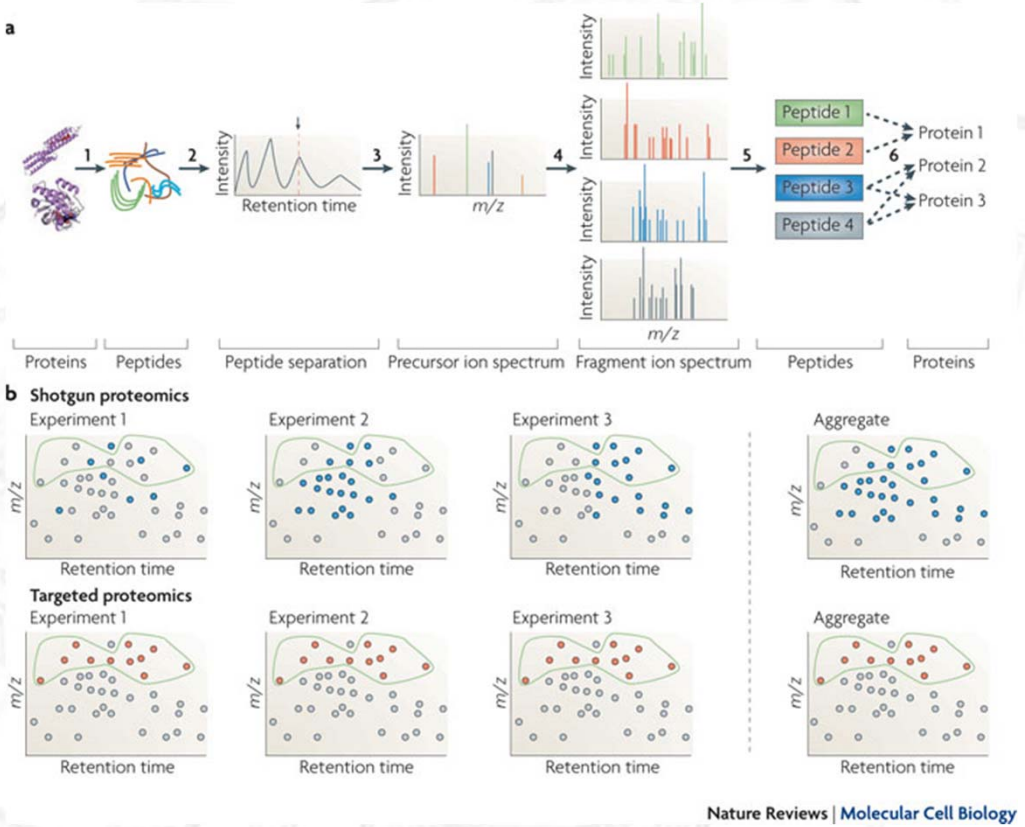
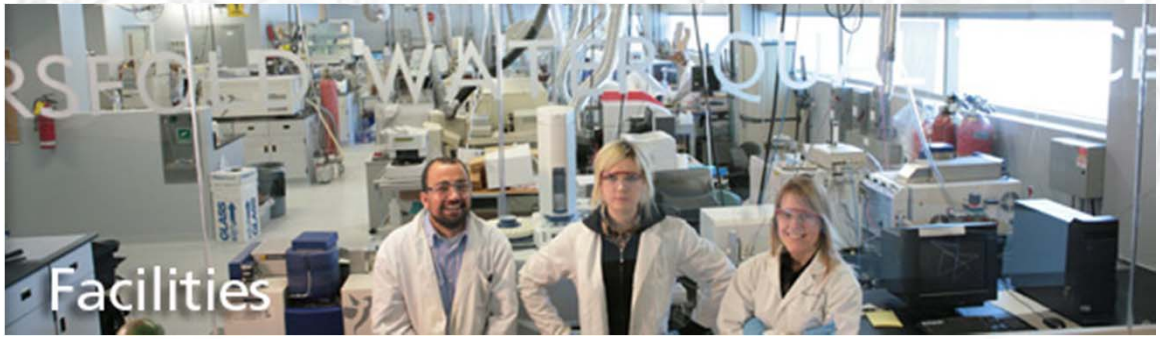
Week 9: MS in Space and Proteomics



Last Time…

- Detectors

- Small Molecule Applications, Environmental: (e.g. TWQC)



Mass Spectrometry in Space

- Possibly the coolest application of small molecule MS is in space…

- Enter Case Study #2: Mass Spectrometry and NASA

- What are the major considerations for MS in space?

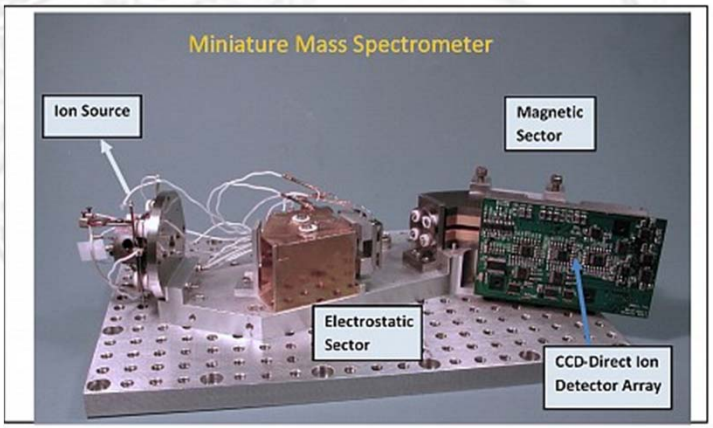
  1. What do we need to be able to do? (Mass limit, accuracy, resolution, sensitivity)

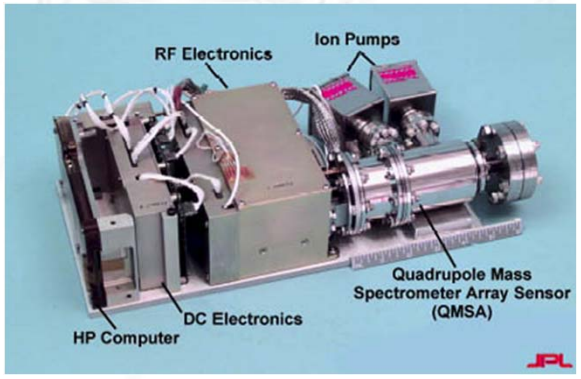
  2. Size and weight of instrument

  3. Power requirements.

Mass Spectrometry and NASA

- Over the years, NASA has had quite a number of MS instruments on board it's spacecraft. Why?
  - Sampling Atmosphere (upper and lower)
  - Sampling Soil
  - Monitoring cabin atmosphere, life support
- Of course, use is space requires that the instrument be miniaturized, which also reduces weight and, generally power consumption





Challenges of Miniaturization

Property	Space	Lab
Weight	~kg	10-100kg
Power	~5-10W	100-1000 W
Size	<50cm	1-10 m
Robust	1g	10g @100Hz
MTBF	Years	hours
UV	100Mcts/s	1ct/s
Source Temp	1-100 keV	0.001 keV
Energy width	100%	.01-1%
Radiation	10-100kRad/y	<1 kRad/y

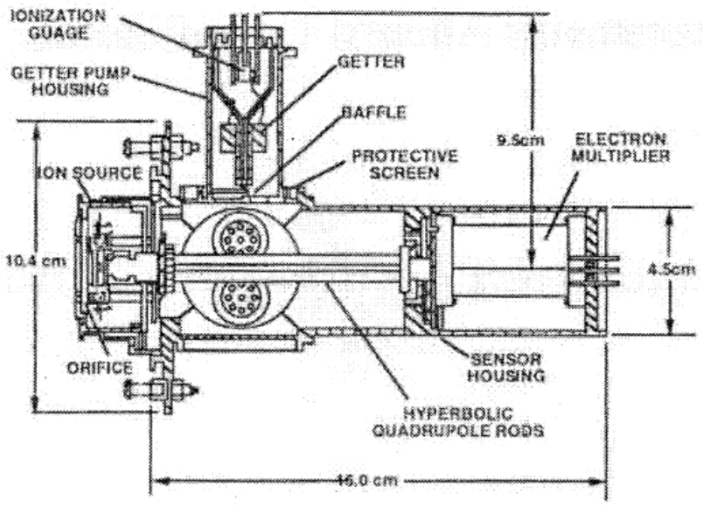
History of MS in Space

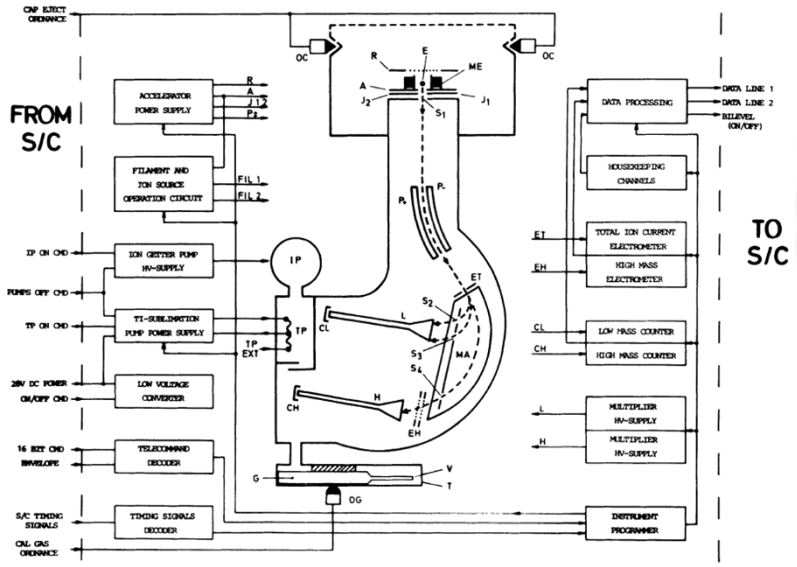
Mass Spectrometer	Year, Mission	Resolution
Ion Traps	1959 Luna 1	< 2
Faraday Cup	1961 Explorer 10	~2
Electrostatic E/Q	1962 Mariner 2	~3
GC Double Sector	1975 Viking 1 and 2	~50
Hyperbolic Quadrupole	1978 Pioneer (Venus U. atmosphere)	~30
Magnetic Sector	1978 Pioneer (Venus L. atmosphere)	~30
Wien Filter	1983 ISEE-3	~5
Magnetic Sector	71 Apollo*, 86 Giotto*	>40, >10
Linear TOF	1984 Ampte	~15
Isochronous TOF	1996 Wind	~100
Reflectron TOF	2004 Rosetta*	>3000

Properties of Some Mini MS Instruments

- Obviously, Mini-MS doesn't really work as well as the big gigantic ones we have in the lab. However:

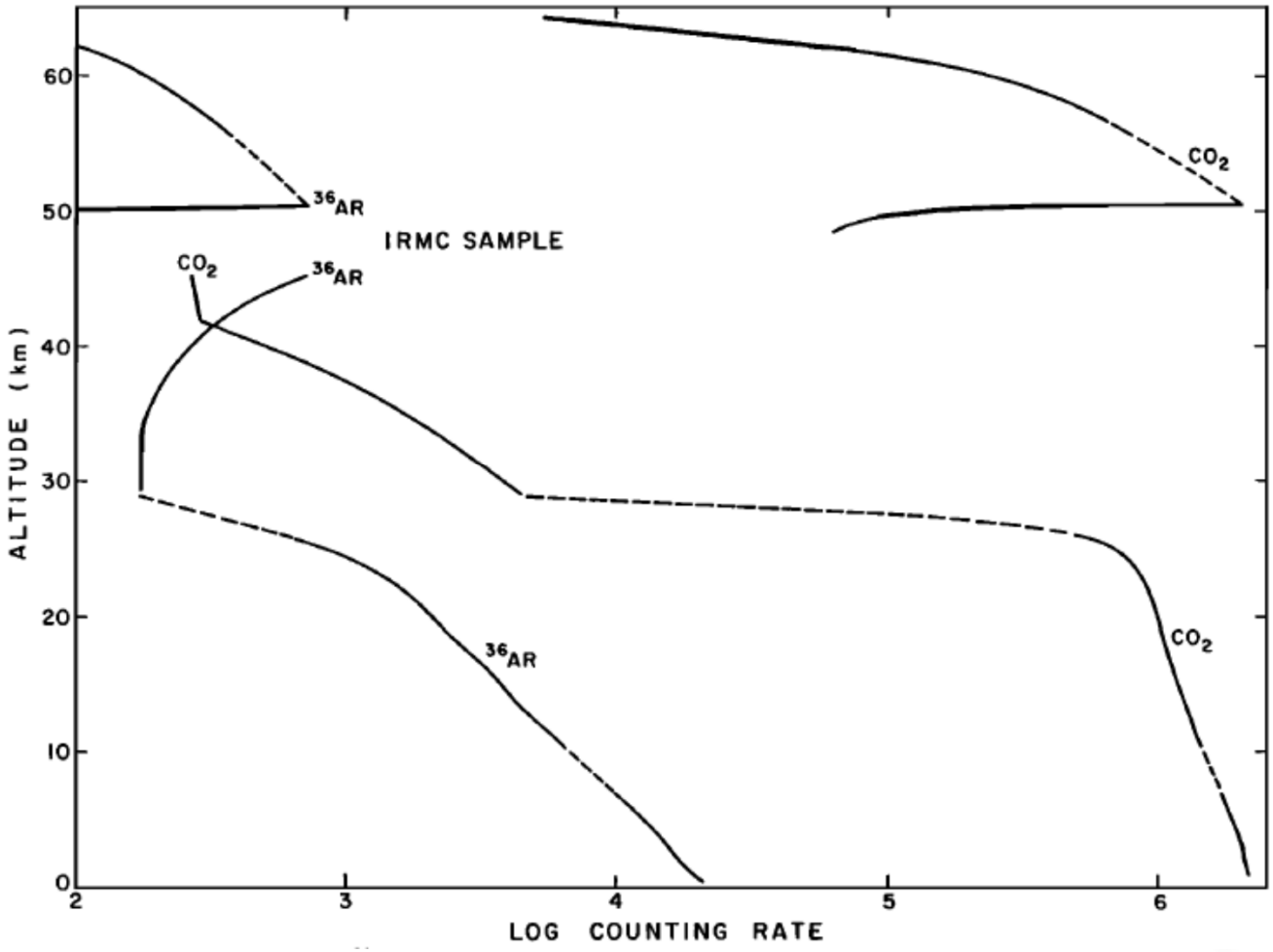
Instrument	Pioneer Venus—Upper Atmosphere	Pioneer Venus—Lower Atmosphere	Mars Viking Lander
Mass Analyzer	quadrupole	magnetic sector	dual sector (B/E)
Detection Limit	N/A	1 ppmv	ppbv-ppmv (soil)
Mass Range	1–46 Th	1–208 Th	12-250 Th
Size	N/A	N/A	0.6 ft^3
Weight	8.4 lbs	24 lbs	45 lbs
Power	12 W	14 W	140





The Venutian Atmosphere…

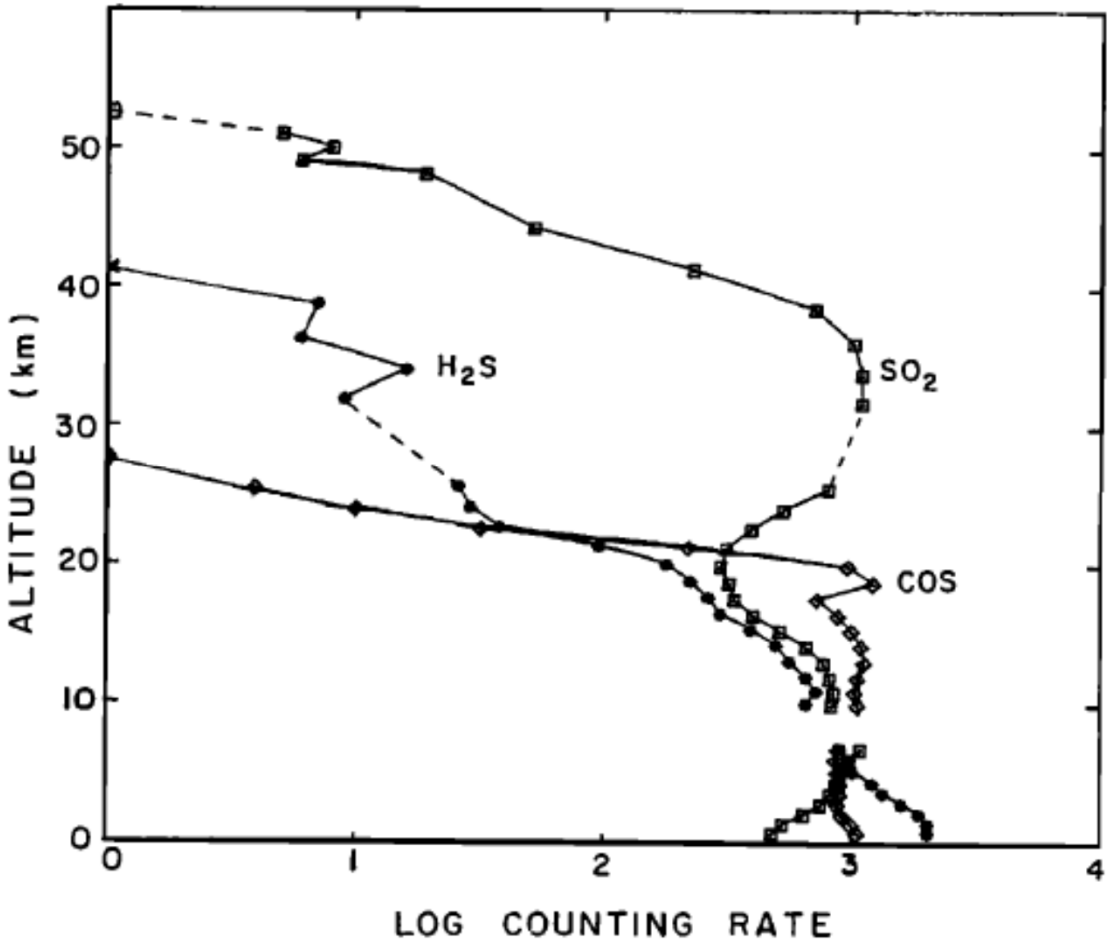
- As it was falling into the atmosphere, the Pioneer Lander made the following measurements:



- Dotted lines are when they were at low EI energies…

- Decreases between 50 and 28 km due to accumulation of H_2SO_4 droplets

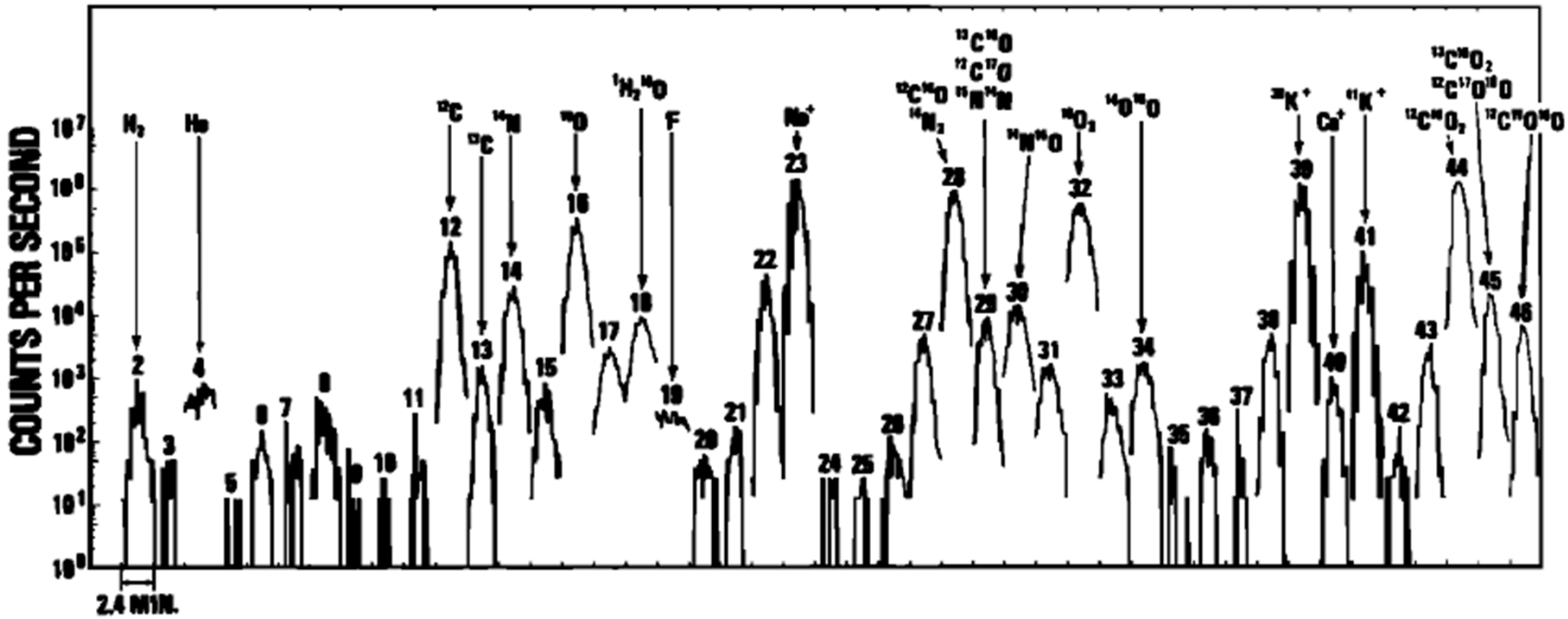
Sulphur Gasses in the Venusian Atmosphere



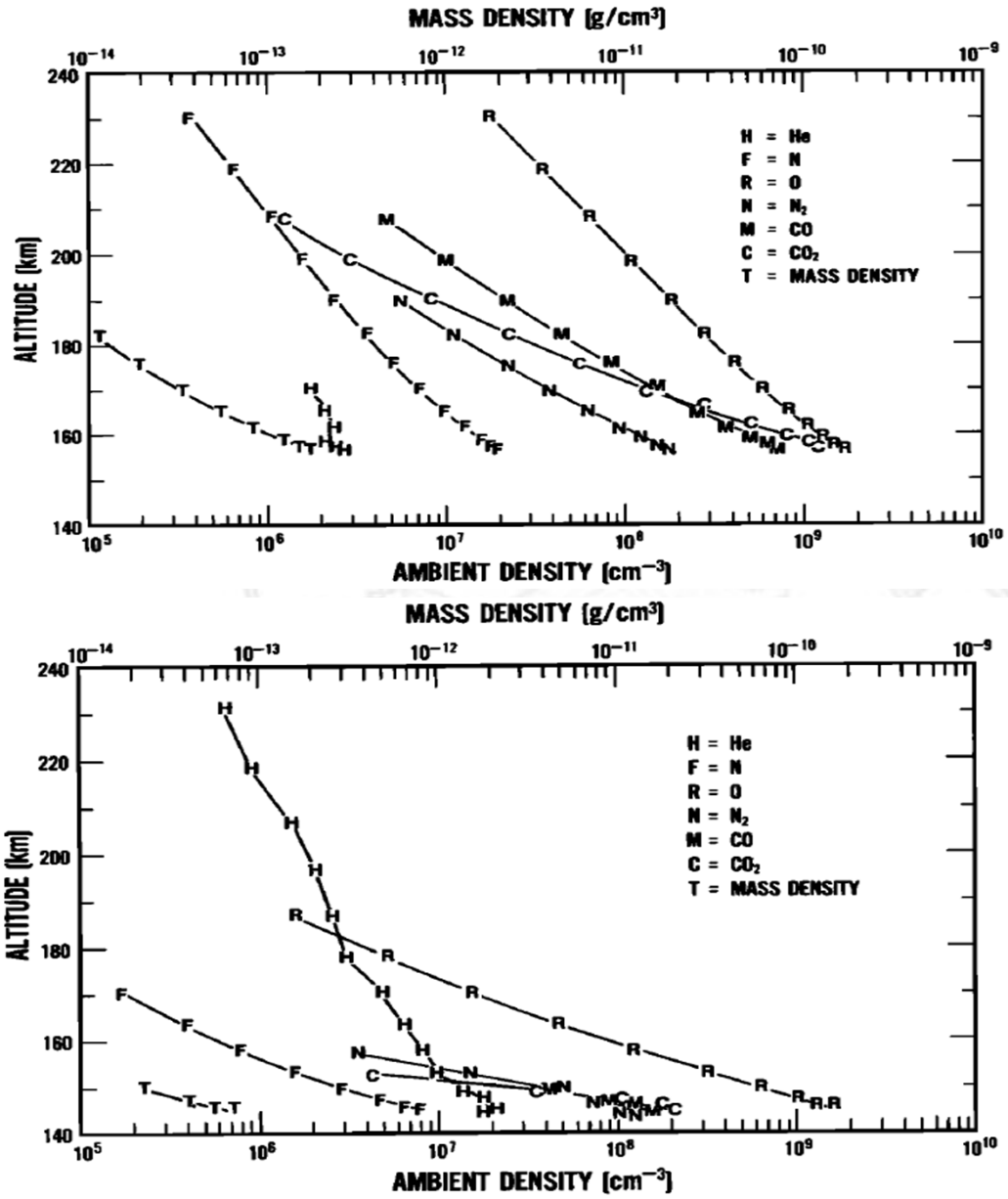
- Peak in SO_2 at 38 km is due to evaporation of H_2SO_4
- Lots of Carbonyl sulfide (COS) in the lower atmosphere

And the Upper Atmosphere…

- Pioneer also had an MS on it's orbiter… A Quadrupole MS, no less…



More of the Venusian Upper Atmosphere



Daytime

- Measurements are much cleaner when you're not falling through the atmosphere...

Nighttime

The Viking GC-MS

- The GC-MS on Viking lander was designed to look for, among other things, organic compounds at the ppb level in soil.

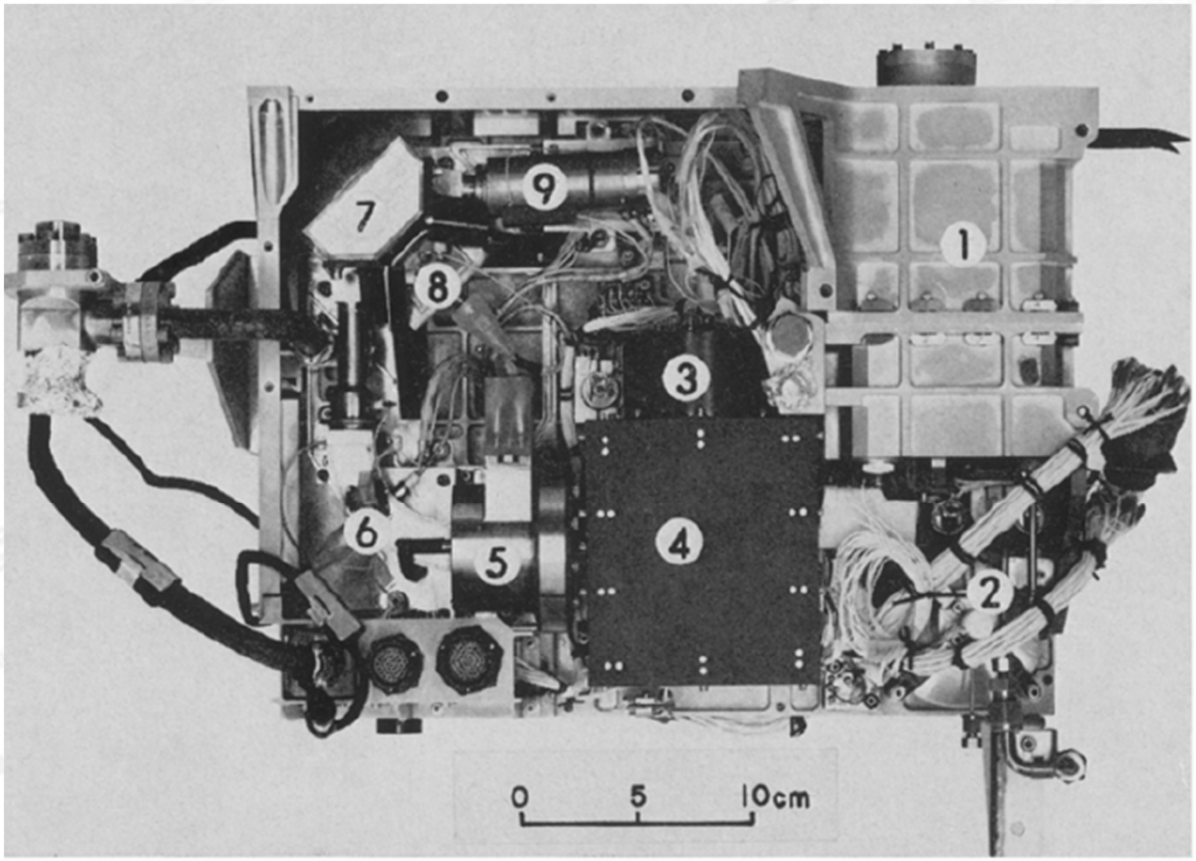
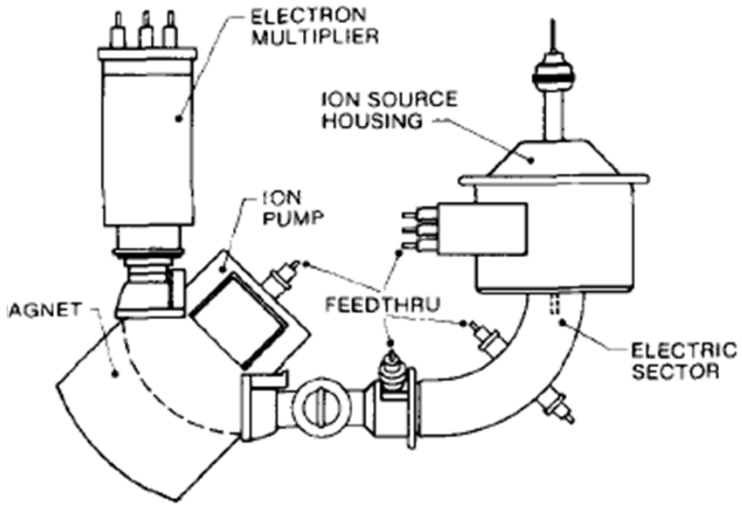
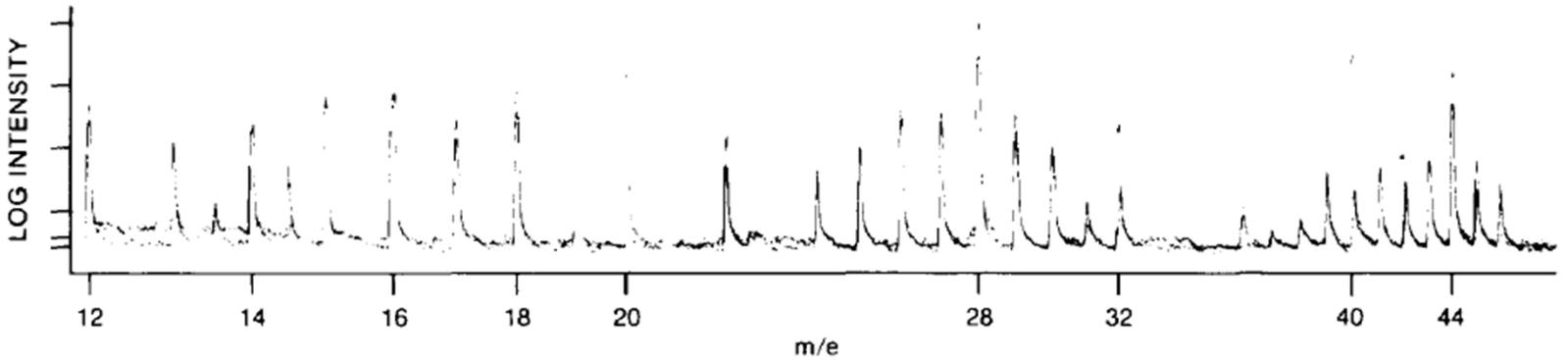


Fig. 2. Development Test Unit of the Viking GC-MS instrument (side view). (1) Sample oven housing. (2) Hydrogen tank. (3) GC-column. (4) Valving, effluent divider, separator (in housing held at 200°C). (5) Ion source housing. (6) Electric sector. (7) Magnet. (8) Ion pump. (9) Electron multiplier.

The Viking GC-MS Cont.





Oh No He didn't! Controversey…

- A huge controversy erupted over the Viking Lander GC-MS. It started with a publication by Rafael Navarro-Gonzalez *et al.* which was entitle thus:

> **The limitations on organic detection in Mars-like soils by thermal volatilization–gas chromatography–MS and their implications for the Viking results**

- The bottom line of the paper was that under very dry conditions (similar to those of martian soil), the Viking GC-MS would have missed organic molecules at the ppb level – an amount that is consistent with low concentrations of microorganisms such as those found in the deep antarctic.

Viking Lander MS Sucks?...

- Here are the results summarized:

Table 1. Total organic matter (TOM) present in different Mars analogs soils and its detection by TV–GC–MS

Soil sample	TOM, μg of C per gram of soil	$\delta^{13}C$	C/N ratio	TV–GC–MS,* 500°C, μg of benzene per gram of soil	TV–GC–MS,* 750°C, μg of benzene per gram of soil
Antarctic Desert					
Dry Valley	20–30	−25.03	0.9	N.D.	N.D.
Dry Valley (sample no. 726)	60–90	−24.34	0.3	N.D.	N.D.
Otway massif mill stream glacier	10–20	−25.13	1.0	N.D.	N.D.
Atacama Desert					
Yungay, Chile (AT02-03A)	20–40	−26.09	8.2	N.D.	1–4
La Joya, Peru (PC03-06)	20–30	−21.04	0.3	N.D.	1–4
Las Juntas, Chile (AT02-22)	400–440	−28.93	16.7	1.0–3.0	70–200
Libyan Desert					
SA05-01	30–40	−23.43	>30	N.D.	N.D.
SA05-02	50–60	−21.62	>30	N.D.	N.D.
SA05-03	60–70	−20.06	>30	N.D.	N.D.
Mojave Desert (DV02-10)	145–260	−24.84	9.5	N.D.	15–100
Minas de Rio Tinto					
Sediment (RT04-01)	1050–1400	−24.34	11.4	5–50	50–100
Evaporite (RT04-02)	1200–1500	−23.34	8.2	7–80	70–100
Panoche Valley (PA04-01)	140–180	−27.37	7.4	N.D.	5–20
NASA Mars-1 martian soil simulant	1200–1400	−24.13	11.2	N.D.	100–150

Detection is Easy…

- The argument was that complex carbon containing molecules present in the soil might have been oxidized to CO_2 in the GC oven in the presence of Fe

$$3Fe_2O_3 + H_2 \rightarrow 2Fe_3O_4 + H_2O \quad [1]$$

and

$$Fe_3O_4 + H_2 \rightarrow 3FeO + H_2O \quad [2]$$

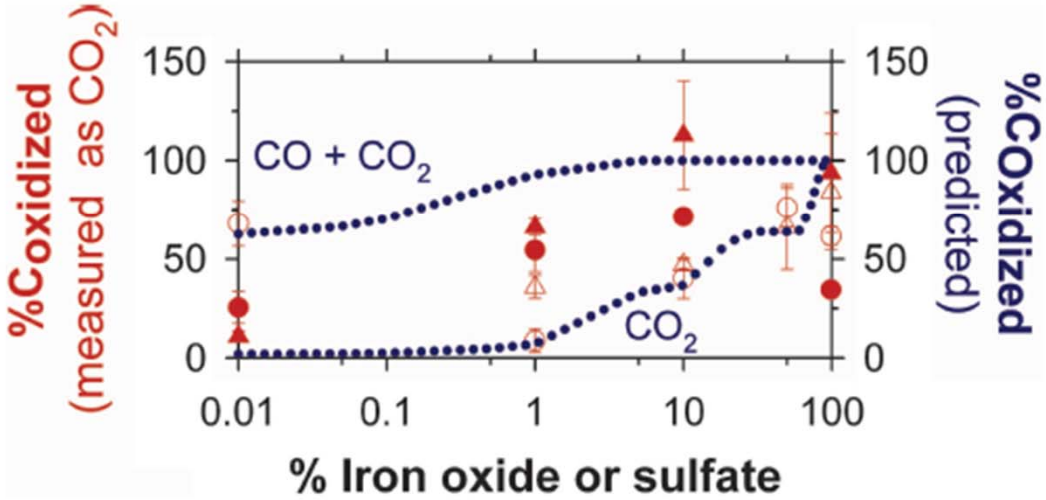


Fig. 5. Oxidation of a 1,000 μg of C from stearic acid with iron species present in silica by flash TV at 750°C in an inert atmosphere composed of helium. Symbols correspond to experimental data, and dotted lines are predicted. Open circles and triangles are Fe_2O_3 and $Fe_2(SO_4)_3$, respectively. Solid symbols indicate values of oxidation with sulfuric acid.

Viking Sucks Rebuttal

- Interestingly, this paper got a lot of press, but few citations. The reason is probably the firm rebuttal that followed from the lead designer of the Viking MS instrument (Klaus Biemann) in a paper entitled:

On the ability of the Viking gas chromatograph–mass spectrometer to detect organic matter

- This paper contains the following scientific smackdown:

Navarro-Gonzalez et al. (18) claim (on page 16092) to have shown "two limitations of the Viking TV [thermal volatilization]– GC–MS for the detection of organic material": (i) that 500 C may be inadequate to release the organic compounds and/or (ii) that these compounds were oxidized during the heating to 500 C by the iron oxides present in the sample.

The first of these statements is contradicted by the results of the extensive tests of the Viking GCMS instrument reiterated above. Although Navarro-Gonzalez et al. (18) cite our paper (3) on the Antarctic soils (their reference 21), they apparently have not read it carefully…

More Smackdown!

- That was followed by this:

The remarkable fact of these measurements and their interpretation is that the lowest level of detection at either temperature is 1 μg/g (1 ppm), i.e., a 1,000-fold poorer sensitivity than the 1 ng/g sample (1 ppb) demonstrated with the Viking engineering breadboard instrument (see above and Fig. 1 and Table 1). The lack of sensitivity seems to be due to the experimental design. The investigators combined three commercially available laboratory instruments, a pyrolizer, a gas chromatograph (using a column suitable only for the separation of low-polarity organic compounds containing seven or fewer carbon atoms), and a quadrupole mass spectrometer scanning from m/z 12–100 or 45–200. For some reason, only benzene was reported, rather than all of the compounds evolved upon heating the sample.

The assumption that benzene is always the major pyrolysis product is naive. It would have been more convincing to present the entire chromatogram, including amounts detected, of at least a few representative experiments as it was done for the Viking GCMS tests (2, 3).

Organics Detected...

- This was followed by these results on the real viking GC-MS showing that it could detect organics... even in Antarctic soil...

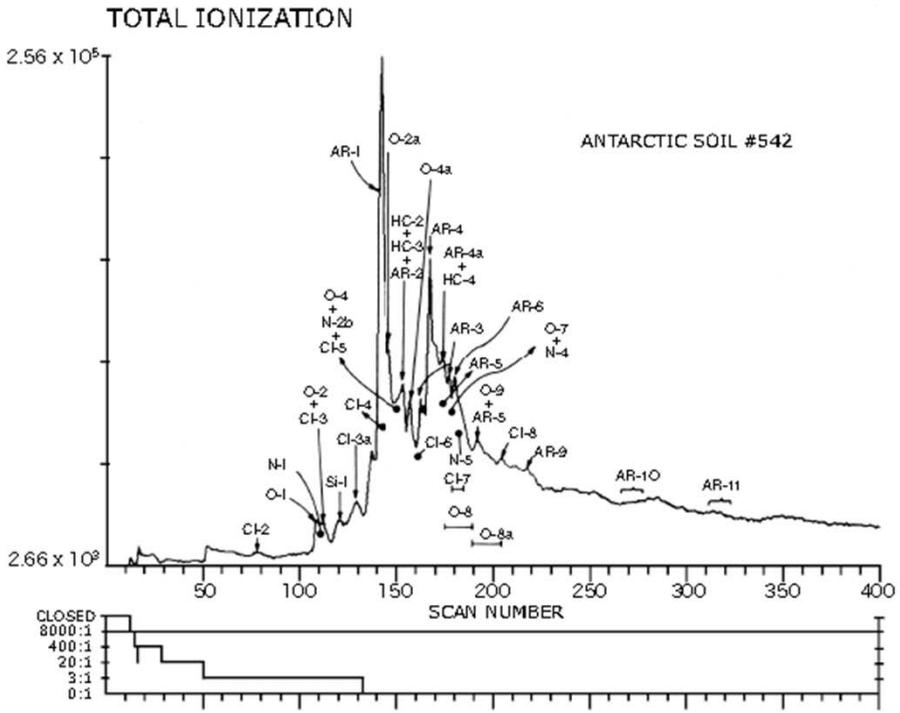
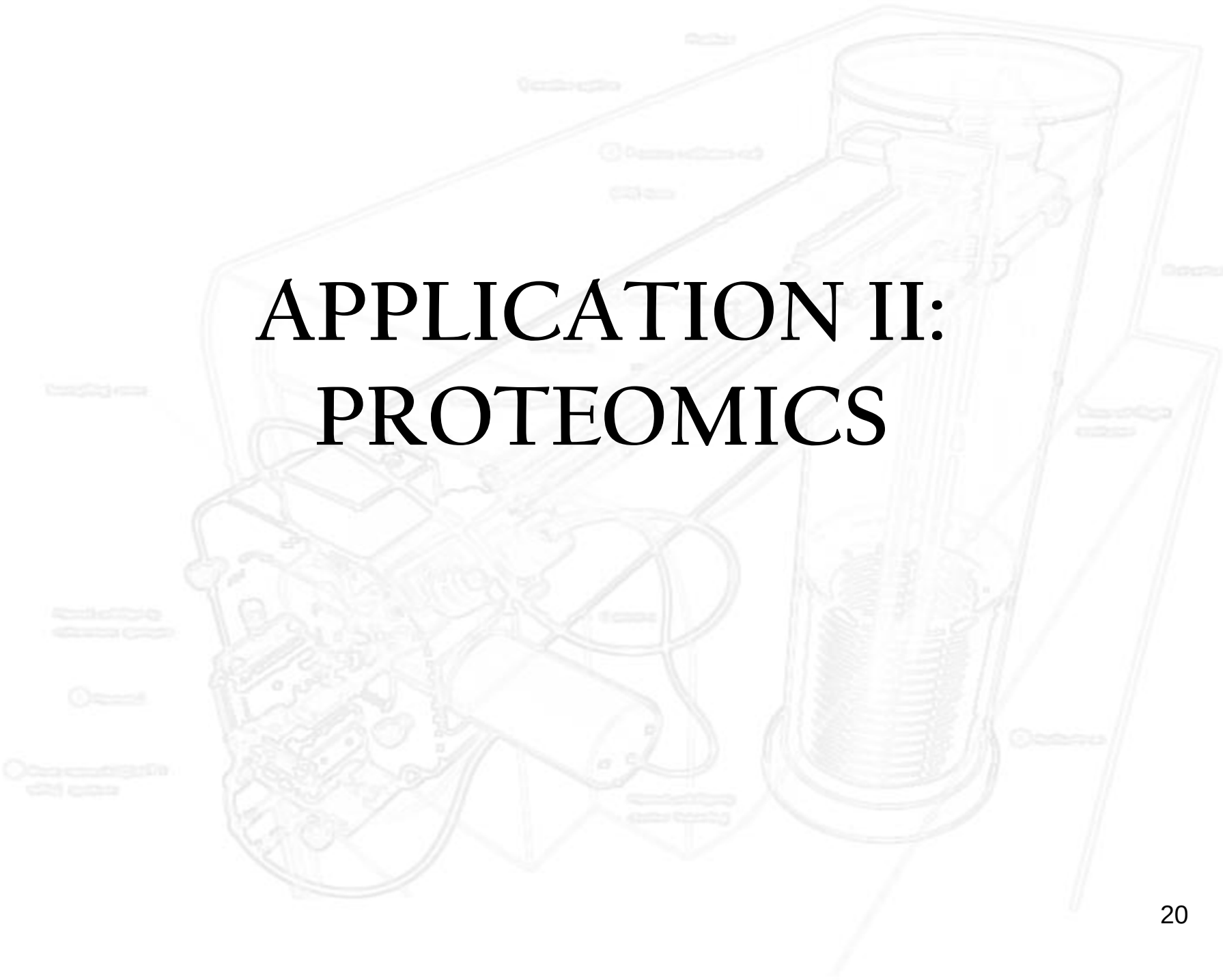


Table 1. Compounds identified (in ppb) by the Viking GCMS upon heating Antarctic soil #542 (8) to 500°C

Code in Fig. 1	Compound name	ppb
AR-1	Benzene	90
AR-2	Toluene	20
AR-3	Phenyl-C_2	90
AR-4	Styrene	100
AR-4a	Methylstyrene	4
AR-5	Phenyl-C_3	20
AR-6	Phenyl-C_4	10
AR-9	Naphthalene	10
AR-10	C_1-naphthalene	2
AR-11	Biphenyl	10
HC-2	Cyclooctane	100
HC-3	Hexane	70
HC-4	Heptane	70
N-1	Acetonitrile	100
N-2b	Vinylacetonitrile	40
N-4	Benzonitrile	20
N-5	Methylbenzonitrile	4
O-1	Furan	20
O-2	Acetone	200
O-2a	Methylmethacrylate	90
O-4	Methylvinylketone	200
O-7	Benzofuran	2
O-8	Phenol	20
O-8a	Cresol	10
O-9	C_1-benzofuran	1

APPLICATION II: PROTEOMICS



And Now for Something Completely Different…

- Proteomics is the 'catchment term' for any research aimed at characterizing the protein complement of the cell (or a subset thereof).

- The term Proteomics comes from Genomics, a field centered on the characterization of the entire gene complement of various organisms.

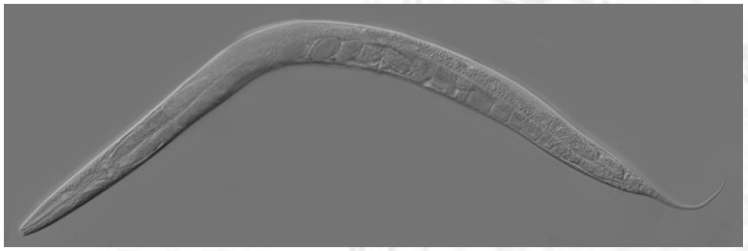
- Genomics was initiated by Sanger (Nobel Laureate x2) who sequenced the entire genome of a bacteriophage in 1977.

- The field of proteomics didn't get started until around 15 years later, mainly due to the development of 2D page electrophoresis and ESI/MALDI MS.

Why the Proteome?

- We care about the proteome because it is there, and NOT the genome level that the complexity of life truly arises.

witness *C. elegans*
(a flatworm)



genes: ~20,100

unique gene products:
~25,600

witness the human
(a primate)

genes: ~25,000

unique gene products:
~477,000

Mass Spectrometry and Proteomics

- Currently, the vast majority of proteomics research efforts are enabled by mass spectrometry

- This is because MS combines extreme selectivity (the ability to distinguish multliple coexisiting species in solution) and very good sensitivity (the ability to detect analytes at low concentrations).

- Sensitivity is needed because some very important proteins exist in the cell at very low copy number. There is also a huge range of copy numbers so that low abundance proteins are often obscured by high abundance proteins.

- Selectivity is needed because proteomic samples invariably involve a large (sometimes massive) number of proteins and/or peptides that need to be simultaneously detected...

Challenge 1: Expression Levels

- Proteins are expresses in a huge range of concentrations in the cell. This can result in masking of low-copy proteins by high-copy ones or simple failure to detect low-copy proteins...

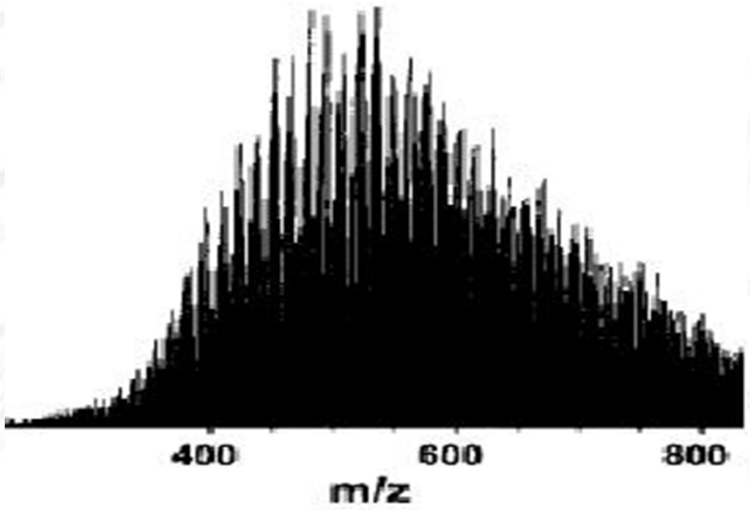
Table 1. Summary of Proteins Detected and Quantified In Each Group of Cellular Abundance

Abundance Range (Copies/Cell)	Group	Proteins Measured	Protein Names[a]	Absolute Quantification[b]
524,288–1,255,722	1	5	YGL008C, YKL060C, YLR355C, YLR249W, YDR382W	YKL060C
262,144–524,288	2	5	YJR104C, YML028W, YMR116C, YCR012W, YER091C	YJL136C
131,072–262,144	3	5	YDR050C, YER165W, YGR192C, YER177W, YNL178W	YLR249W
65,536–131,072	4	5	YBR127C, YHR183W, YKL182W, YHR208W, YDL126C	
32,768–65,536	5	5	YLR058C, YML008C, YIL078W, YAL012W, YGR204W	YHR183W, YLR058C,
16,384–32,768	6	5	YBR249C, YJR105W, YNR016C, YLR216C, YGR209C	YBR249C
8,192–16,384	7	5	YJL136C, YDR368W, YJL130C, YOR007C, YMR099C	YJL026W
4,096–8,192	8	5	YKR048C, YER006W, YML086C, YKR001C, YER003C	
2,048–4,096	9	5	YFL014W, YDR129C, YPL235W, YOL140W, YMR170C	YEL031W, YHR107C, YPR118W, YJR051W
1,024–2,048	10	10	YDL021W, YML100W, YKL150W, YEL031W, YGL202W, YDL017W, YGR080W, YPL049C, YGL248W, YEL011W	YMR170C, YCL017C
512–1,024	11	10	YHR107C, YGL100W, YBR208C, YPR118W, YJL172W, YBR283C, YCR088W, YGR256W, YJL026W, YCL030C	YOL116W
256–512	12	10	YCL017C, YOL116W, YNL161W, YJR051W, YKL068W, YHR138C, YGR232W, YMR199W, YOR267C, YJR134C	YGL248W
128–256	13	10	YKL141W, YHR074W, YLR330W, YDR436W, YKL129C, YOR020C, YBR117C, YBR125C, YKL073W, YOL022C	YIL084C, YML109W
<128	14	15	YLL040C, YNL014W, YML109W, YIL092W, YIL084C, YKL145W, YKL075C, YIL002C, YHR015W,YPL008W, YGL006W, YKR031C, YLR035C, YNR067C, YOR093C	YGL006W, YNR067C, YKR031C
No expression detected[a]	15	15	YDR381W, YNL208W, YHR020W, YNL160W, YEL024W, YJL008C, YJL111W, YFL037W, YDR023W, YJR123W, YLR340W, YJR077C, YDR321W, YCL018W, YER055C	
Below QOD (<50 copies/cell)[a]	16	6	YBR006W, YCL043C, YDR150W, YOR120W, YJL167W, YBR006W	
Western blot band not quantifiable[a]	17	6	YHR029C, YNL055C, YJL080C, YDL140C, YIR006C, YGR284C	
Never observed in publicly accessible proteomics data sets[c]	18	10	YDL017W, YOL116W, YBR117C, YIL092W, YKL075C, YIL002C, YHR015W, YPL026C, YLR035C, YOR093C	

Cell *138*, 795–806, August 21, 2009 ©2009 Elsevier Inc.

Challenge 2: Number of Proteins

- Obviously, there are a huge number of proteins in the cell. If we tried to just look at them by MS, we might see something like this (at best):



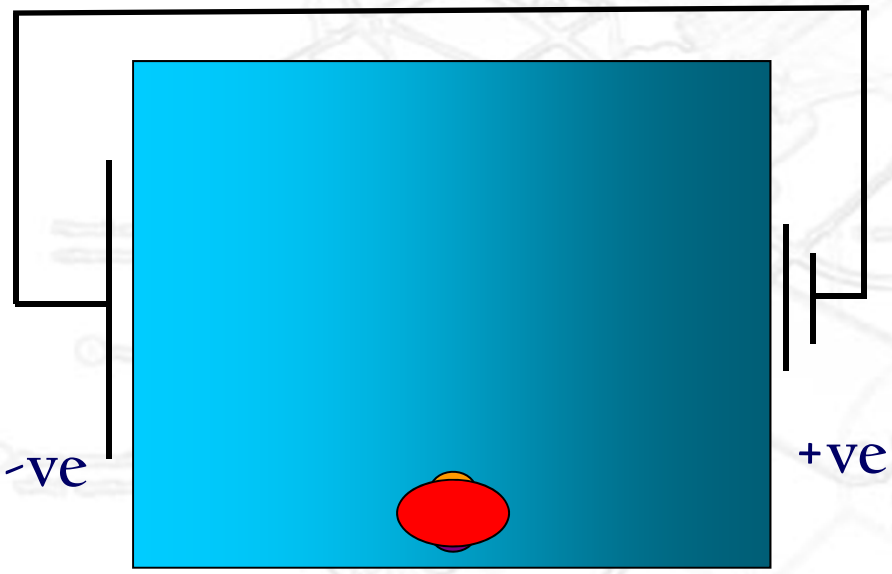
- This is where separation techniques come in…

- The first of these was 2D gel electrophoresis, which is carried out on whole proteins (vide infra).

2D Electrophoresis

- It's impossible to talk about proteomics without mentioning the separation technique that got it all started…

- Step 1: Separate the proteins based on their unique pI (the pH at which they are neutral) by making a pH gradient in the gel.

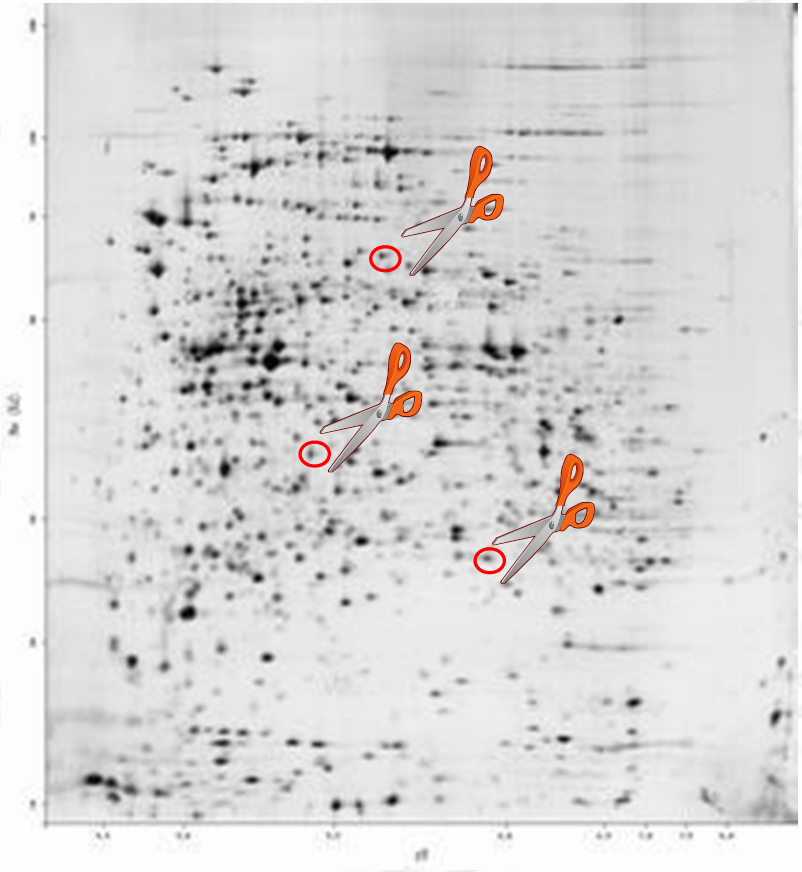


- Step 2: Apply detergent (SDS) so that all are negatively charged in proportion to their mass.

- Step 3: Separate by mass.

2D Electrophoresis Cont.

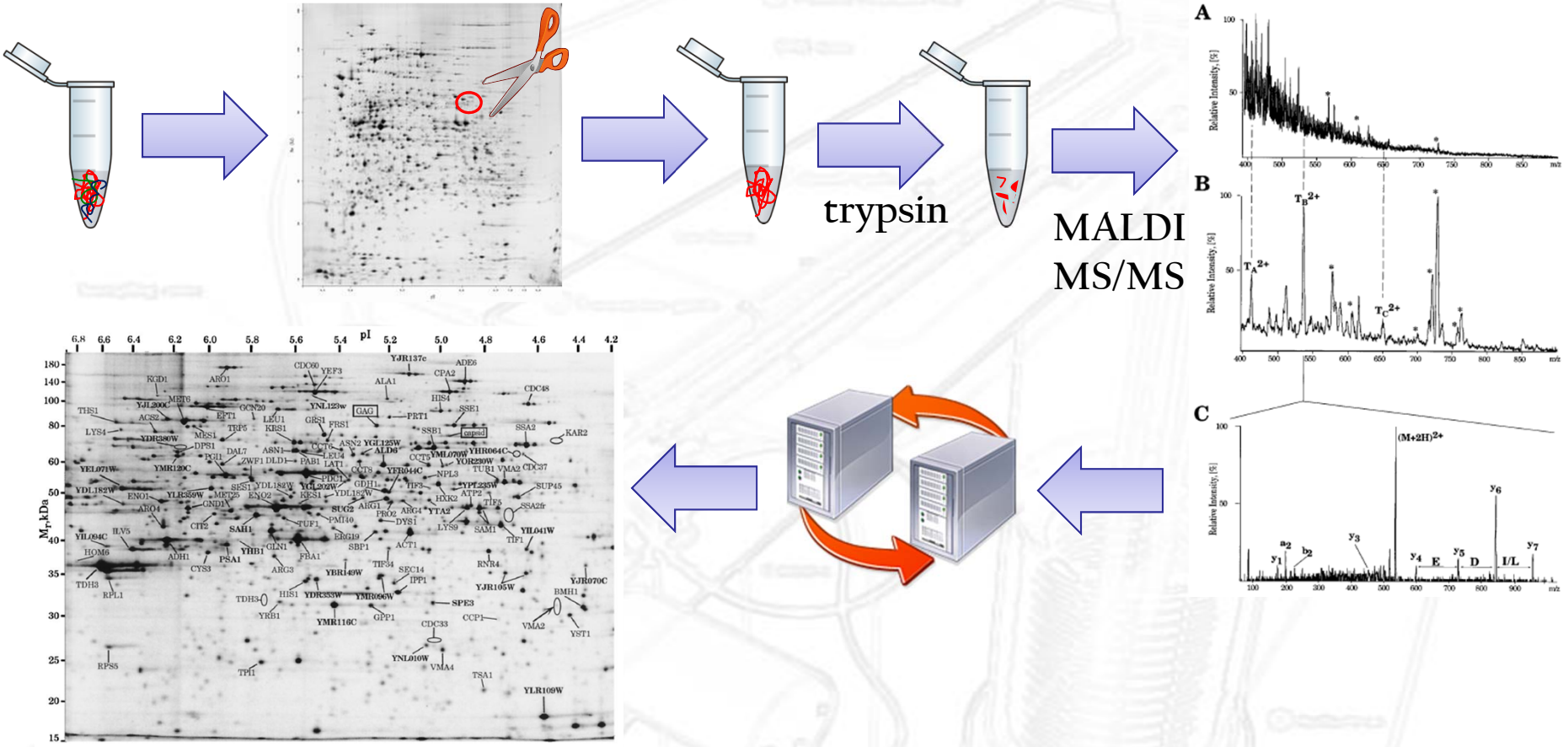
- In order to make the same 'spot' on the 2D gel, two proteins would have to have the same pI and molecular weight!



- We can then cut out these 'spots'…

- And do what we want with them, most likely resolvation followed by trypsin digestion to yield peptides followed by nano-ESI or MALDI-MS

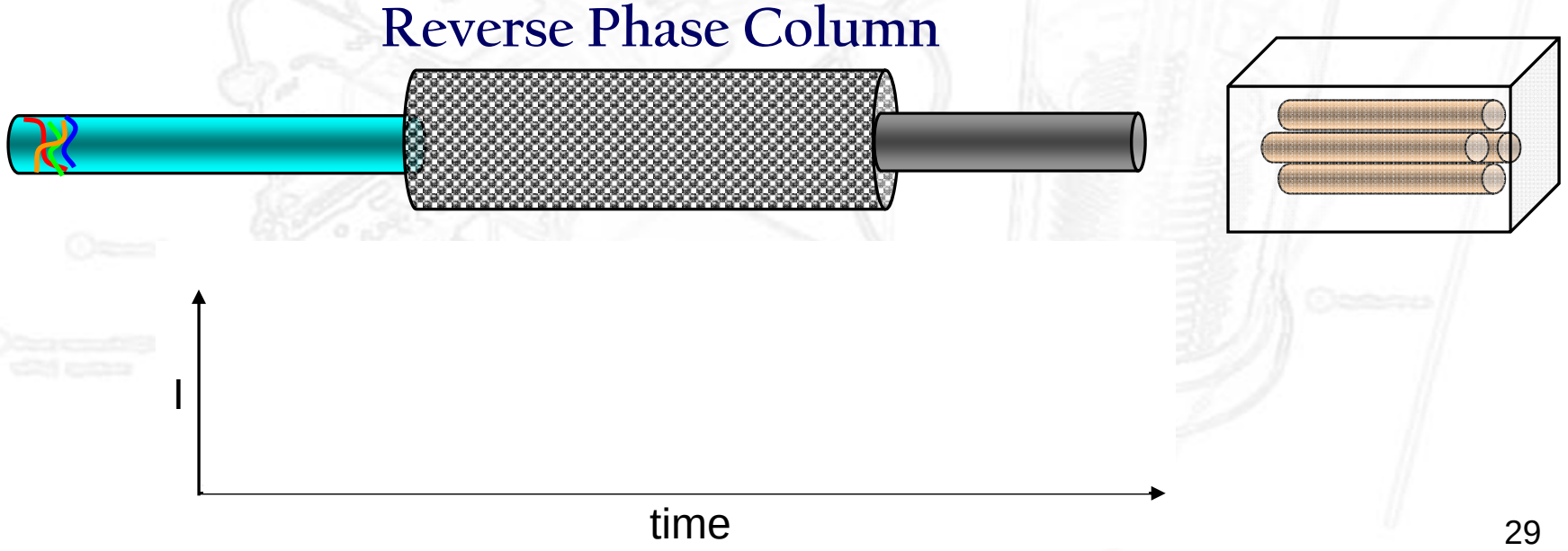
Workflows: Flavors of Proteomics



- The above corresponds to a proteomic workflow which really defines the nature of the proteomic experiment.

Interlude: LC-MS

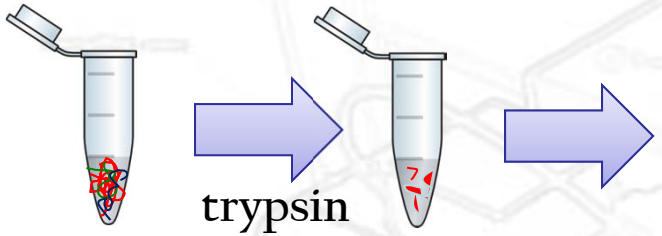
- The 2D gel workflow is actually quite labor intensive. These days most people do LC-MS for their proteomics, where the separation technique can be directly coupled to nanospray ESI-MS.
- For separating peptide, people typically use 'reversed phase' HPLC which relies on hydrophobic interactions between peptides and modified silica beads…



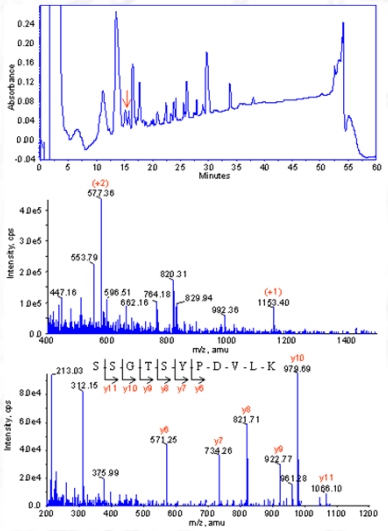
Types of Proteomic Experiment:

- In general, proteomics can be subdivided into two types based more-or-less on whether enzymatic digestion is used:

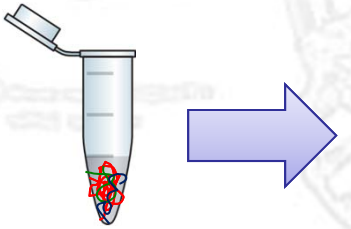
- Bottom up:

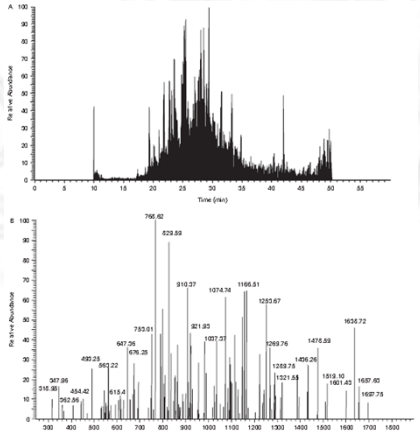


LC/MS



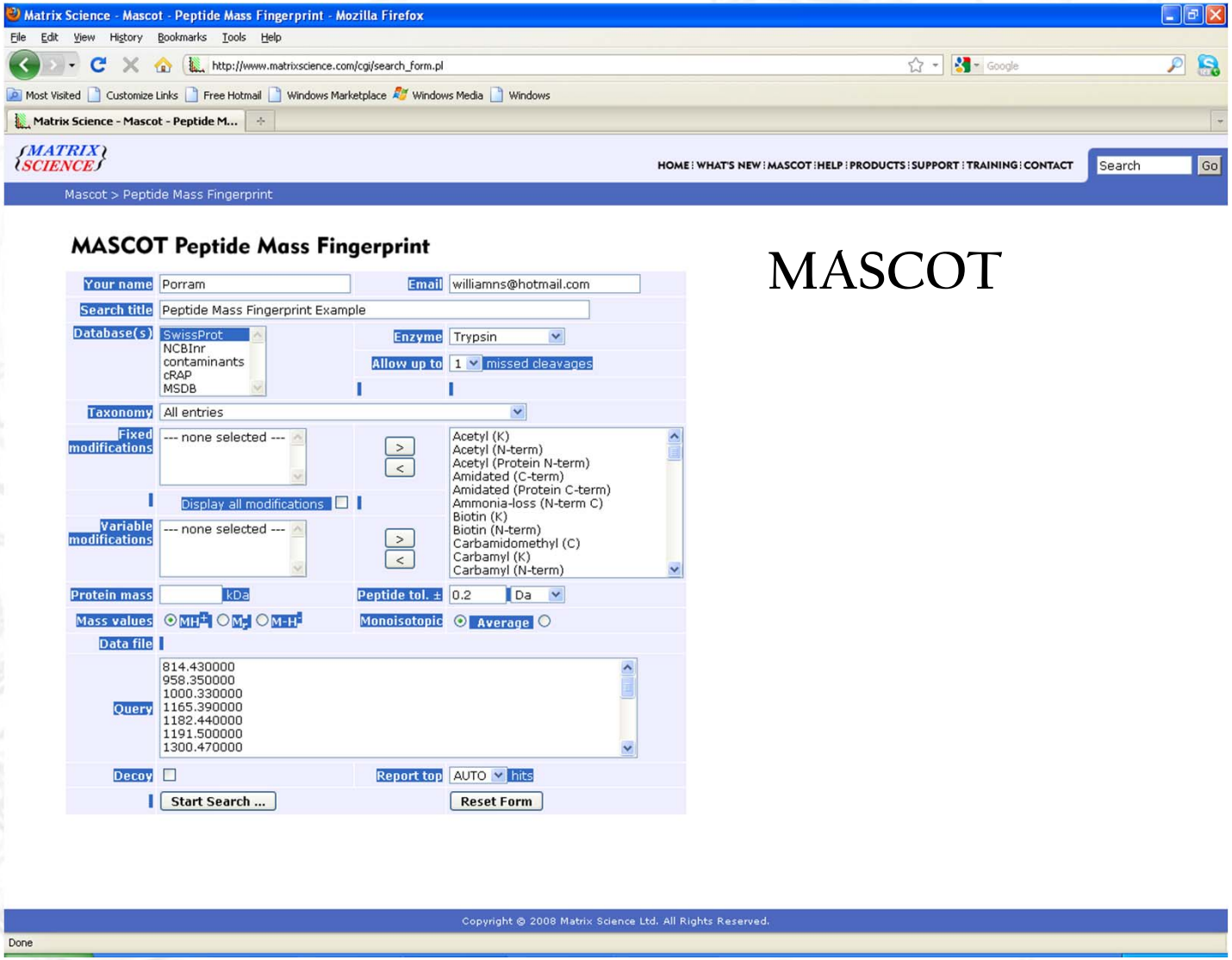
- Top Down:





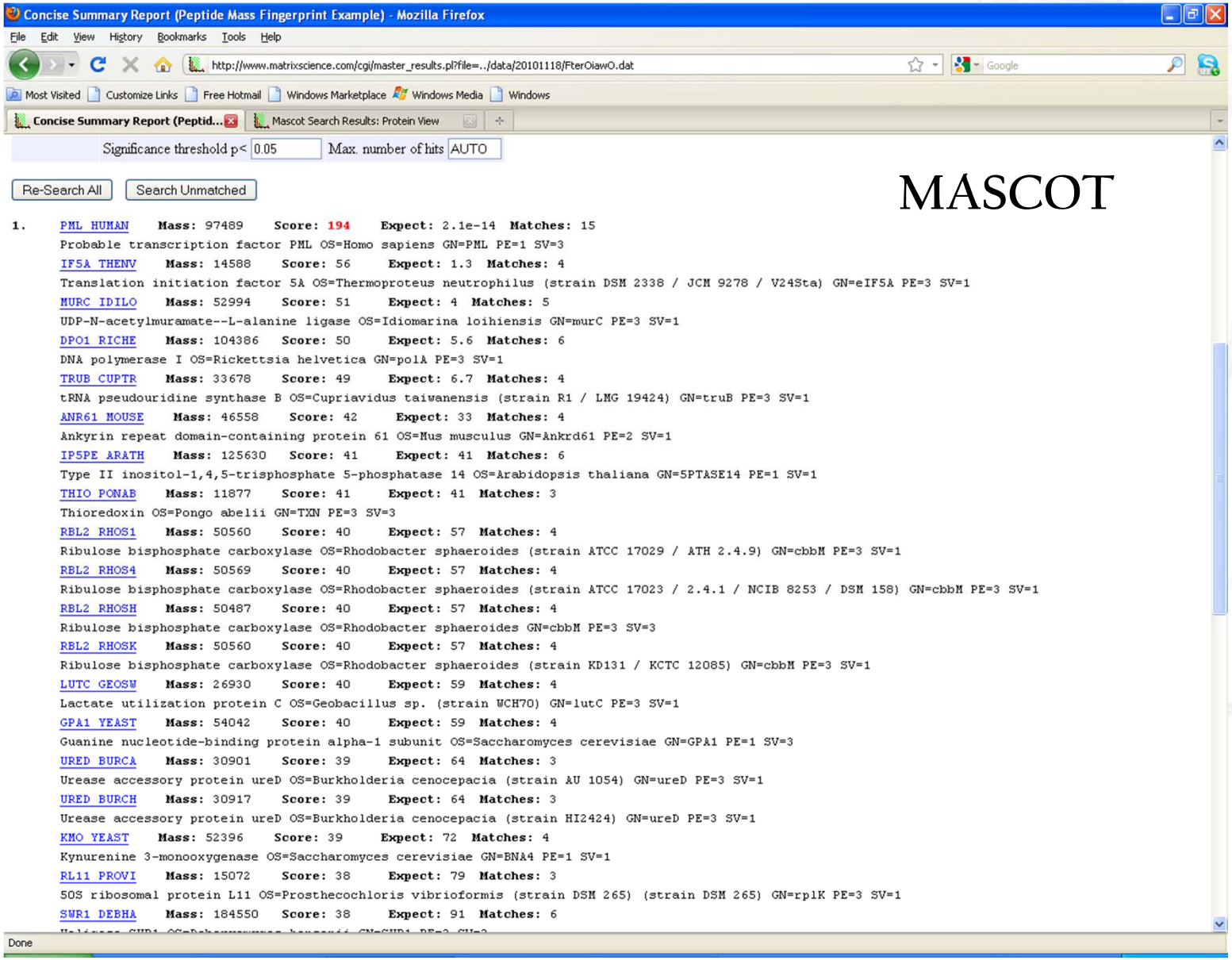
Proteomics Databases...

- In a typical proteomics experiment, proteins are 'identified' based on the presence of a small number of peptides whose sequence is within the overall protein sequence... for example:



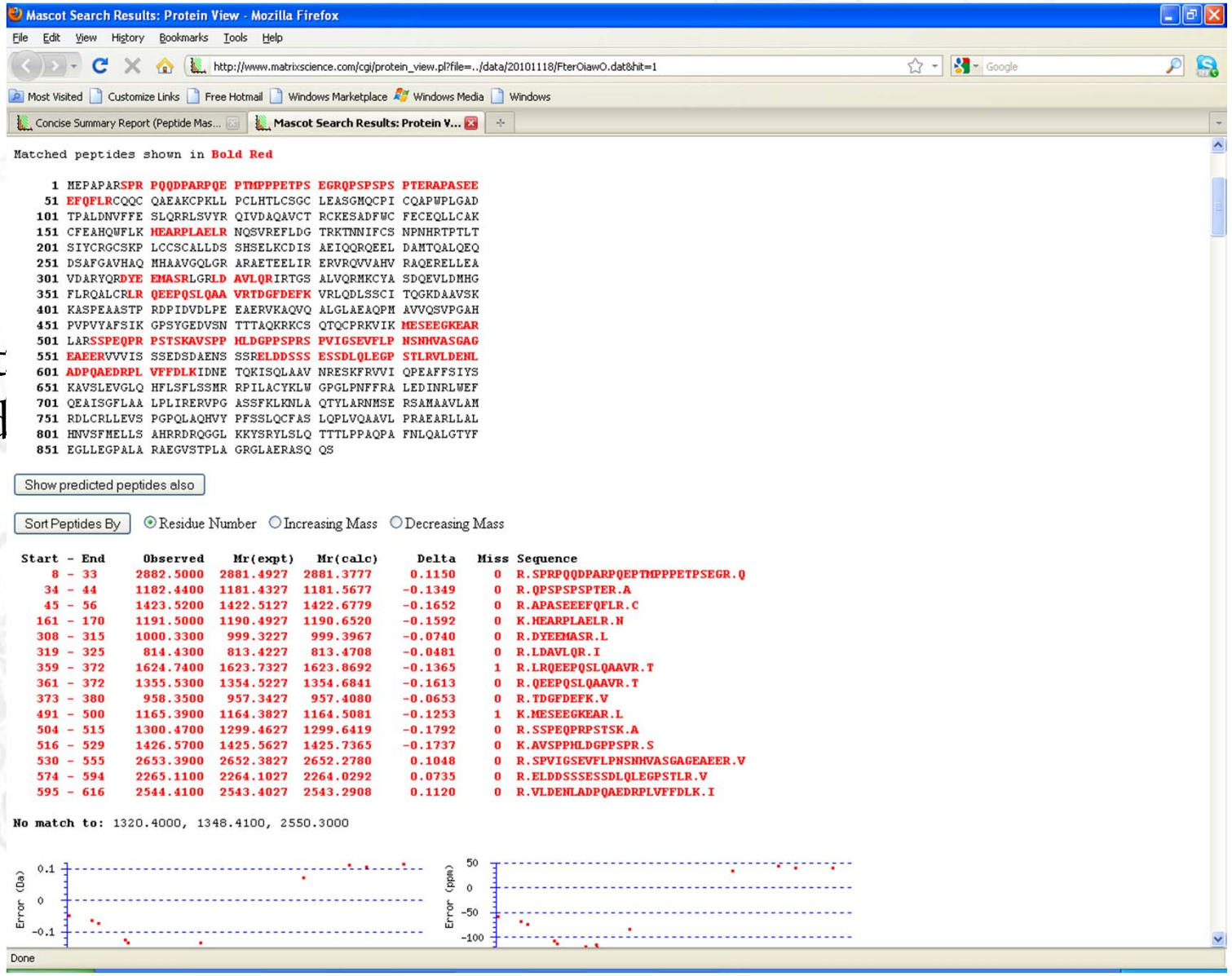
MASCOT

More databases



More databases

- In the background, mascot decided that the masses that we input corresponded to the following peptides:



Applications of Proteomics

- There are a huge number of applications for proteomics including:
  - Characterization of the proteome vs. genome
  - Monitoring cellular metabolism in response to stimuli
  - Early detection of disease/cancer (clinical tests)
  - Highly specific identification of disease/cancer (quantitative)
  - Post-translational modifications (*e.g.* phosphorylation, epigenetics)

Case Study: Samuel Lunefeld Research Institute @ Mount Sinai Hospital in T.O.

- The SLRI is one of Toronto's biggest research institutes with close ties to U of T and the university hospital network.

- The institute has many areas of expertise, but there is a focus on clinically oriented systems biology, a big part of which is proteomics studies...

- As an example, we'll look at some of the proteomics studies coming out probably the biggest name group at SLRI, the Pawson group...

Pawson Group: Cancer Detection

Cell-Specific Information Processing in Segregating Populations of Eph Receptor Ephrin–Expressing Cells

Claus Jørgensen,[1] Andrew Sherman,[1,2] Ginny I. Chen,[1,2] Adrian Pasculescu,[1] Alexei Poliakov,[3] Marilyn Hsiung,[1] Brett Larsen,[1] David G. Wilkinson,[3] Rune Linding,[4]* Tony Pawson[1,2]*