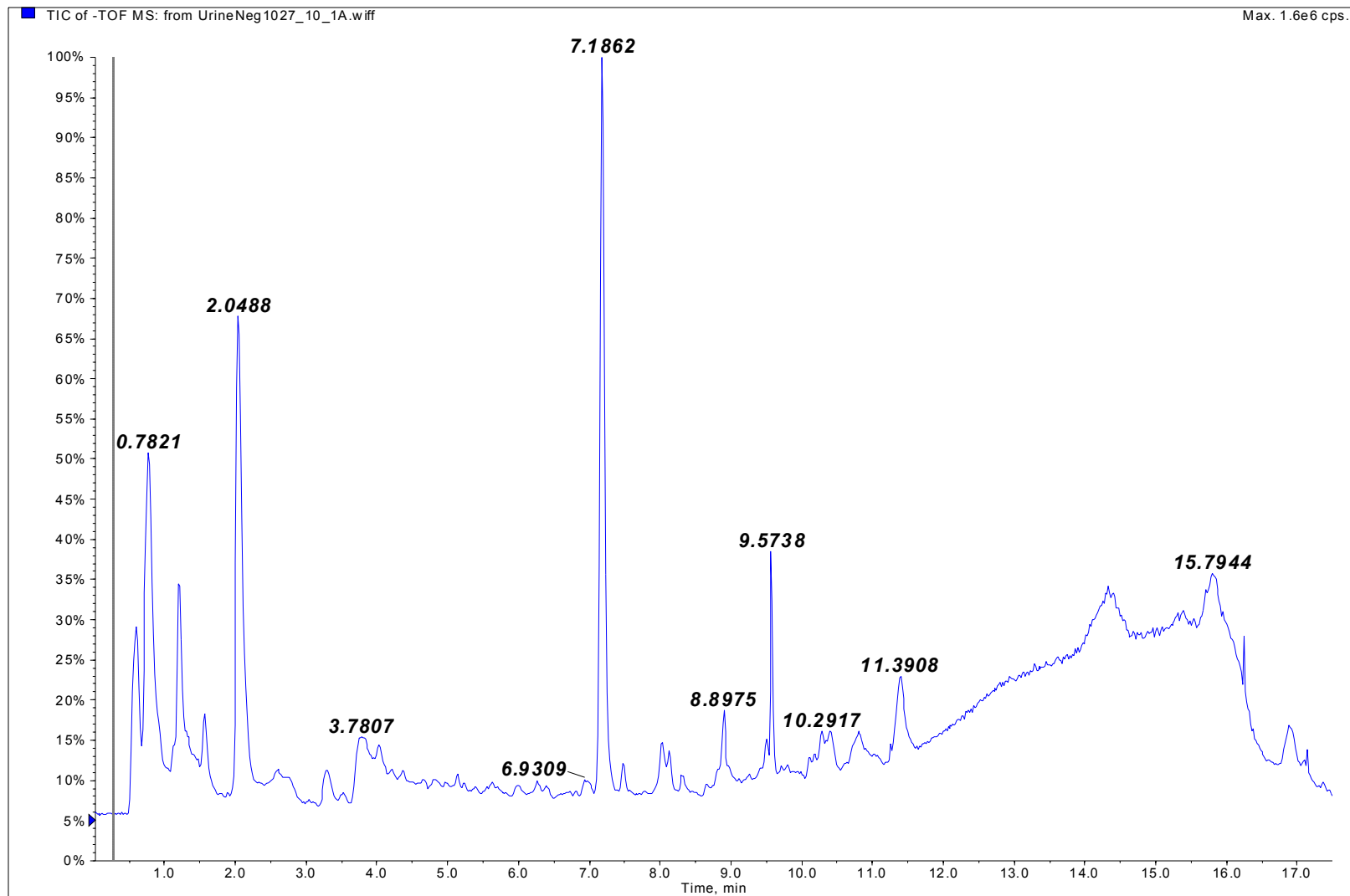
# Reverse Search – Database Search

- Same approach as formula confirmation but use a database containing formulas and optionally retention times
- For each compound in database, extract appropriate ion chromatograms from data file
- Filter results by mass error, abundance and retention time difference

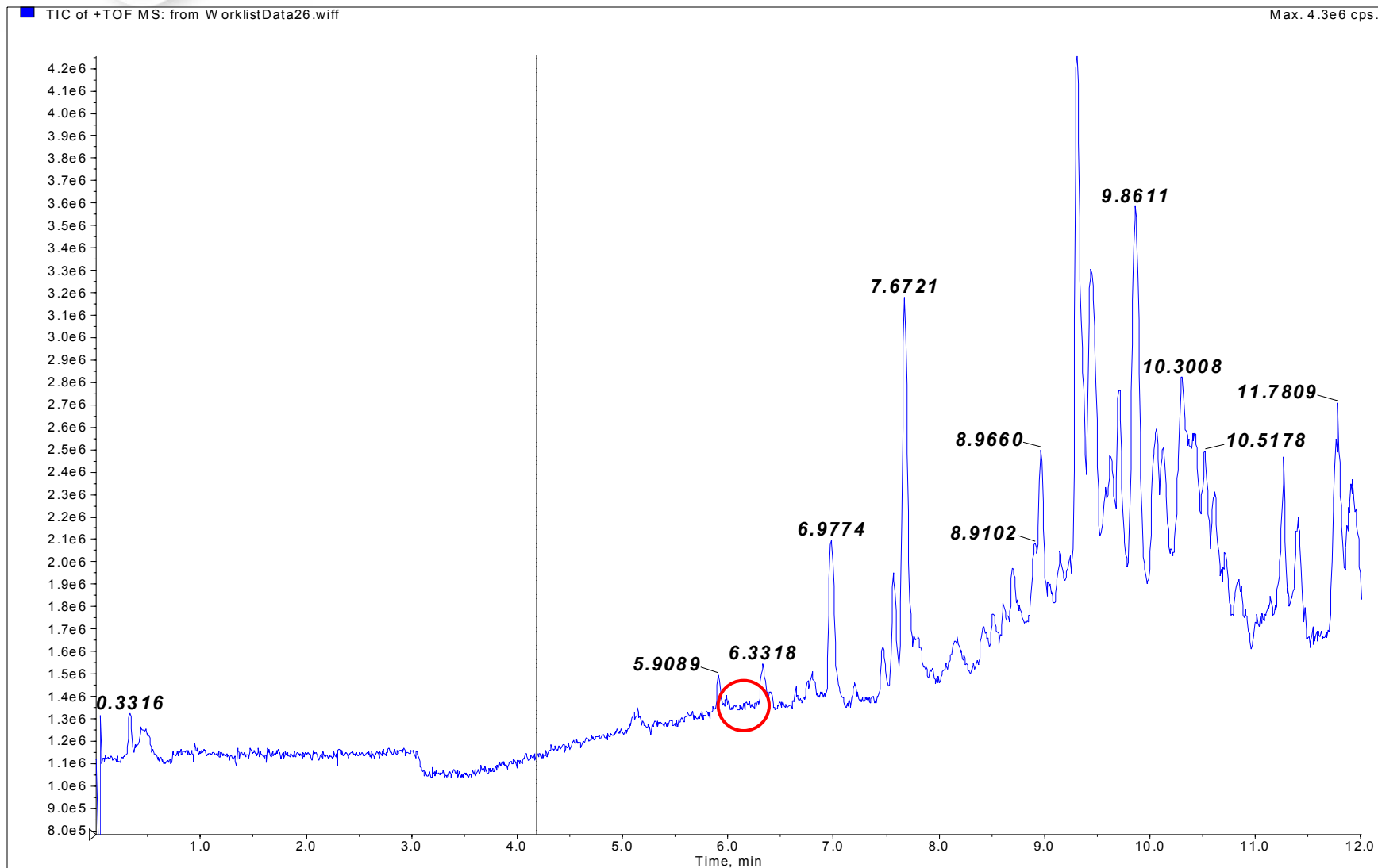
# Limitations

- Forward Search relies on compound being present as a distinct peak in a chromatographic signal
- Reverse search times become significant with large databases
- (~25 minutes for 100 compound database)
- In complex samples, usually only care about a few components
  - How to find the critical few?

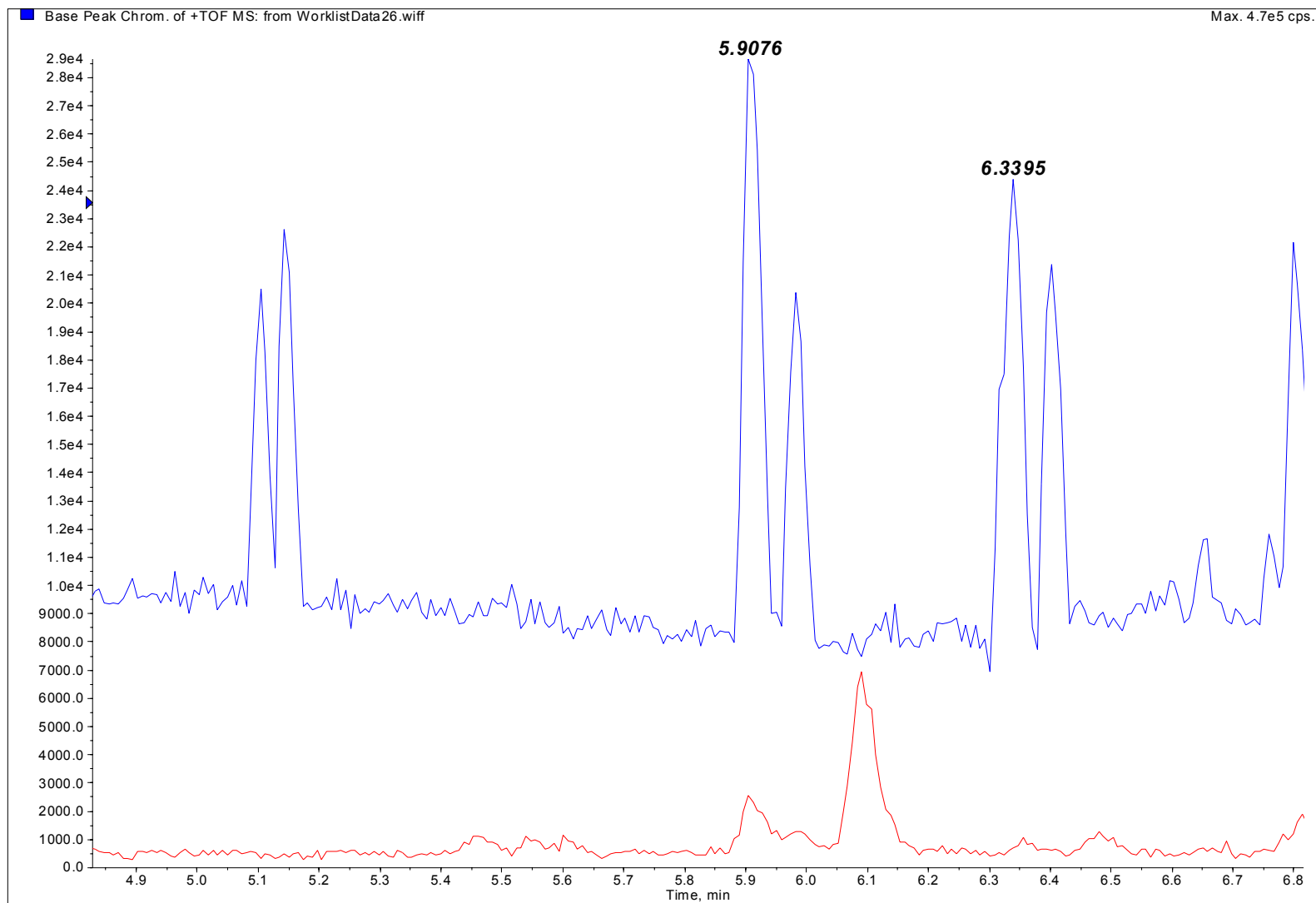
# Metabolites in Rat Urine



# Pesticides in Plant Extract



# Example - Diuron



# Finding the Critical Information

- A TOF spectrum contains 40 – 100,000 data points
- At two spectra/second, a 20 minute analysis has 2400 spectra
- Dealing with 100 million mass abundance pairs
- Most have no value to the chemist
- Solution:
  - Reduce data to ion sets indicative of compounds (molecular features)

# Mass Hunter Algorithm

- Detect spectral peaks for each spectrum
- Pool peaks, divide into  $m/z$  bins and strip out baseline data
- Remove small peaks to optimize computational performance
- Cluster peaks into time- $m/z$  2D space to form 3D peaks. Each peak represents an ion
- Group ions at same time into isotope clusters
- For each isotope cluster, determine charge carrier ( $H^+$ ,  $Na^+$ , etc), multimer number (monomer, dimer, etc) and charge state
- Group all isotope clusters at same time attributable to one neutral mass to form a molecular feature
- All features eluting at same time are a group

# Identifying Features

C:\PE Sciex Data\Projects\Pesticides Italy\Data\WorklistData26.wiff

File Batch Settings Process Help

Export features: 1100 / groups:566

Group	RT	mass	vol	m/z	defect	#ion	minZ	maxZ	#z	width	#sl
901	90	6.762	420.1554	2.67	421.1627	0.16	3	1	1	1	0.053
902	90	6.762		0.37	136.0412		1	0	0	1	0.049
903	90	6.761	317.1384	4.22	318.1457	0.14	3	1	1	1	0.047
904	90	6.761	300.1125	4.19	301.1192	0.11	8	1	1	1	0.039
905	90	6.760		0.84	168.0663		1	0	0	1	0.055
906	317	6.758	312.2315	0.71	313.2343	0.23	3	1	1	1	0.081
907	296	6.729	238.1572	1.36	239.1645	0.16	5	1	1	1	0.044
908	296	6.727		0.34	221.1537		1	0	0	1	0.047
909	106	6.700	264.1741	4.62		0.17	6	1	2	2	0.044
910	87	6.852	358.0699	4.51	359.0772	0.07	6	1	1	1	0.042
911	424	6.644	168.1152	0.42	169.1224	0.12	2	1	1	1	0.053
912	266	6.632	308.1999	1.32	309.2067	0.20	4	1	1	1	0.054
913	289	6.629	222.0913	0.76	223.0988	0.09	3	1	1	1	0.058
914	289	6.627		0.28	149.0239		1	0	0	1	0.066
915	289	6.627	568.2860	0.43	569.2932	0.29	2	1	1	1	0.082
916	289	6.627		0.32	177.0565		1	0	0	1	0.061
917	415	6.582	296.0691	0.46	297.0764	0.07	2	1	1	1	0.047
918	444	6.543	328.0618	0.41	329.0690	0.06	2	1	1	1	0.051
919	170	6.538		0.59	294.2286		1	0	0	1	0.062
920	170	6.537	276.1954	2.29	277.2021	0.20	4	1	1	1	0.062
921	506	6.532		0.27	171.1382		1	0	0	1	0.060
922	467	6.506	206.1309	0.40	207.1383	0.13	3	1	1	1	0.056
923	277	6.497	282.1853	0.74	283.1926	0.19	3	1	1	1	0.067
924	277	6.496	224.1415	1.11	225.1481	0.14	3	1	1	1	0.054
925	375	6.486	381.2506	0.52	382.2578	0.25	2	1	1	1	0.108

Export Group #90

charge state	RT	m/z	m0	vol	width
1 M	6.761		317.1384	4.22	0.047
2 M+H	6.763	318.1457	317.1384	3.53	0.052
3 M+H+1	6.762	319.1499		0.63	0.051
4 M+H+2	6.759	320.1531		0.06	0.039
5					
6 M	6.761		300.1125	4.19	0.039
7 M+H	6.762	301.1192	300.1119	1.72	0.049
8 M+H+1	6.762	302.1225		0.30	0.052
9					
10 M+Na	6.762	323.1016	300.1131	1.65	0.048
11 M+Na+1	6.764	324.1052		0.29	0.050
12 M+Na+2	6.752	325.1059		0.02	0.015
13					
14 M+K	6.765	339.0768	300.1136	0.06	0.037
15					
16 2M+Na	6.759	623.2137	300.1127	0.11	0.032
17 2M+Na+1	6.761	624.2161		0.05	0.031
18					
19 M	6.762		420.1554	2.67	0.053
20 M+H	6.762	421.1627	420.1554	2.09	0.056
21 M+H+1	6.764	422.1664		0.47	0.049
22 M+H+2	6.761	423.1744		0.11	0.054
23					

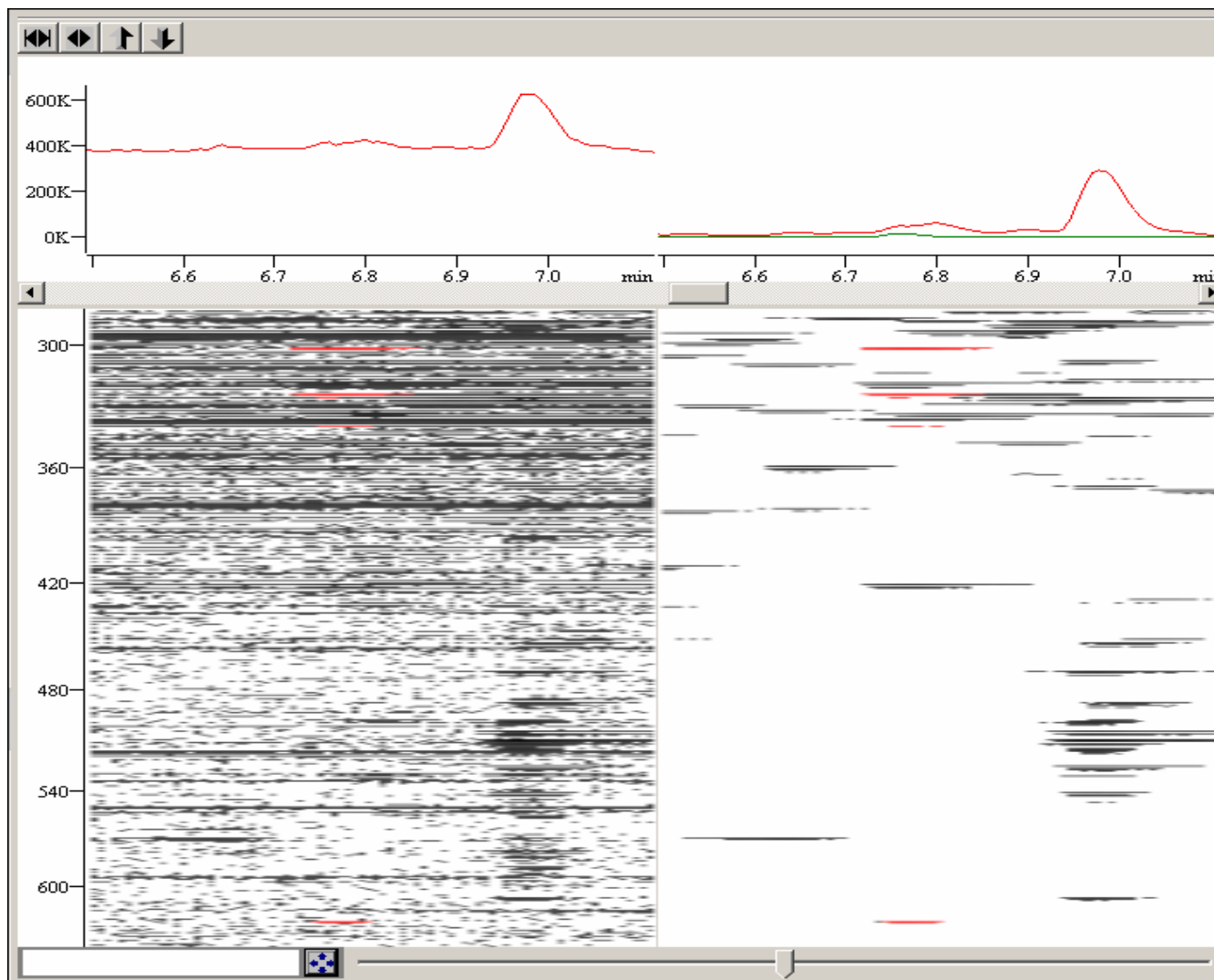
# Results of Data Reduction

	Group	RT	mass	vol	m/z	defect	#ion	minZ	maxZ	#z	width	#sl ▲
901	90	6.762	420.1554	2.67	421.1627	0.16	3	1	1	1	0.053	
902	90	6.762		0.37	136.0412		1	0	0	1	0.049	
903	90	6.761	317.1384	4.22	318.1457	0.14	3	1	1	1	0.047	
904	90	6.761	300.1125	4.19	301.1192	0.11	8	1	1	1	0.039	
905	90	6.760		0.84	168.0663		1	0	0	1	0.055	
906	317	6.758	312.2315	0.71	313.2343	0.23	3	1	1	1	0.081	
907	296	6.729	238.1572	1.36	239.1645	0.16	5	1	1	1	0.044	
908	296	6.727		0.34	221.1537		1	0	0	1	0.047	
909	106	6.700	264.1741	4.62		0.17	6	1	2	2	0.044	
910	87	6.652	358.0699	4.51	359.0772	0.07	6	1	1	1	0.042	
911	424	6.644	168.1152	0.42	169.1224	0.12	2	1	1	1	0.053	
912	266	6.632	308.1999	1.32	309.2067	0.20	4	1	1	1	0.054	
913	289	6.629	222.0913	0.76	223.0988	0.09	3	1	1	1	0.058	
914	289	6.627		0.28	149.0239		1	0	0	1	0.066	
915	289	6.627	568.2860	0.43	569.2932	0.29	2	1	1	1	0.082	
916	289	6.627		0.32	177.0565		1	0	0	1	0.061	
917	415	6.582	296.0691	0.46	297.0764	0.07	2	1	1	1	0.047	
918	444	6.543	328.0618	0.41	329.0690	0.06	2	1	1	1	0.051	
919	170	6.538		0.59	294.2286		1	0	0	1	0.062	
920	170	6.537	276.1954	2.29	277.2021	0.20	4	1	1	1	0.062	
921	506	6.532		0.27	171.1382		1	0	0	1	0.060	
922	467	6.506	206.1309	0.40	207.1383	0.13	3	1	1	1	0.056	
923	277	6.497	282.1853	0.74	283.1926	0.19	3	1	1	1	0.067	
924	277	6.496	224.1415	1.11	225.1481	0.14	3	1	1	1	0.054	
925	375	6.486	381.2506	0.52	382.2578	0.25	2	1	1	1	0.108	

# Results for 6.76 minutes

	charge state	RT	m/z	m0	vol	width
1	M	6.761		317.1384	4.22	0.047
2	M+H	6.763	318.1457	317.1384	3.53	0.052
3	M+H+1	6.762	319.1499		0.63	0.051
4	M+H+2	6.759	320.1531		0.06	0.039
5						
6	M	6.761		300.1125	4.19	0.039
7	M+H	6.762	301.1192	300.1119	1.72	0.049
8	M+H+1	6.762	302.1225		0.30	0.052
9						
10	M+Na	6.762	323.1016	300.1131	1.65	0.048
11	M+Na+1	6.764	324.1052		0.29	0.050
12	M+Na+2	6.752	325.1059		0.02	0.015
13						
14	M+K	6.765	339.0768	300.1136	0.06	0.037
15						
16	2M+Na	6.759	623.2137	300.1127	0.11	0.032
17	2M+Na+1	6.761	624.2161		0.05	0.031
18						
19	M	6.762		420.1554	2.67	0.053
20	M+H	6.762	421.1627	420.1554	2.09	0.056
21	M+H+1	6.764	422.1664		0.47	0.049
22	M+H+2	6.761	423.1744		0.11	0.054
23						

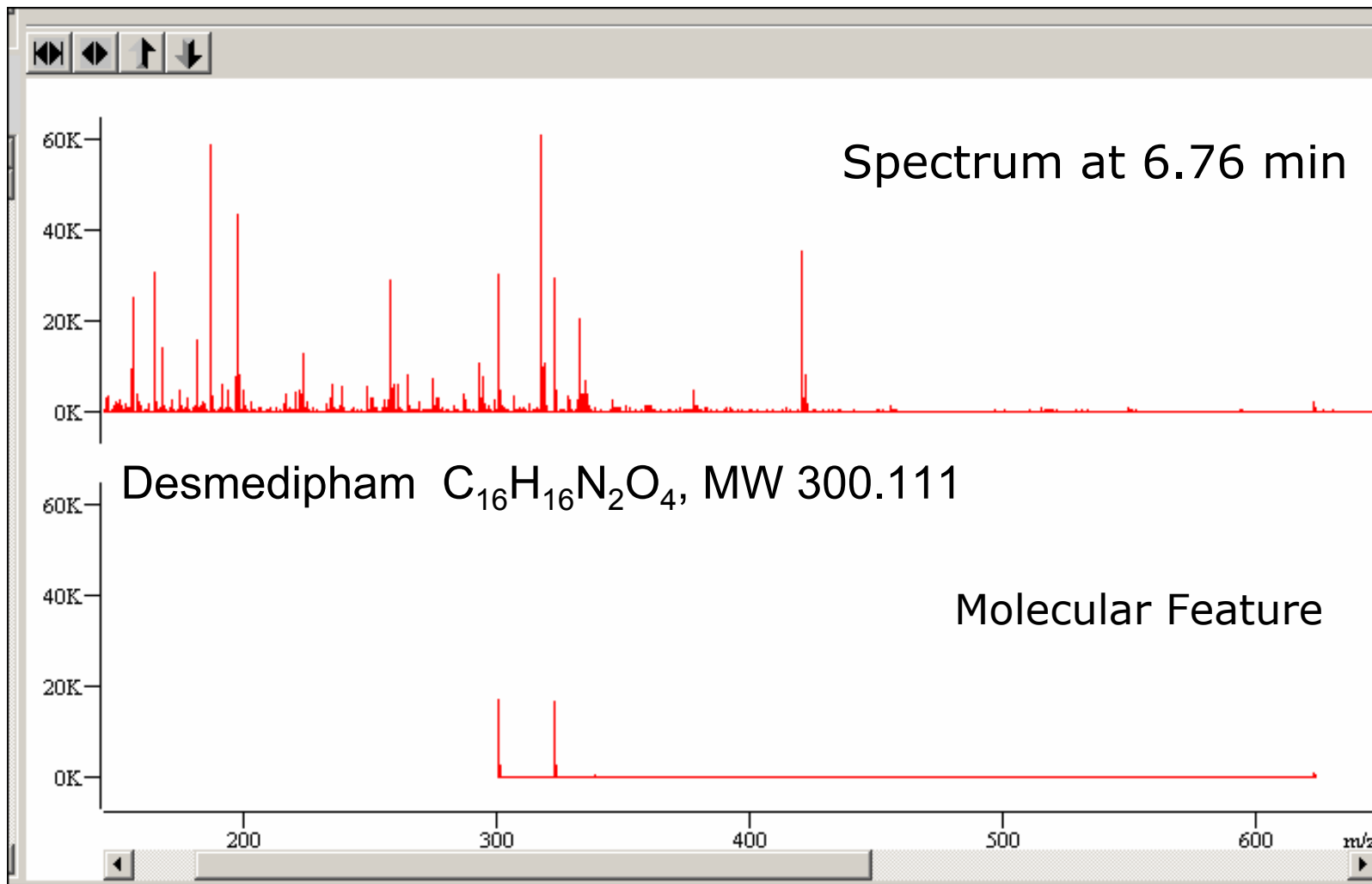
# 2D Imaging



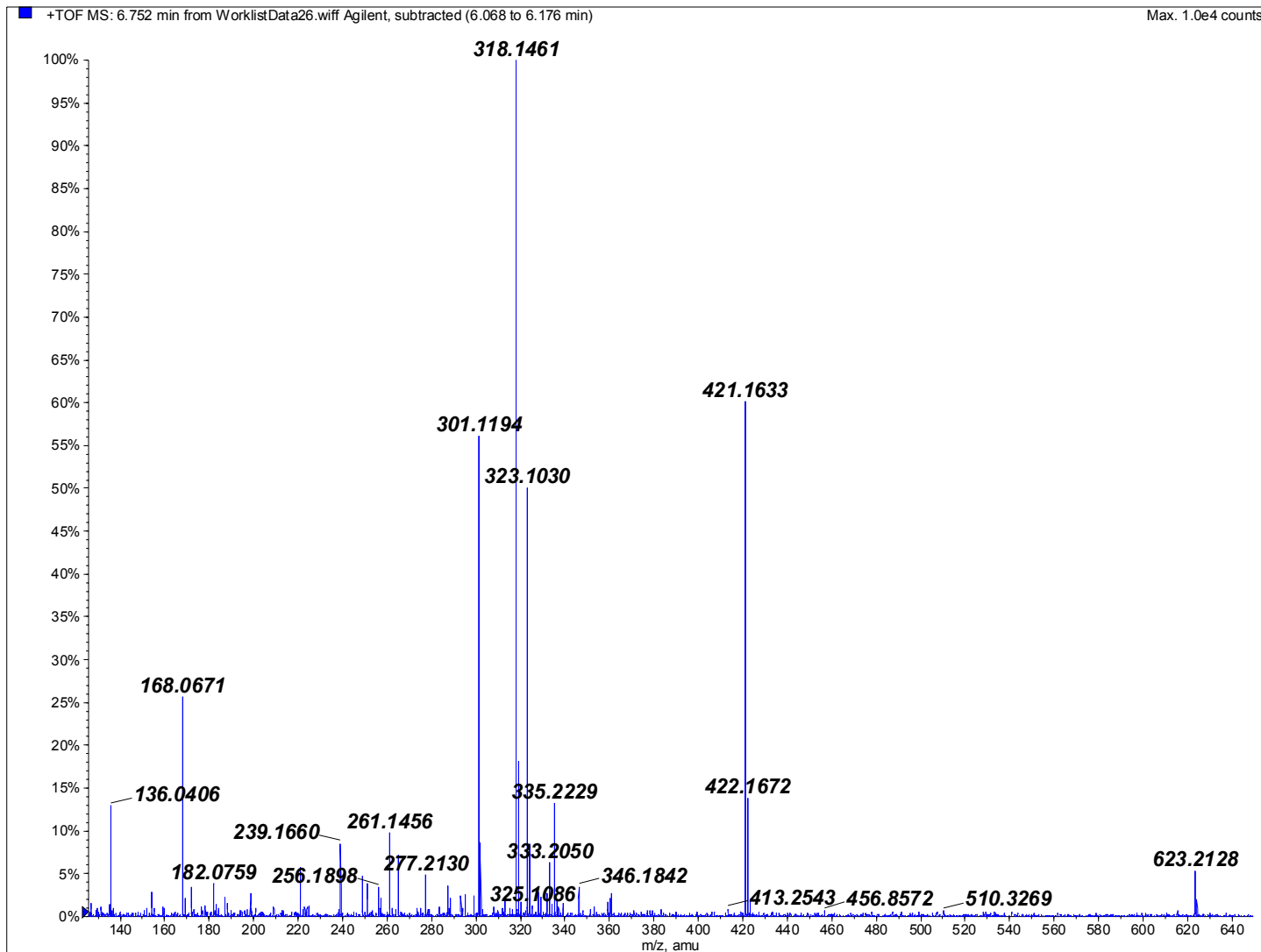
Original  
Data

Reduced  
Data

# Spectral Data



# Original Spectrum



# Narrowing the Results

**Extraction Parameters**

Data Ranges

	Min	Max	
RT	0.00	20000.00	min
M/Z	0	1000000	Da

Spectral Peak Detection

S/N threshold: 5

Feature Detection

Max spectral peaks to use: 500 × 1000

Peptidic (averagine)  
 Single charge only

Adducts

**Positive ions**

salt dominated

[ ] Add

K  
 Na

[ ] Delete

**Negative ions**

[ ] Add

[ ] Delete

OK Cancel

**Display Filters**

Feature Position

Use all the available data

	Min	Max	
RT	0	2000	min.
Mass	0	100000	Da

Special Masses

Exclude  Limit to these Accuracy: 0.0050 Da

Mass Defect (amu)

Min	Max
-0.50	0.50

Charge State

Any  Mult. charge required  Mult. charge forbidden

Isotope Pattern

Formula: Custom

formula: C12BrN

normalized height error: 0.1

Doubtful Feature Exclusion

features w/o isotopes  
 features with unknown mass

Abundance

Relative Absolute Top

Largest: 1000000 Compounds

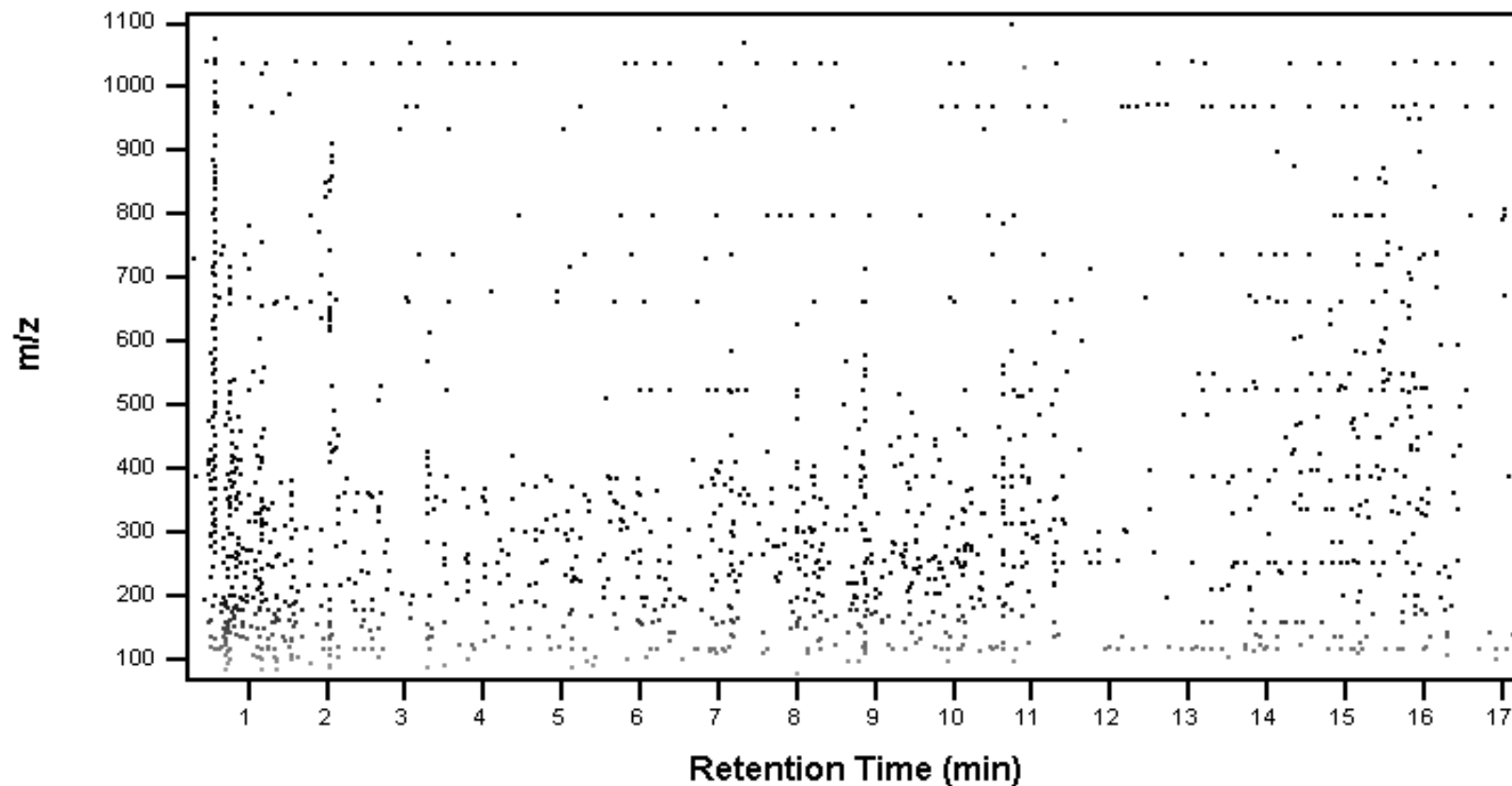
OK Cancel

# Using the Information

- Use the calculated molecular weight of a feature to:
  - Propose empirical formulas
  - Match against target compound database
    - Metabolites
    - Forensics
- Compare the features found to other samples
  - Look for similarities
  - Look for differences

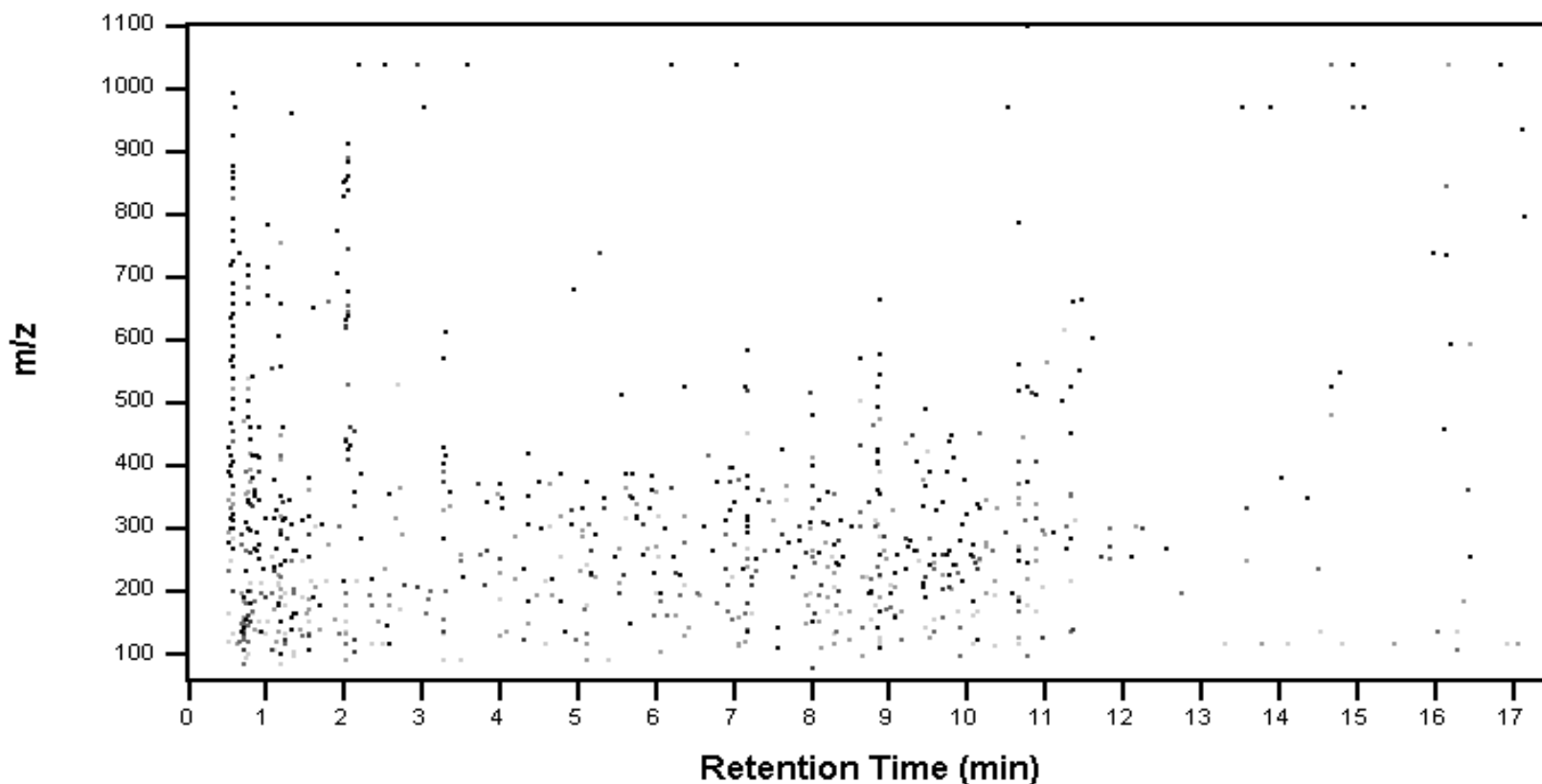
# Molecular Features

Approximately 1500 Molecular Features  
Identified in a single Sample Analysis



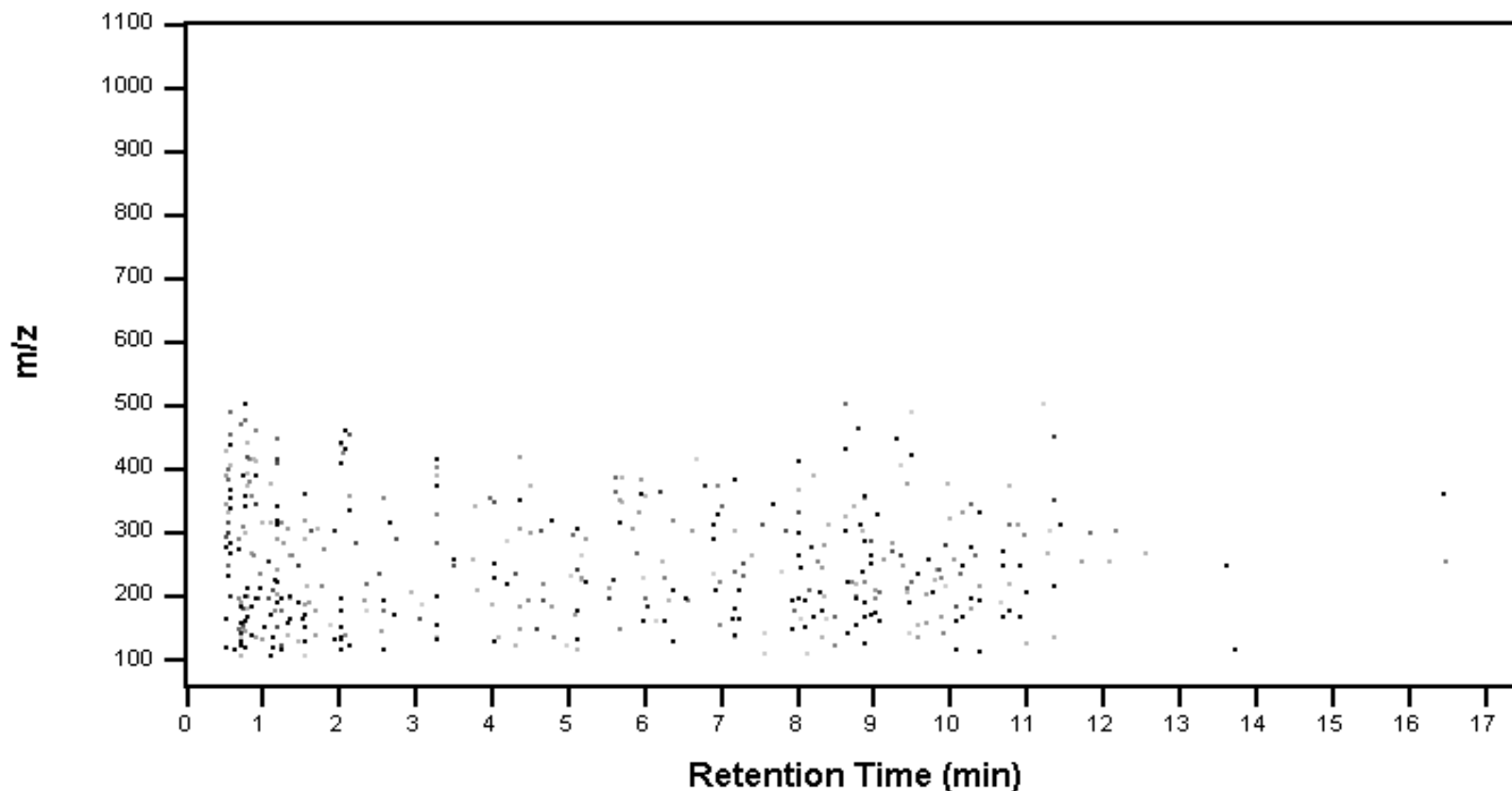
# Molecule Features Common to All Sample A Analyses

836 Molecular Features were common to all Sample A analyses

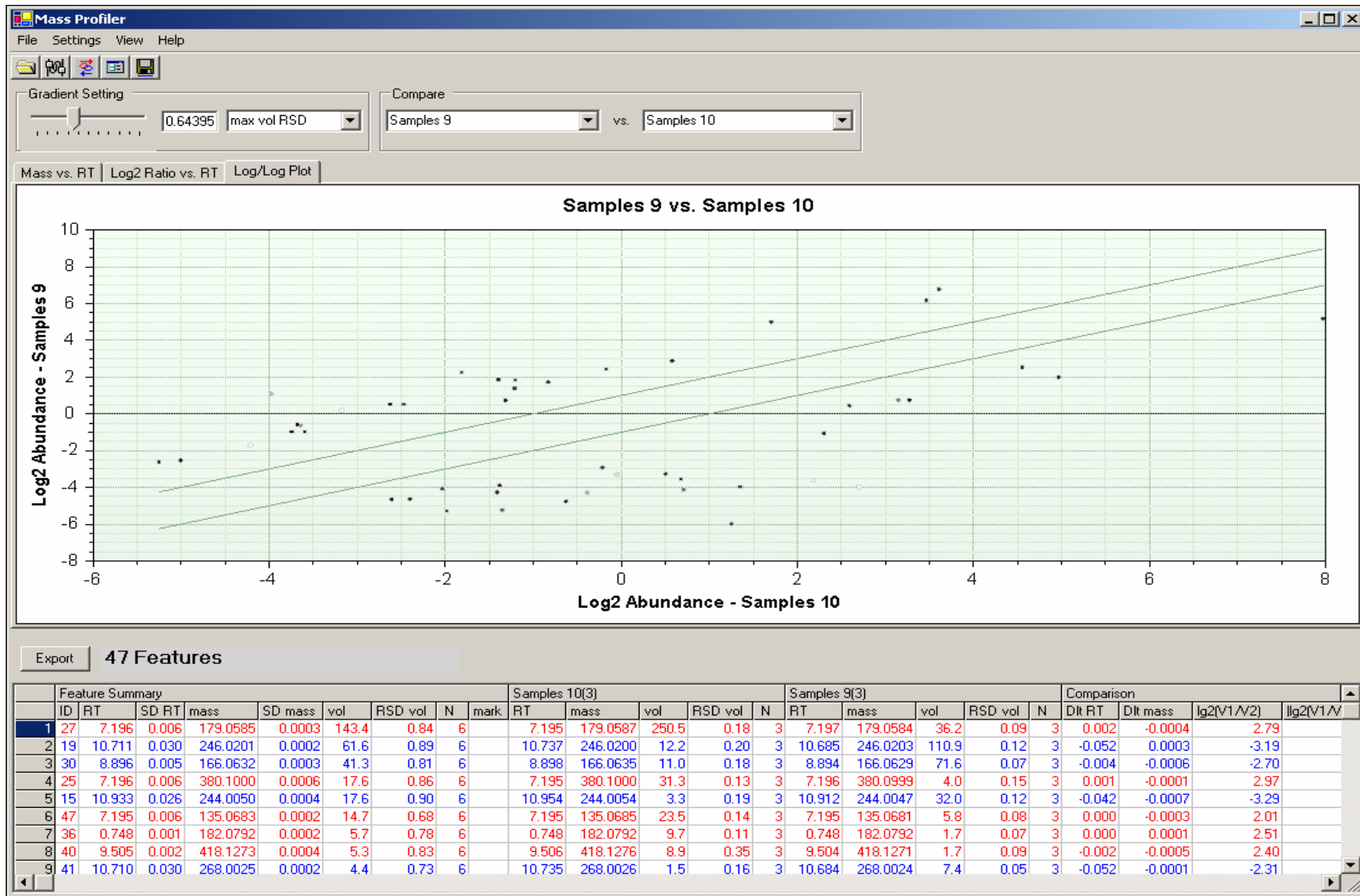


# Molecular Features Common to all A & B Samples

441 Molecular Features were found to be present in all A and B Samples



# Comparison Details



# Identifying Biomarker

**Information for Feature # 30**

Export Chemistry Calculate Compositions

	chemical form	dm(amu)	dm(ppm)	DBE	score
1	C9H10O3	0.0000	0.3	5.0	87
2	C2H10N6OS	0.0007	4.4	1.0	79
3	C5H6N6O	-0.0026	-15.9	6.0	43
4	C6H14O3S	0.0034	20.6	0.0	78

Export All Samples

	ID	Name	RT	mass	vol
1	1	UrineNeg1027_10_1	8.893	166.0635	8.82
2	2	UrineNeg1027_10_2	8.899	166.0635	12.69
3	3	UrineNeg1027_10_3	8.902	166.0635	11.61
4	4	UrineNeg1027_9_1A	8.898	166.0629	74.15
5	5	UrineNeg1027_9_2A	8.889	166.0629	65.55
6	6	UrineNeg1027_9_3A	8.894	166.0630	75.07

Export UrineNeg1027\_10\_3A\_0\_0\_0

	charge state	RT	m/z	m0	vol	width
1	M	8.902		166.0635	11.61	0.069
2	M-H	8.901	165.0562	166.0634	10.58	0.084
3	M-H+1	8.900	166.0594	166.0637	0.87	0.088
4	M-H+2	8.901	167.0600	166.0621	0.08	0.058
5						
6	2M-H	8.908	331.1195	166.0634	0.08	0.047
7						

Samples 10

**UrineNeg1027\_10\_1A\_0\_0\_0**

**UrineNeg1027\_10\_2A\_0\_0\_0**

**UrineNeg1027\_10\_3A\_0\_0\_0**

Samples 9

**UrineNeg1027\_9\_1A\_0\_0\_0**

**UrineNeg1027\_9\_2A\_0\_0\_0**

**UrineNeg1027\_9\_3A\_0\_0\_0**

# Applications

- Create database of commonly found “features”
- Focus on similarities
- Focus on differences
- Create time/precursor ion inclusion lists for MS/MS experiments
- Use to calculate MW of multiply charged molecules when charge cluster isotopes are resolved

# Summary

- It is possible to reduce the data in very large files containing accurate mass information and produce a list of “features” indicative of compounds even when not observable as peaks in a chromatogram.
- By statistically comparing batches of these reduced data sets, information on the key compounds in complex samples can be extracted and used for further experiments