Mass Spectrometry for Metabolomics and Proteomics

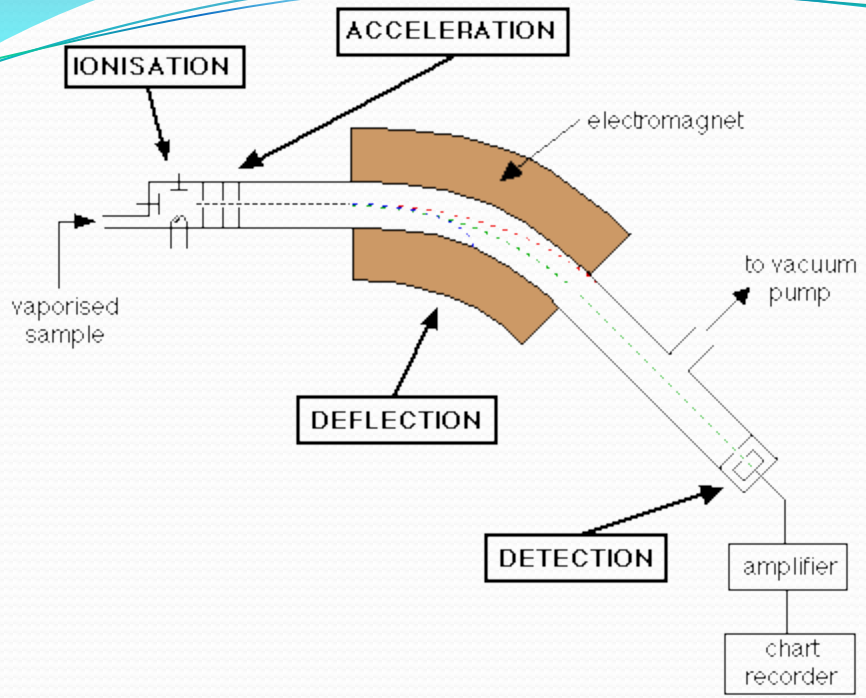
BST 226
Statistical Methods for Bioinformatics
David M. Rocke

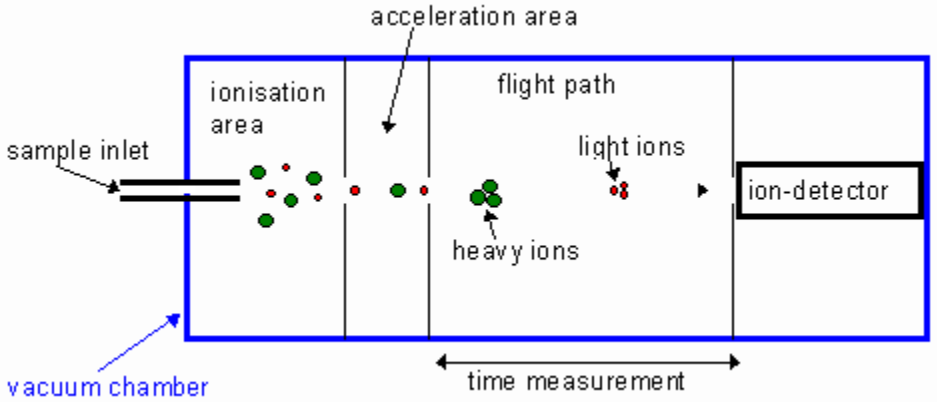
Mass Spectrometry

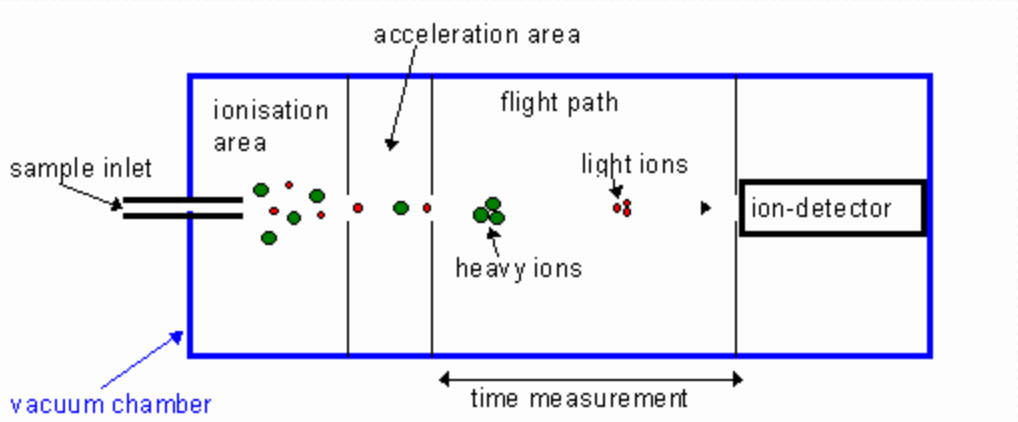
- Mass spectrometry (mass spec, MS) comprises a set of instrumental methods that can identify compounds by molecular weight and can potentially quantify by using the peak height or peak area.
- If the substance to be analyzed is not already a gas, it needs to be vaporized and also ionized so that the molecules are charged, most commonly by elimination of an electron.
- The behavior of the ion can be measured in terms of the mass-to-charge ratio (m/z) because a 100 Dalton + molecule behaves in a magnetic field like a 200 Dalton ++ molecule.

Physics

- One method is to detect the amount of deflection as the ions traverse an evacuated tube in a magnetic field. The amount of deflection at the end of the tube is smaller when m/z is larger. Varying the magnetic field yields a spectrum of total ion current at each field strength = at each m/z
- Time-of-flight MS is used for larger molecules. The ions are accelerated at the beginning of the tube, and the velocity larger for smaller m/z, so the time of flight is larger. The spectrum is collected over time and the x-axis converted to m/z.
- Varieties of ion trap MS circulate the ions in a magnetic field and generate a signal that can be deconvolved with a Fourier transform.

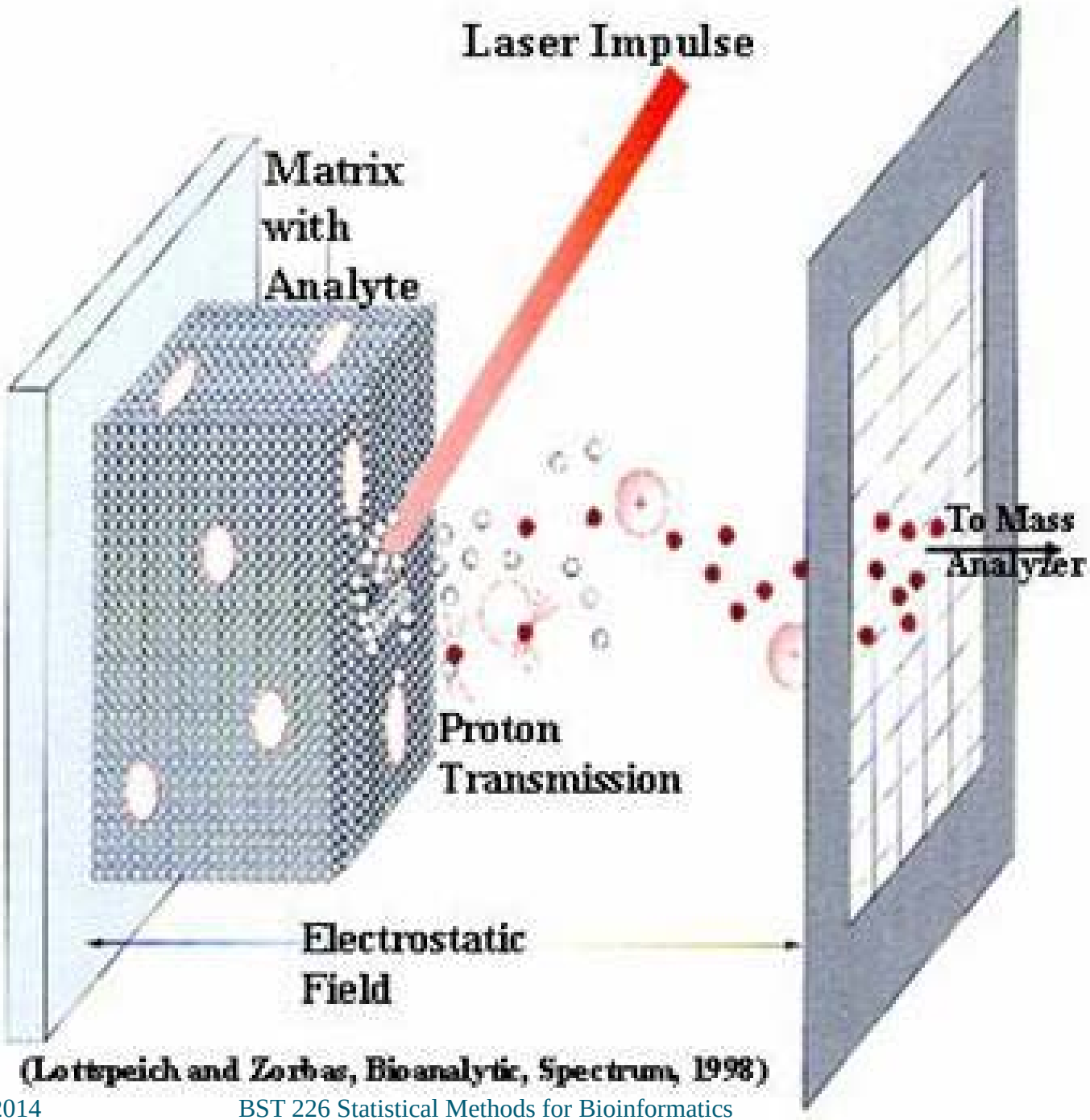

IONISATION
ACCELERATION
electromagnet
vaporised sample
to vacuum pump
DEFLECTION
DETECTION
amplifier
chart recorder


acceleration area
ionisation area
flight path
sample inlet
light ions
ion-detector
heavy ions
vacuum chamber
time measurement


acceleration area
ionisation area
flight path
sample inlet
light ions
ion-detector
heavy ions
vacuum chamber
time measurement

MALDI

- Matrix Assisted Laser Desorption Ionization
- Sample is embedded in matrix on a slide
- Laser energy absorbed by matrix, ionizes and vaporizes matrix and sample.
- Can fragment compounds
- Can doubly or triply ionize compound—Look for peak at half or a third of the expected m/z
- Results of different shots can be very different—multiple shots for each sample


Laser Impulse
Matrix
with
Analyte
To Mass
Analyzer
Proton
Transmission
Electrostatic
Field
(Lottspeich and Zorbas, Bioanalytic, Spectrum, 1998)

Electrospray Ionization

- Electrospray ionization uses electricity to disperse a liquid
- High voltage is applied to a liquid supplied through an emitter (usually a glass or metallic capillary).
- This leads to the formation of small and highly charged liquid droplets, which are radially dispersed due to Coulomb repulsion.
- This is often used between a separation stage (liquid chromatography) and the mass spec analyzer.

Separation Technologies

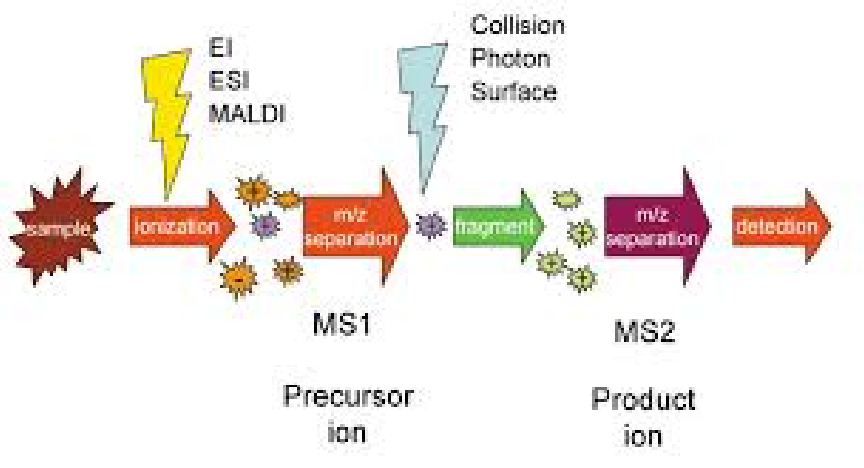
- Various chemical methods can be used to produce a sample that contains mostly the class of analyte of interest.
- These can be small molecule organic compounds, lipids, oxy-lipids, sugars, proteins, etc.
- Gas or liquid chromatography will separate the sample by various chemical properties and then aliquots will be serially analyzed by mass spec.
- This can help differentiate molecules that have the same weight because they elute at different times.

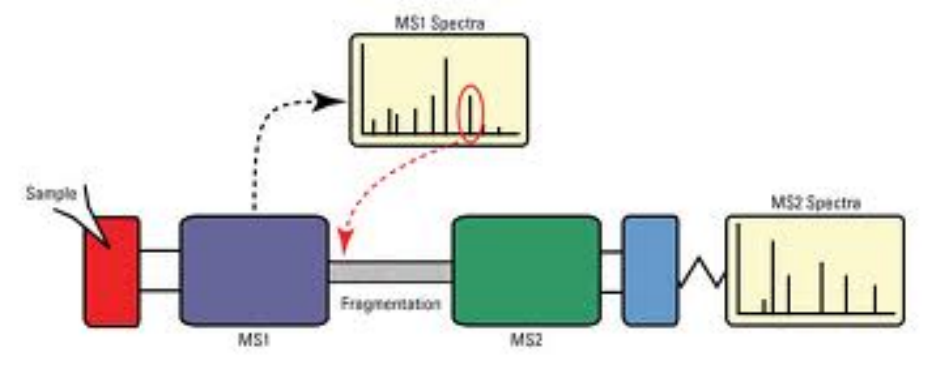
Metabolomics

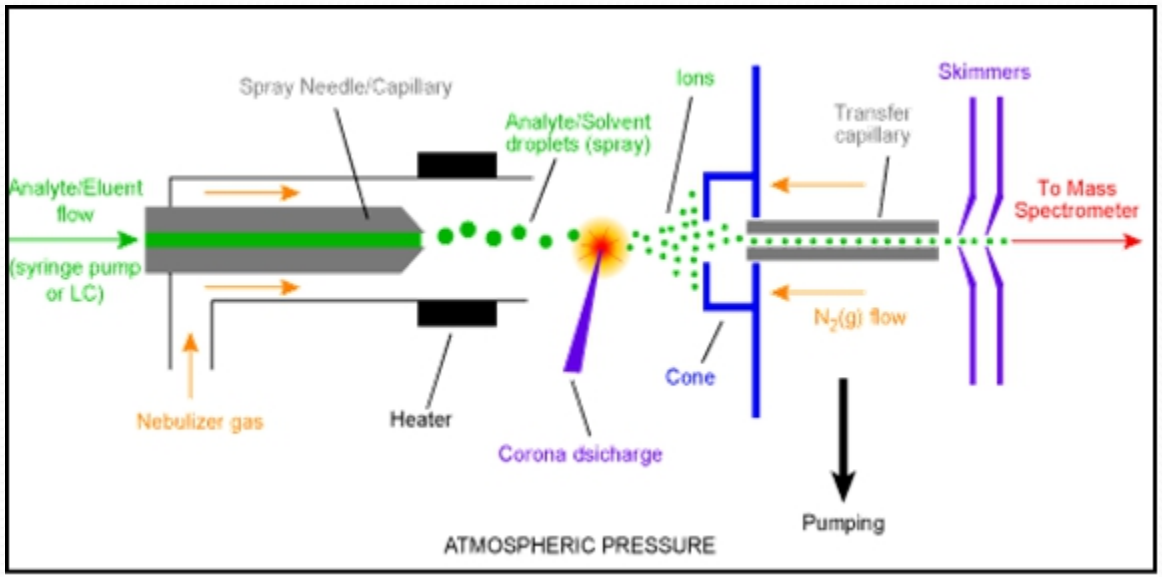
- Mass spectrometry (e.g., LC/ToF-MS can detect and measure relative amounts of small molecules much more easily than with proteins
- Lipids, saccharides, and others
- For example we can speciate 500 lipids including di- and tri-glycerides species
- For example, we can measure over a hundred compounds in the arachidonic acid pathway including prostaglandins, COX1, COX2, and LOX products, and CyP450 pathways eicosanoids.
- Can measure enzymes, substrate, and product

Proteomics

- Proteins can be very large molecules, up to several thousand amino acids or hundreds of thousands of Daltons.
- These must be broken up into much smaller peptides, usually chemically with a protease, before they can possibly be ionized.
- These peptides are then usually further fragmented in a second chamber and the spectrum of fragment sizes recorded. This is called tandem mass spec or MS/MS


EI
ESI
MALDI
Collision
Photon
Surface
sample
ionization
m/z separation
fragment
m/z separation
detection
MS1
MS2
Precursor ion
Product ion


MS1 Spectra
Sample
MS2 Spectra
Fragmentation
MS1
MS2


Spray Needle/Capillary
Analyte/Solvent droplets (spray)
Ions
Skimmers
Transfer capillary
Analyte/Eluent flow
(syringe pump or LC)
To Mass Spectrometer
N_2(g) flow
Cone
Nebulizer gas
Heater
Corona dsicharge
Pumping
ATMOSPHERIC PRESSURE

Processing Spectra

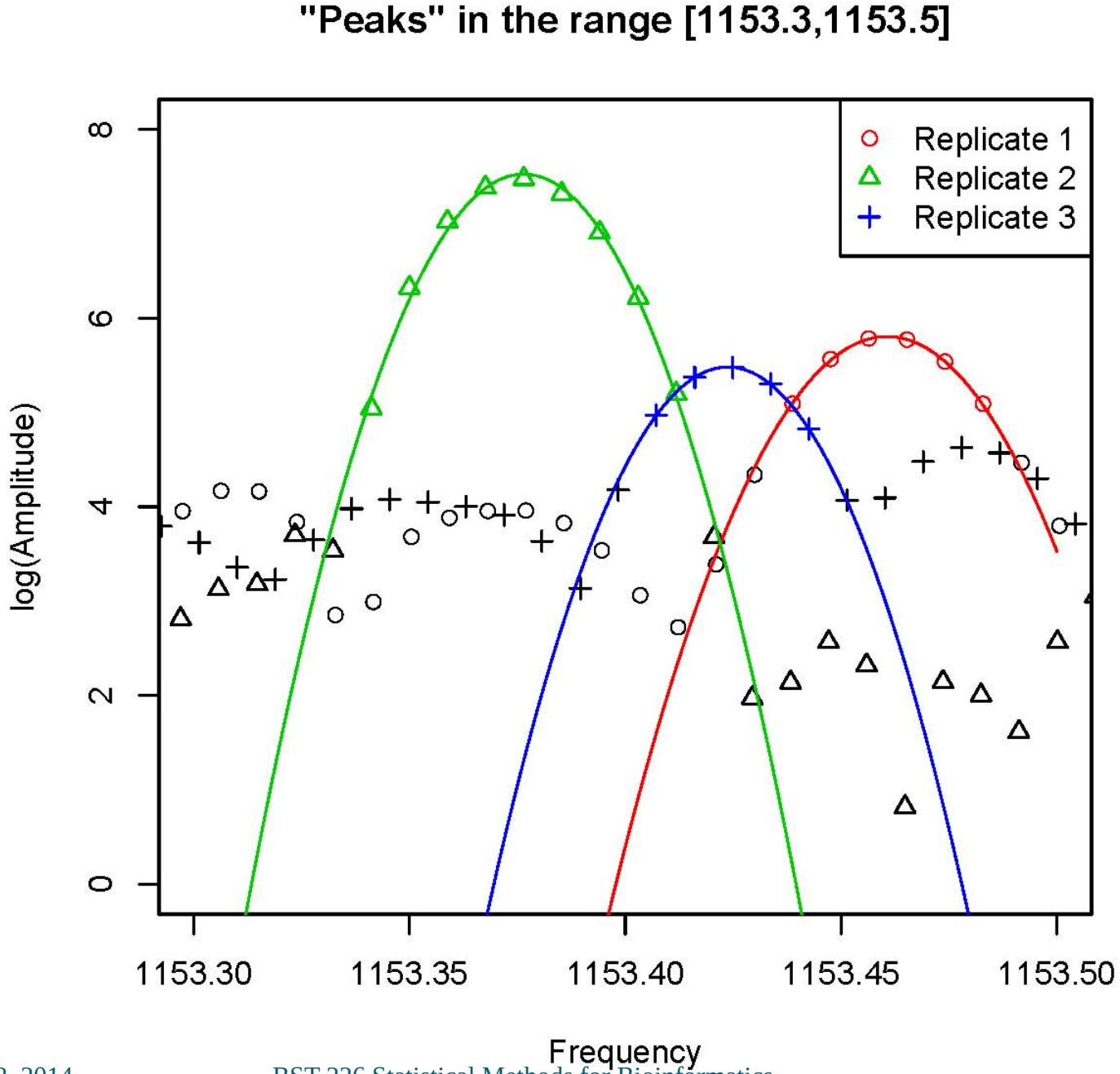
- Baseline estimation
- Peak identification
- Mass calibration
- Data transformation
- Normalizing across spectra
- Peak quantitation
- Identification of isotope and shoulder peaks
- FTICRMS R package

Baseline Correction

- In regions of the spectrum where no analyte is present, the signal should fluctuate around a baseline, which should be set to zero
- In peak regions, the estimated baseline forms the base for peak heights or area.
- For some technologies, the baseline tends to be relatively flat, but for others, such as NMR and FT-ICR MS, the baseline is curved or wavy.
- It is important to get the baseline right to find peaks and accurately measure them.

Peak Identification

- In some cases, a peak is just a point on the spectrum where the points to the left and right have lower magnitudes.
- But this can result in false multiple peaks from jagged signal.
- For FT-ICR MS, the peaks look quadratic on the log scale which allows accurate identification.


"Peaks" in the range [1153.3,1153.5]
Replicate 1
Replicate 2
Replicate 3
log(Amplitude)
0
2
4
6
8
1153.30
1153.35
1153.40
1153.45
1153.50
Frequency

Mass Calibration

- This is supposedly done by in-machine software using known large peaks
- In principle, the mass accuracy of FTICRMS at 1000 Daltons should be 0.001 Daltons or better
- In practice, as shown on the previous slide, the same compound can be as much as 0.1 Daltons apart
- Because the theoretical mass accuracy of FT-ICR is so high, this mass calibration process can be done very effectively.
- The accuracy of ToF MS and NMR is much worse.

Data Transformation

- Data from MS and most other technologies needs to be transformed, usually to a log scale to make analysis effective.
- Care needs to be taken at the low end .
  - Logs of zero and negative numbers are not defined
  - Signal fluctuating around a zero baseline will often be negative.
- Shifted log (add a constant), generalized log, and other methods can be used.

Statistical Analysis

- Use only peaks that are present in a minimum number of samples, impute peaks for samples in which peaks are not detected
- Appropriate statistical analysis per peak depending on the design (e.g., one-way ANOVA, two-way ANOVA, linear regression).
- Correct for multiple comparisons using Benjamini Hochberg False Discovery Rate control methods
- Determine signatures by a variety of methods: logistic regression, PAM, SVM

Proteomics

- Quantitative proteomics by MS/MS is much more difficult since we never actually see the signal from a whole protein (in contrast to ELISA or Luminex where we do).
- The process of analysis is to take each peptide whose mass is measured in stage 1 and assign it to a protein, if possible, by the fragment spectrum in stage 2
- For each protein (and for humans we have a kind of a list in advance) we count the unique peptides that map to that protein and use that as the score for the protein.
- This can then be analyzed with (for example) overdispersed Poisson regression.