Mass Spectrometry Analysis of Proteins Using Electron Transfer Dissociation

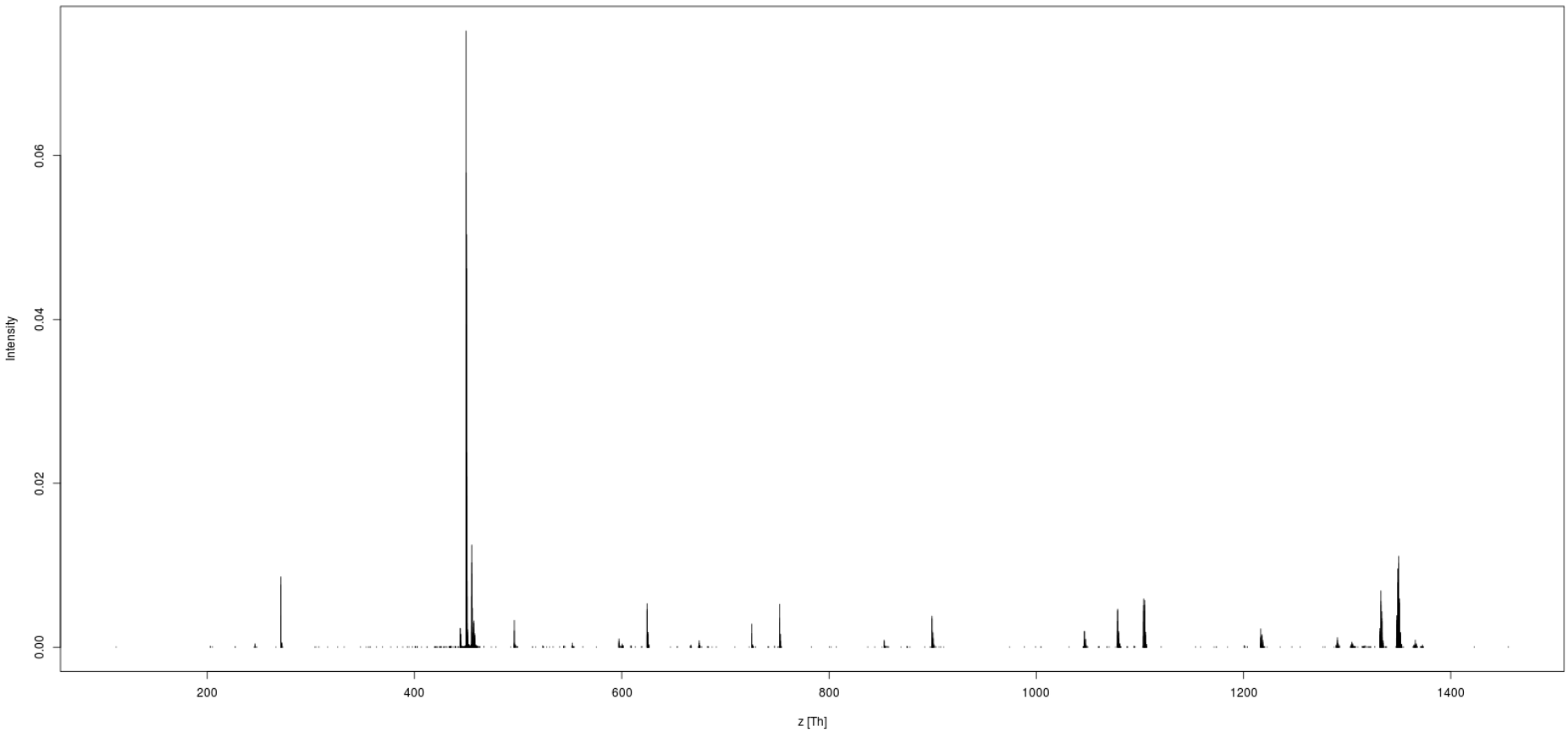
Statistical Model

Mateusz Łącki

Uniwersytet Warszawski

17 October 2013

Data from a Mass Spectrometer for a given substance.



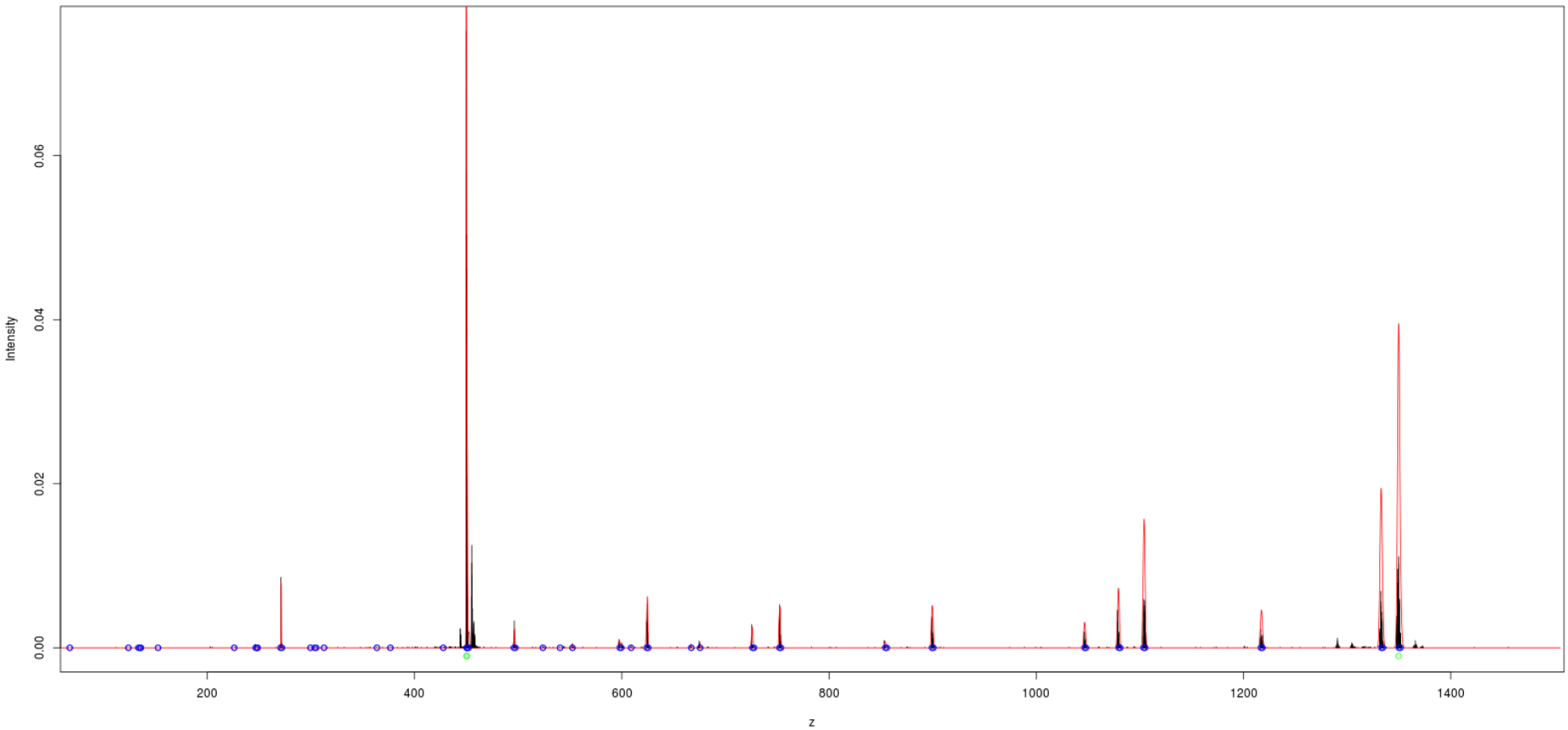
Today's Agenda

- Understanding ETD statics
- Next step : understand dynamics
- Quantitative approach : potential characterisation of peptides through the use of MASS SPEC. But not certain yet how to do it.

Project MassTodon: Debriefing

- Answers to:
  - What results from applying several actions on many molecules of a given type? Actions such as
    - Cleavage of peptide bond
    - Lost of proton
    - Gain of electron
  - Which reaction trails are most probable?
- Maturity?
  - Non-optimised reactions enumeration.
  - Non-calibrated fitting procedures.
- But who cares? To some extent - it works (for one substance).

Theoretically Explained Spectrum

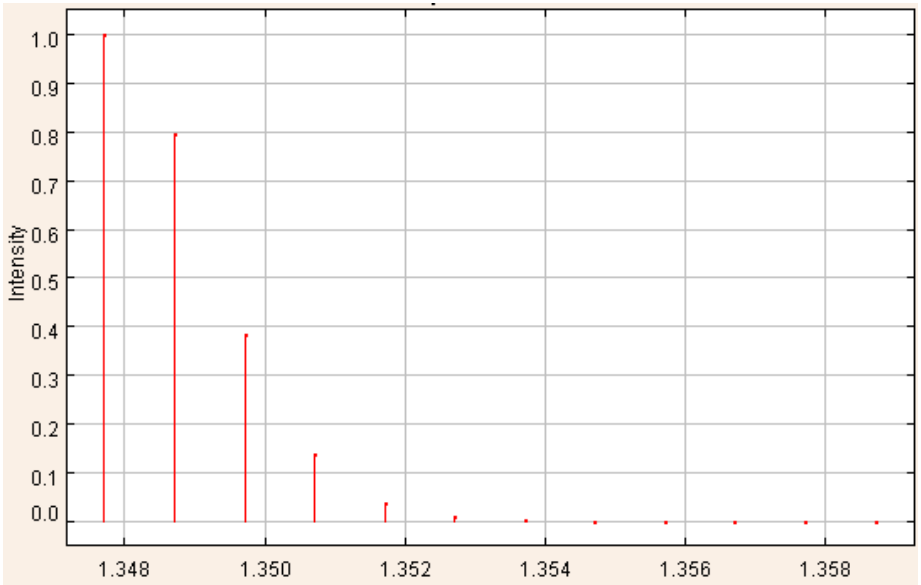


Today's Agenda

1. What is our motivation?
2. Statistical Modelling
3. Generating DAGF
4. What remains to be done?

Multinomial Model

- What is being explained?
  - Distributions of masses
  - Deviations from monoisotopic peaks



Multinomial Model

- Modelling isotope distributions
  - $\mathcal{P}(^{16}\mathrm{O}) = 99.757$
  - $\mathcal{P}(^{17}\mathrm{O}) = 0.038$
  - $\mathcal{P}(^{18}\mathrm{O}) = 0.205$
- Molecule = $\mathrm{C_cH_hO_oN_nS_s}$
- Assumptions
  - Isotope variant of a single atom from $\mathrm{C_cH_hO_oN_nS_s}$(e.g. C) independent of isotope variants of other atoms

    i.e. for a molecule with 2 atoms

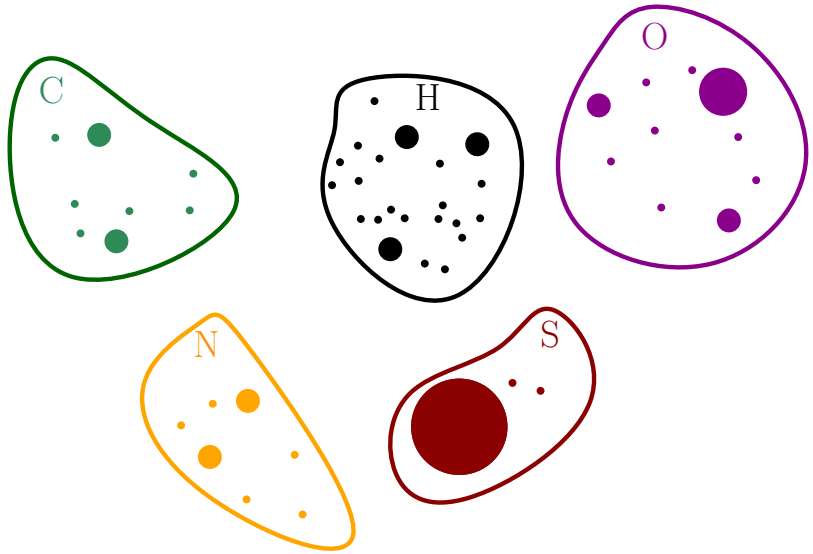
    $$\mathcal{P}(^{13}\mathrm{C}^{17}\mathrm{O}) = \mathcal{P}(^{13}\mathrm{C}) \times \mathcal{P}(^{17}\mathrm{O})$$

  - We cannot discern among isomers

    $$^{13}\mathrm{C}^{17}\mathrm{O}^{17}\mathrm{O} \simeq {}^{17}\mathrm{O}^{13}\mathrm{C}^{17}\mathrm{O} \simeq {}^{17}\mathrm{O}^{17}\mathrm{O}^{13}\mathrm{C}$$

Multinomial Model

- Chemical compound = list of sets of atoms



$$\mathcal{P}(\underbrace{c_{12}\,{}^{12}C,\ c_{13}\,{}^{13}C}_{c_{12}+c_{13}=c},\ \underbrace{h_1\,{}^{1}H,\ h_2\,{}^{2}H}_{h_1+h_2=h},\ \ldots,\ s_{36}\,{}^{36}S) =$$

$$\binom{c}{c_{12}, c_{13}} \mathcal{P}({}^{12}C)^{c_{12}} \mathcal{P}({}^{13}C)^{c_{13}} \binom{h}{h_1, h_2} \mathcal{P}({}^{1}H)^{h_1} \mathcal{P}({}^{2}H)^{h_2} \ldots$$

Multinomial Model and Molecular Mass

- Atomic mass m of a molecule $R = (c_{12}, c_{13}, \ldots s_{36})$
  - ⋆ $m_R = c_{12} m_{^{12}C} + \cdots + s_{36} m_{^{36}S}$
  - ⋆ $\mathcal{P}(m_R = m) = \sum \mathcal{P}(c_{12}\,{}^{12}C, \ldots, s_{36}\,{}^{36}S)$

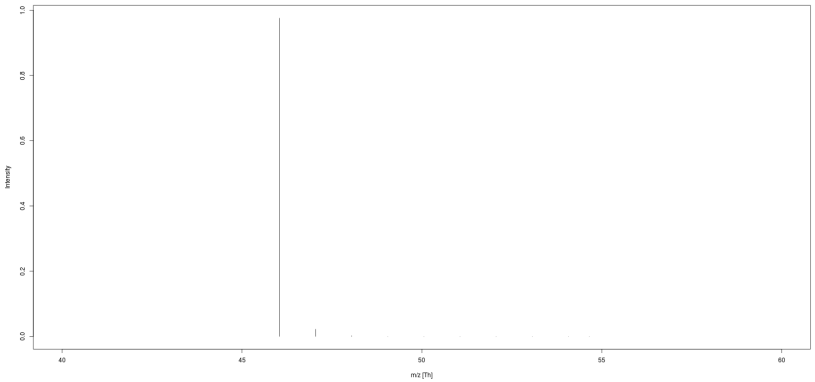
  s.t. $m_R = c_{12} m_{^{12}C} + \cdots + s_{36} m_{^{36}S}$
  - Good news: theory operates on chemical formulas
    - given a formula F, we calculate it's mass probability.
    - No need to solve these equations!
- Cool thing
  - Masses of neutrons for different elements are not equal!
  - Modern mass specs can already discern them

  ★ To some extent - we neglect that for now

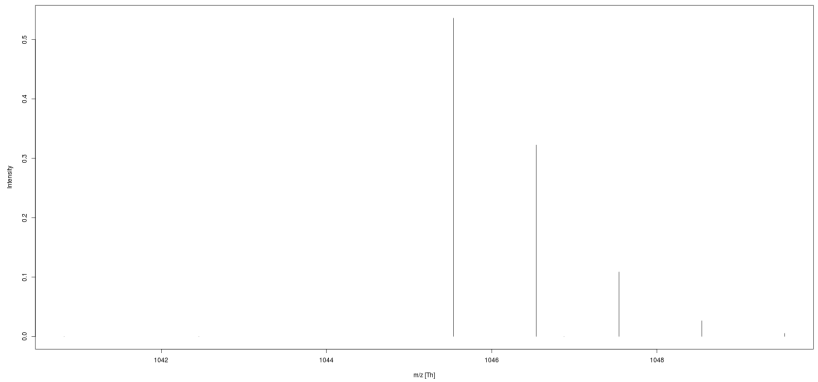
Visualising Multinomial Model

- BRAIN software - Piotr Dittwald®

Ethanol C_2H_5OH

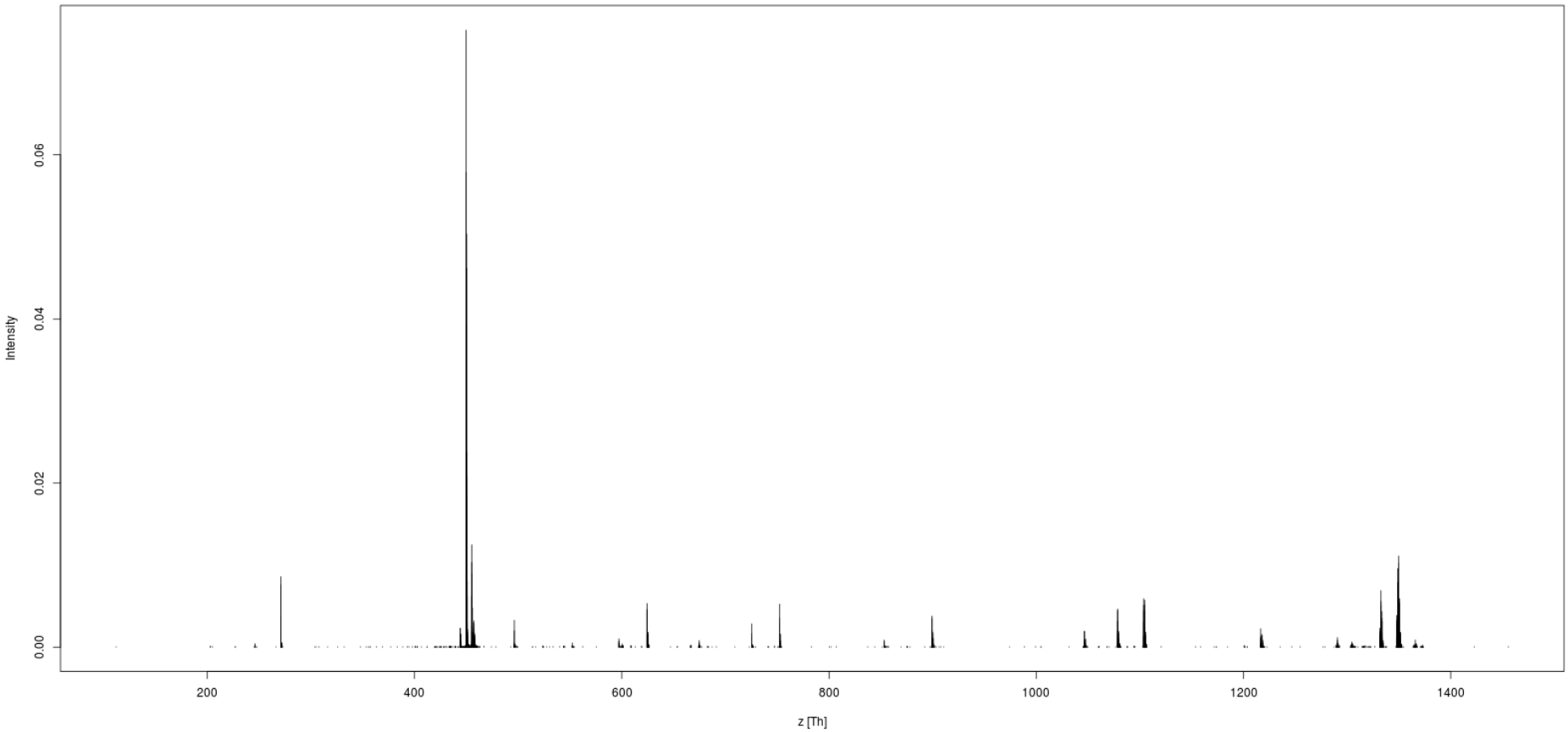


Angiotensine II $C_{50}H_{71}N_{13}O_{12}$

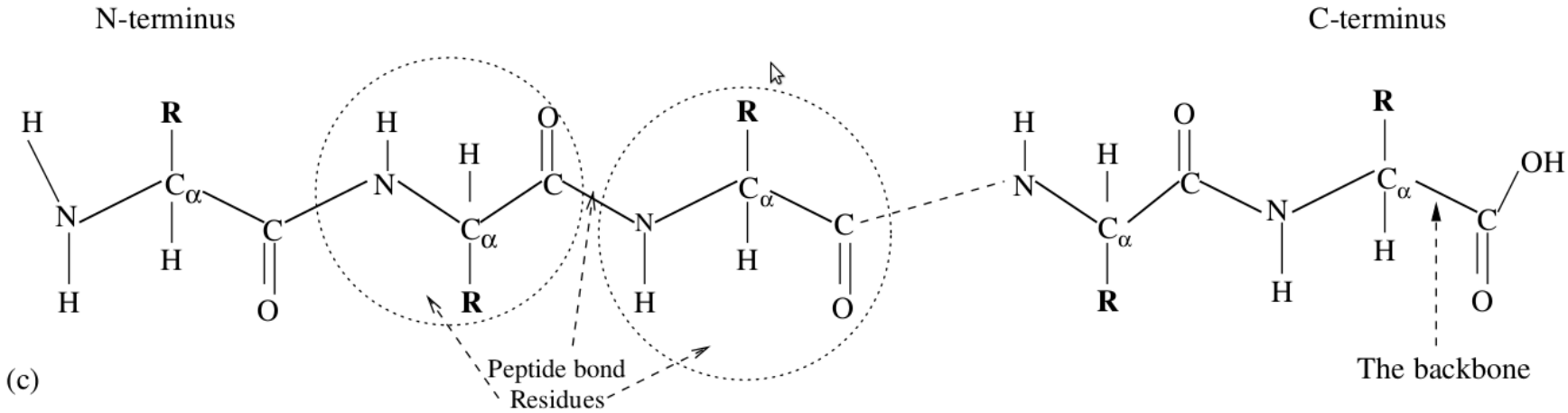


- Alas! Real spectra are multimodial!
  - fragmentation? (ETD)
  - charge reductions? (ETD, ETnoD, PTR)

Mass Spec results for substance P



Polymer as a sequence of Amino Acids

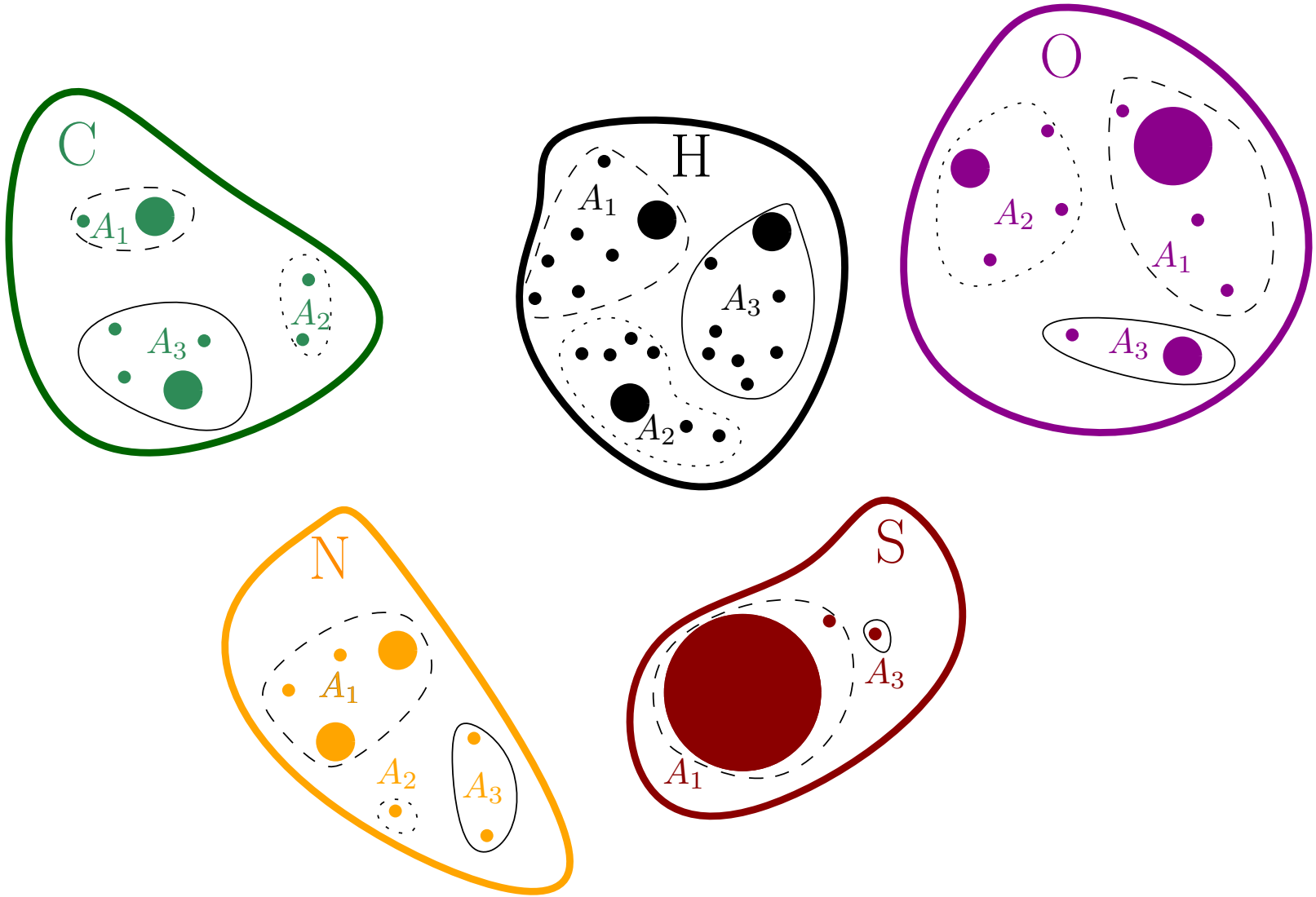


- Extra structure in our model must be added

  $C_cH_hO_oN_nS_s = A_1 A_2 \ldots A_k$

  $A_i \in \{$ Alanine, Cysteine, Aspartic Acid, Glutamic Acid, ...$\}$

Molecule Subdivided into Amino Acids



Electron Transfer Disociation

- Result: random cleavage of the peptide in two subsequences

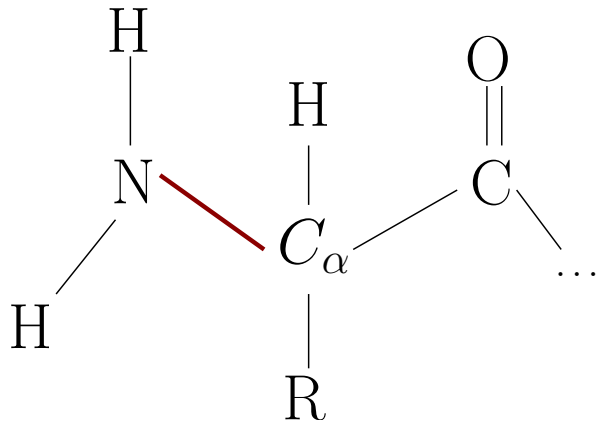
$$A_1 A_2 \ldots A_k \rightarrow \underbrace{A_1 \ldots A_L}_{C \text{ fragment}} \oplus \underbrace{A_{L+1} \ldots A_k}_{Z \text{ fragment}}$$

- L = index of the cleaved peptide bond (N-terminus to C-terminus)

- Assumption
  - Cleavage independent of isotope composition

$$\mathcal{P}(A_1 \ldots A_L = a_1 \ldots a_L | L = l) =$$

$$\mathcal{P}(A_1 \ldots A_l = a_1 \ldots a_l)$$



- Minor complication
  - Cleavage solely on A_1

Reactions considered by Frederik, our fellow chemist

Hypothesis:

- ⋆ Empirical spectrum = Result of Several Reactions

- ETD

$$[M+nH]^{n+} + A^{\cdot -} \longrightarrow [C+xH]^{x+} + [Z+(n-x)H]^{(n-x-1)\cdot} + A$$

- PTR

$$[M+nH]^{n+} + A^{\cdot -} \longrightarrow [M+(n-1)H]^{(n-1)+} + AH$$

- ETnoD

$$[M+nH]^{n+} + A^{\cdot -} \longrightarrow [M+nH]^{(n-1)+\cdot} + A$$

- Not good description of rules

  e.g. Concatenating reactions: $ETD \rightarrow PTR \rightarrow PTR \rightarrow ETnoD$

Correct Rules

- Additional Description of $C_cH_hO_oN_nS_s$- (p, q)
  - p - protonisation
    - n^o of extra protons
    - charge state of a given molecule
    - adds weight to molecule
  - q - neutralised protonisation
    - n^o of extra protons paired with electrons
    - only adds weight
- Problem Input $(A_1A_2 \ldots A_k, p, q)$

Reactions revised: post-doc in accountancy

- Problem Input $(A_1 A_2 \dots A_k,\ p,\ q)$
- Some Partial Reactions
  - ♣ ETD
    $\rightarrow\ (A_1 \dots A_L,\ p_1,\ q_1)$
    $\rightarrow\ (A_{L+1} \dots A_k,\ p_2,\ q_2)$
    s.t. $p_1 + p_2 = p - 1$ and $q_1 + q_2 = q$ and $q_2 \geq 0$
  - ♢ PTR
    $\rightarrow\ (A_1 \dots A_k,\ p-1,\ q)$
  - ♡ ETnoD
    $\rightarrow\ (A_1 \dots A_k,\ p-1,\ q+1)$
  - ♠ *HTR*
    $\rightarrow\ (A_1 \dots A_L,\ p_1,\ q_1)$
    $\rightarrow\ (A_{L+1} \dots A_k,\ p_2,\ q_2)$
    s.t. $p_1 + p_2 = p$ and $q_1 + q_2 = q + 1$ and $q_2 \geq 1$

Algorithm?

- Inputs:
  - $S = \left[M = A_1 \dots A_k,\ p = \text{Maximal Charge},\ q = 0\right]$
  - Partial Reactions $=$

$$= \{\mathfrak{Id},\ \clubsuit^{C}_{L,p_1,q_1},\ \clubsuit^{Z}_{L,p_2,q_2},\ \diamondsuit,\ \heartsuit,\ \spadesuit^{C}_{L,\tilde{p}_1,\tilde{q}_1},\ \spadesuit^{Z}_{L,\tilde{p}_2,\tilde{q}_2}\}$$

  - Reactions $=$

$$= \{r_1 r_2 \dots r_k : r_i \text{ is a Partial Reactions and is OK}\}$$

e.g. $\clubsuit^{C}_{L,2,1}\heartsuit$, $\heartsuit\diamondsuit$ might be valid reactions if

- S has enough charges
- M long enough
- there were q reactions before ET resulting in proton neutralisation

  - Empirical Spectrum $y = \left\{\left(\frac{m_j}{z_j}, I_j\right)\right\}_{j=1}^{J}$

Precising the hypothesis

- Observations:
  - Each reaction is also a triplet $R = [F, p, q]$
  - $F \sim$ Multinomial Distribution
  - m_F - corresponding mass distribution

$$\frac{m_R}{z_R} = \frac{m_F + p + q}{p}$$

Hypothesis

$\star$ $y = \sum_{R \in \text{Reactions}} \alpha_R \frac{m_R}{z_R} + \text{Error}$

s.t. $\alpha_R \geq 0$ and $\sum_R \alpha_R \leq 1$

- Problem: need for software that
  - finds Reactions
  - estimates α_R so that error is smallest possible

Project massTodon

- Operating

Statics: Directed Acyclic Graph of Formulas

- What can be discerned?
- Certain Trails

Statics: Directed Acyclic Graph of Formulas

Whole lotta of things to do.

- Algorithmically
  - Optimise the DAG generation
    - Take into account numerous repetitions.
    - Example : PTR->ETnoD = ETnoD->PTR.
  - Calibrate fitting procedures: stick spectra more jazzy among chemical homies, and the hood does not give a shit about approximations.
  - Derive quick procedures for stick spectra generation: modify BRAIN.
- Experimentally
  - More substances to analyze
  - Mixtures of substances
- Pragmatically
  - What if the substance is not known?


Ingvar Eidhammer, Kristian Flikka, Lennart Martens, Svein-Ole Mikalsen, *Computational Methods for Mass Spectrometry Proteomics*. Wiley-Interscience, 2007.


Igor Kaltashov, Stephen J. Eyles *Mass Spectrometry in Biophysics: Conformation and Dynamics of Biomolecules*. Wiley-Interscience, 2005.


Prof. Gavin E. Reid *Mass Analyzers*. Lecture slides from the First International Mass Spectrometry School, 2013.