Statistical modeling in molecular medicine: proteomics

Anna Gambin

Institute of Informatics,
University of Warsaw

outline

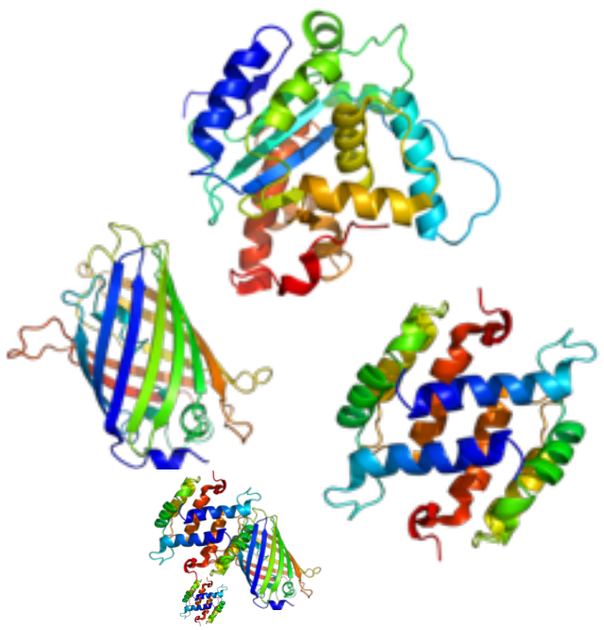
- **masSpec basics**
- **modeling isotopic distribution**

- **modeling exopeptidase activity**
- incorporating MEROPS data
- peptidase activity in time

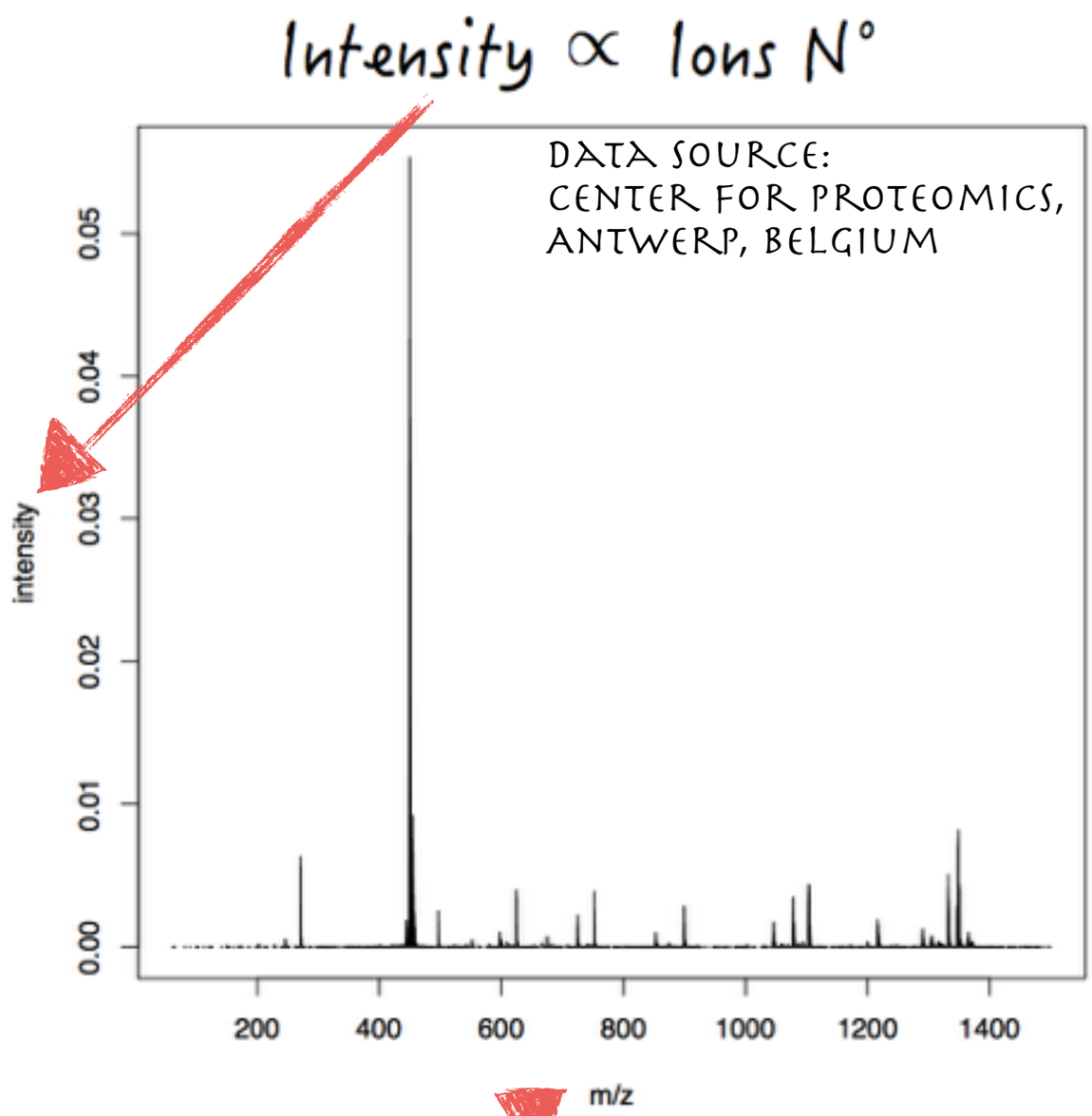
- **modeling electron transfer dissociation**
- deconvolution of spectra
- modeling fragmentation

Mass Spectrometry

Proteins

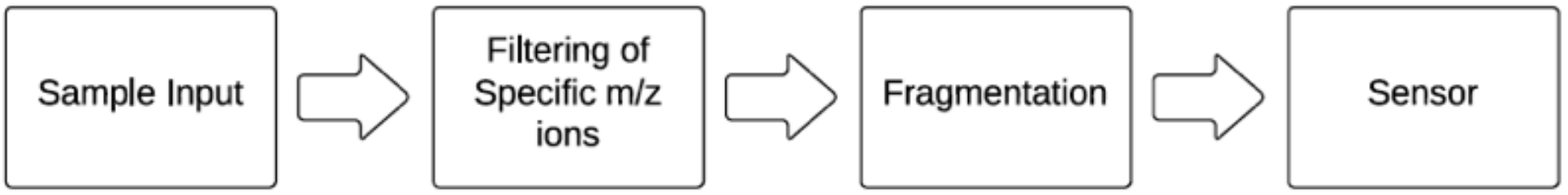


Intensity ∝ Ions N°



Mass/charge [Th]

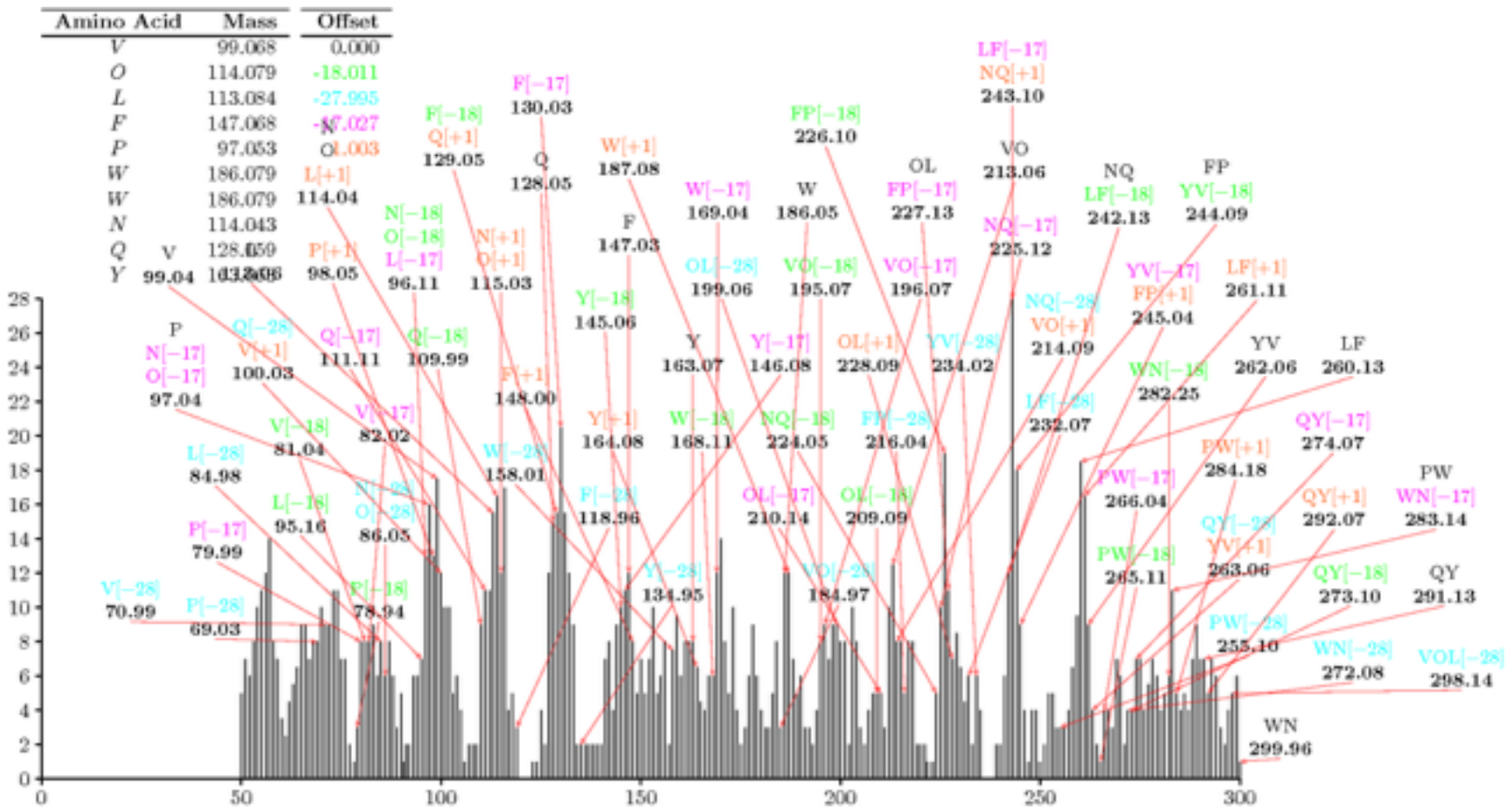
What is a Mass Spectrometer?



- A balance comprising
  - a dust-cleaner
  - a sieve
  - a knife
  - and a tracking system.

- The Mass Spectrometer
  - ionises the sample
  - manipulates the electrostatic field to move ions
  - fragments the ions
  - registers the ions

Identifying proteins is complicated



- there are plenty of proteins in a sample
- proteins are frequently fragmented
- even a single protein has a complicated signal

Chemical compounds are made of different isotopes

13.0033 - 12 = 1.0033 [Da]

differences in frequencies of observation

Element	Isotope	Extra Neutrons	Mass [Da]	Probability
Carbon	^{12}C	0	12	0.9893
	^{13}C	1	13.0033	0.0107
Hydrogen	^{1}H	0	1.0078	0.999885
	^{2}H	1	2.0141	0.000115
Nitrogen	^{14}N	0	14.0031	0.99632
	^{15}N	1	15.0001	0.00368
Oxygen	^{16}O	0	15.9949	0.99757
	^{17}O	1	16.9991	0.00038
	^{18}O	2	17.9992	0.00205
Sulfur	^{32}S	0	31.9721	0.9493
	^{33}S	1	32.9714	0.0076
	^{34}S	2	33.9679	0.0429
	^{36}S	4	35.9671	0.0002

2

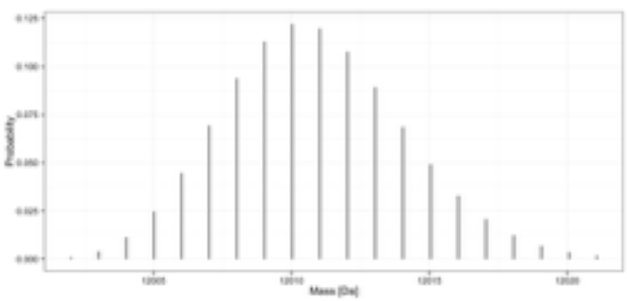
elements have different numbers of stable isotopes

3

4

isotopes of different elements differ in mass differences

32.9714 - 31.9721 = 0,9993 [Da]



isotopic envelope

huge number of isotopologues

A chemical formula

$$\mathrm{C}_c\mathrm{H}_h\mathrm{N}_n\mathrm{O}_o\mathrm{S}_s$$

can be composed out of isotopes

in

$$\prod_{e\in\{\mathrm{C,H,N,O,S}\}} \binom{n_e + i_e - 1}{n_e} \approx \prod_{e\in\{\mathrm{C,H,N,O,S}\}} \frac{e^{i_e-1}}{\sqrt{2\pi(i_e-1)}}\left(\frac{n_e}{i_e-1}+1\right)^{i_e-1}$$

$$^{12}\mathrm{C}_{c_0}\ ^{13}\mathrm{C}_{c_1}\ ^{1}\mathrm{H}_{h_0}\ ^{2}\mathrm{H}_{h_1}\ ^{14}\mathrm{N}_{n_0}\ ^{15}\mathrm{N}_{n_1}\ ^{16}\mathrm{O}_{o_0}\ ^{17}\mathrm{O}_{o_1}\ ^{18}\mathrm{O}_{o_2}\ ^{32}\mathrm{S}_{s_0}\ ^{33}\mathrm{S}_{s_1}\ ^{34}\mathrm{S}_{s_2}\ ^{36}\mathrm{S}_{s_4}$$

ways, where

- n_e - N° of atoms of element e
- i_e - N° of isotopes of element e

important observation

some **isotopic variants** are more **probable** than others

$$P(^{12}C_{c_0}\ ^{13}C_{c_1}\ ^{1}H_{h_0}\ ^{2}H_{h_1}\ ^{14}N_{n_0}\ ^{15}N_{n_1}\ ^{16}O_{o_0}\ ^{17}O_{o_1}\ ^{18}O_{o_2}\ ^{32}S_{s_0}\ ^{33}S_{s_1}\ ^{34}S_{s_2}\ ^{36}S_{s_4}) = ?$$

Assume

1) variants of isotopes of atoms are independent

2) elements vary in abundances of isotopes

$$P\left({}^{12}C_{c_0}\ {}^{13}C_{c_1}\ {}^{1}H_{h_0}\ {}^{2}H_{h_1}\ {}^{14}N_{n_0}\ {}^{15}N_{n_1}\ {}^{16}O_{o_0}\ {}^{17}O_{o_1}\ {}^{18}O_{o_2}\ {}^{32}S_{s_0}\ {}^{33}S_{s_1}\ {}^{34}S_{s_2}\ {}^{36}S_{s_4}\right) =$$

$$\binom{c}{c_0, c_1} \mathcal{P}({}^{12}C)^{c_0} \mathcal{P}({}^{13}C)^{c_1} \binom{h}{h_0, h_1} \mathcal{P}({}^{1}H)^{h_0} \mathcal{P}({}^{2}H)^{h_1} \binom{n}{n_0, n_1} \mathcal{P}({}^{14}N)^{n_0} \mathcal{P}({}^{15}N)^{n_1} \times$$

$$\binom{n}{o_0, o_1, o_2} \mathcal{P}({}^{16}O)^{o_0} \mathcal{P}({}^{17}O)^{o_1} \mathcal{P}({}^{18}O)^{o_2} \binom{s}{s_0, s_1, s_2, s_4} \mathcal{P}({}^{32}S)^{s_0} \mathcal{P}({}^{33}S)^{s_1} \mathcal{P}({}^{34}S)^{s_2} \mathcal{P}({}^{36}S)^{s_4}$$

frequencies of isotopes

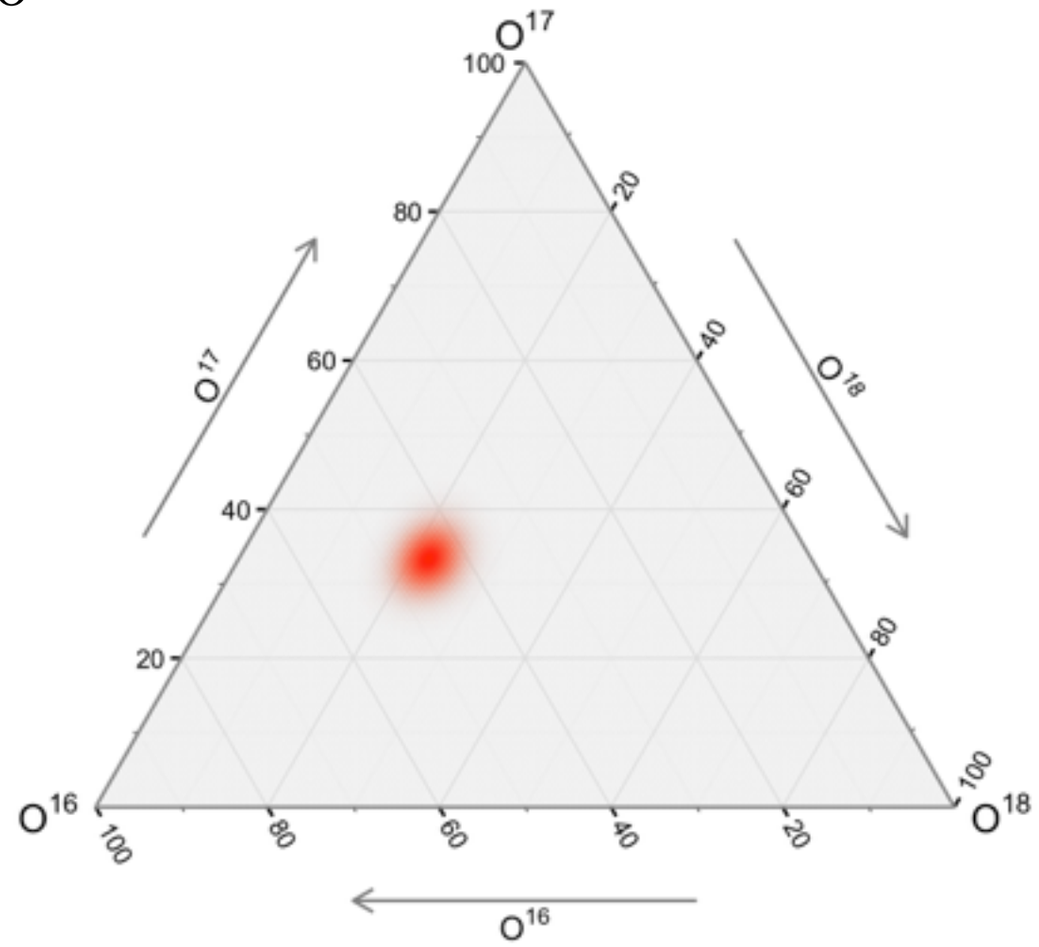
IUPAC
INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

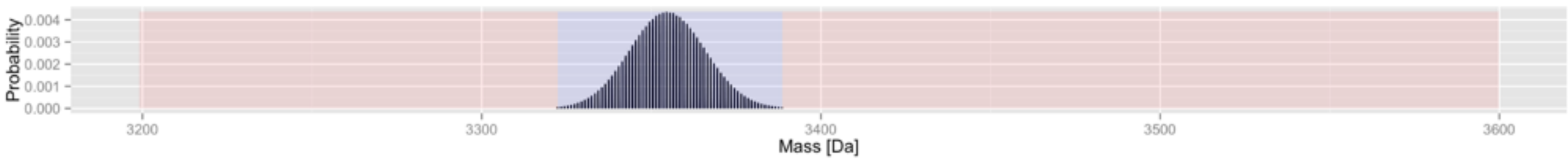
$$\mathcal{P}(^{16}\mathrm{O}) = \frac{4}{9} \quad \mathcal{P}(^{17}\mathrm{O}) = \frac{3}{9} \quad \mathcal{P}(^{18}\mathrm{O}) = \frac{2}{9}$$ (not real world values!)

200 oxygen atoms

$$o_0 + o_1 + o_2 = 200$$

→ 20301 variants → whereas 1043 bear 99% prob.

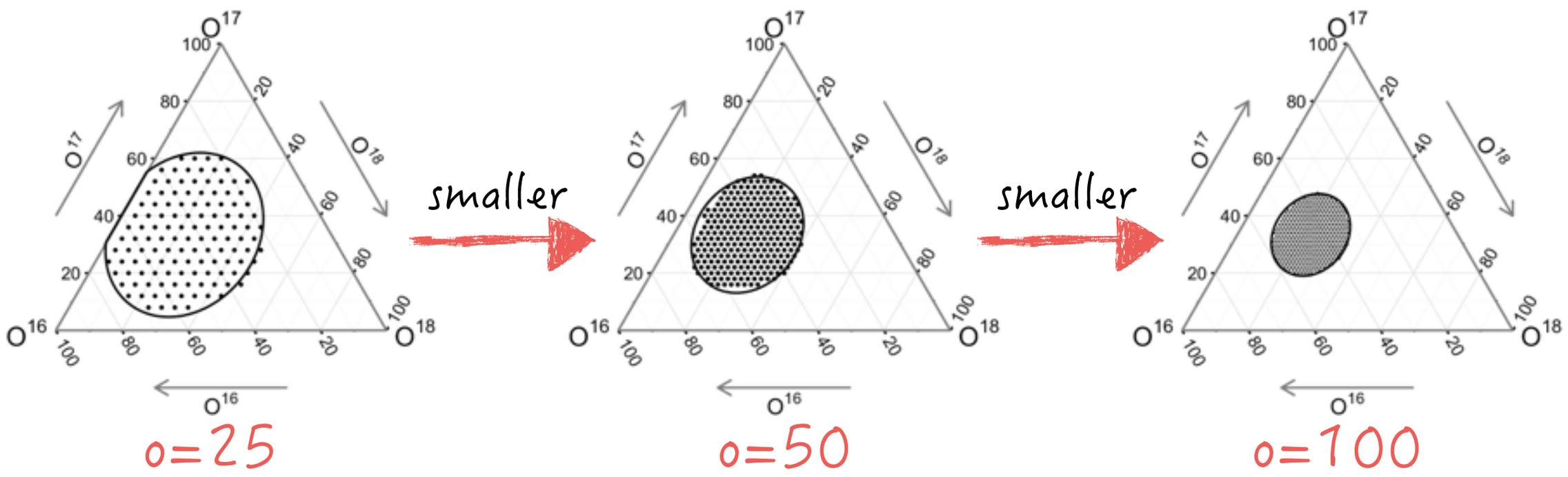


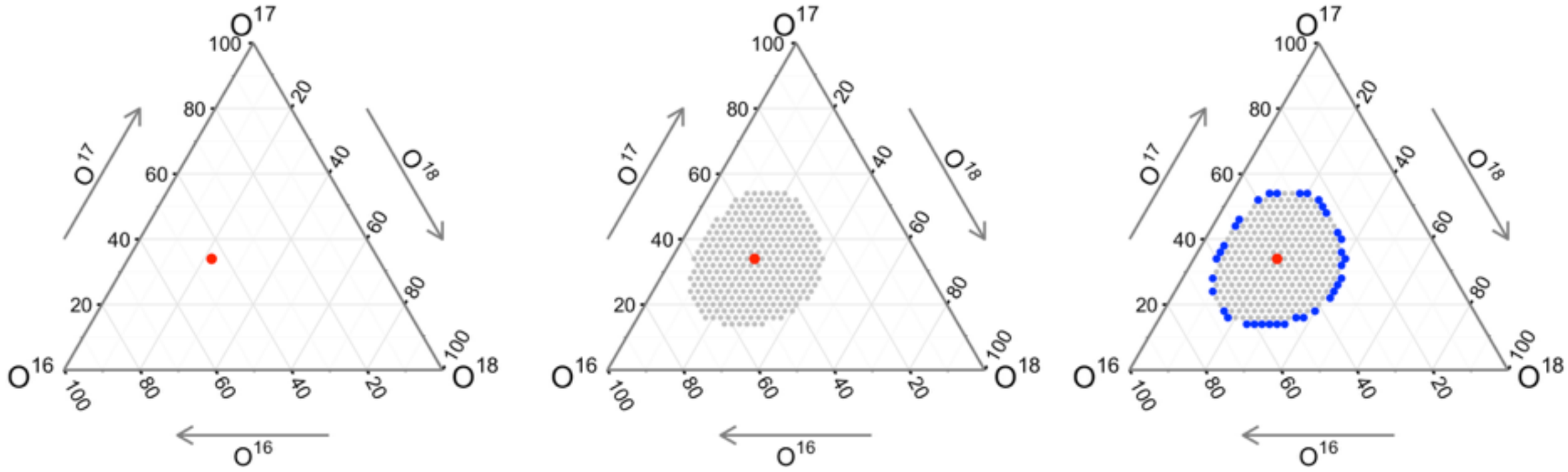


How much we gain by considering the smallest set with a fixed probability ?

$$\#\left(\begin{matrix}\text{PROBABLE}\\ \text{VARIANTS}\end{matrix}\right) \approx C_{\text{lattice}}\Big(\prod_{\text{Elements}} n_e^{\frac{i_e-1}{2}}\sqrt{\det\Delta_e}\Big) q^k_{\chi^2(k)}\frac{\pi^{k/2}}{\Gamma(k/2+1)} \propto$$

$$\prod_{\text{Elements}} n_e^{\frac{i_e-1}{2}} \quad \text{VS} \quad \prod_{\text{Elements}} n_e^{i_e-1}$$





To get the smallest set with probability P:

Find the **most probable variant**

while **Total Probability** < P :

Get **layer** so that p> P(v)>=qp where

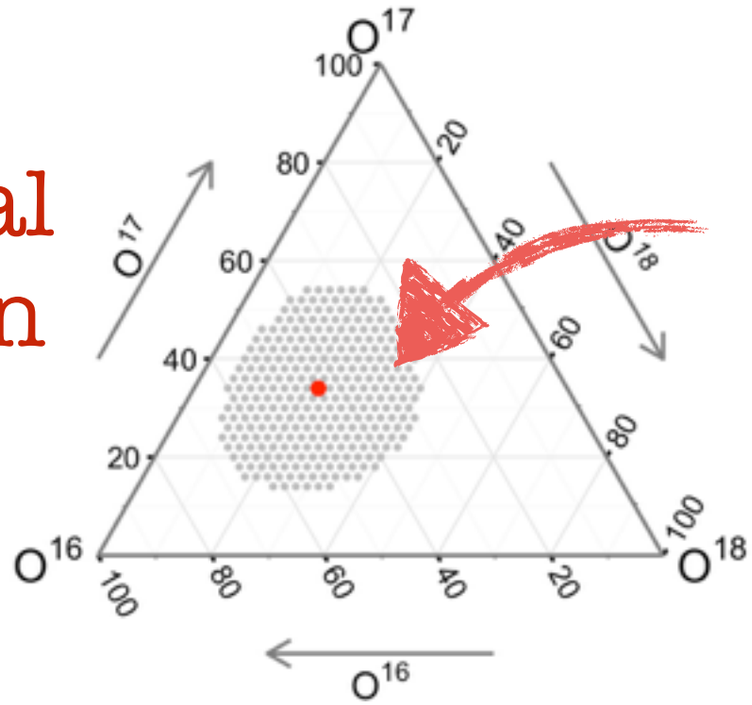
$p = P(v_{\text{min previous layer}})$

Trim the **least probable variants** from the last layer so that **Total Probability** >= P

Monotonic Expansion Property:

For each v set {W: P(W)>=P(v) } is adjacent to v

multinomial distribution

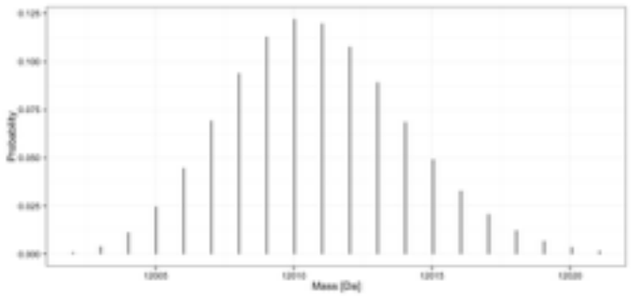


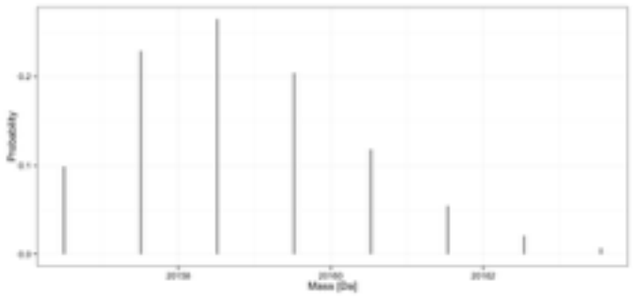
Smallest set with current **Total Probability**

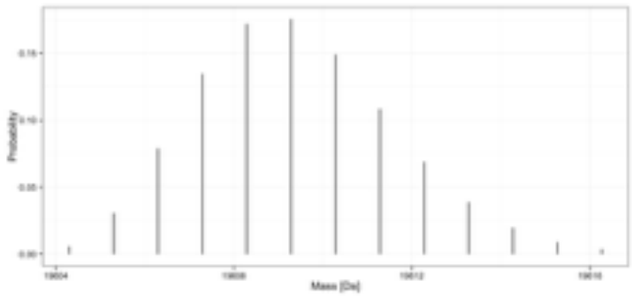
$$\binom{c}{c_0, c_1}\mathcal{P}(^{12}C)^{c_0}\mathcal{P}(^{13}C)^{c_1}\binom{h}{h_0, h_1}\mathcal{P}(^{1}H)^{h_0}\mathcal{P}(^{2}H)^{h_1}\binom{n}{n_0, n_1}\mathcal{P}(^{14}N)^{n_0}\mathcal{P}(^{15}N)^{n_1}$$

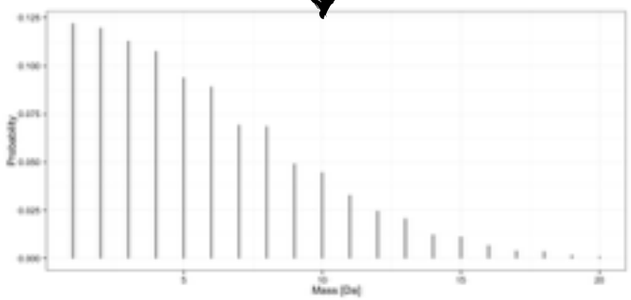
$$\binom{n}{o_0, o_1, o_2}\mathcal{P}(^{16}O)^{o_0}\mathcal{P}(^{17}O)^{o_1}\mathcal{P}(^{18}O)^{o_2}\binom{s}{s_0, s_1, s_2, s_4}\mathcal{P}(^{32}S)^{s_0}\mathcal{P}(^{33}S)^{s_1}\mathcal{P}(^{34}S)^{s_2}\mathcal{P}(^{36}S)^{s_4}$$

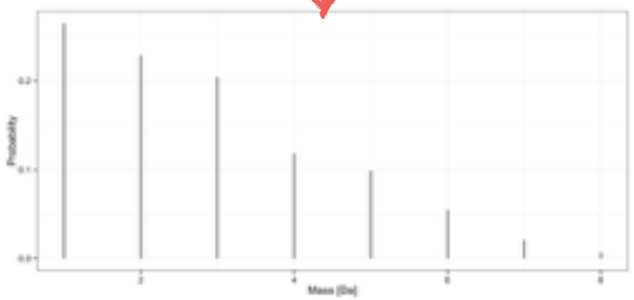
$$\binom{c}{c_0, c_1}\mathcal{P}(^{12}C)^{c_0}\mathcal{P}(^{13}C)^{c_1}\binom{h}{h_0, h_1}\mathcal{P}(^{1}H)^{h_0}\mathcal{P}(^{2}H)^{h_1}\binom{n}{n_0, n_1}\mathcal{P}(^{14}N)^{n_0}\mathcal{P}(^{15}N)^{n_1}$$

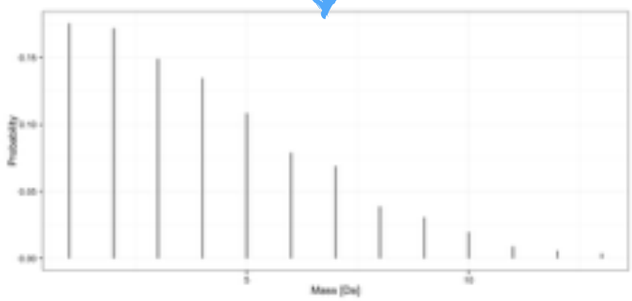












$(x_k\ y_l\ z_m)$

$(x_{k+1}\ y_l\ z_m)$

$(x_k\ y_{l+1}\ z_m)$

$(x_k\ y_l\ z_{m+1})$

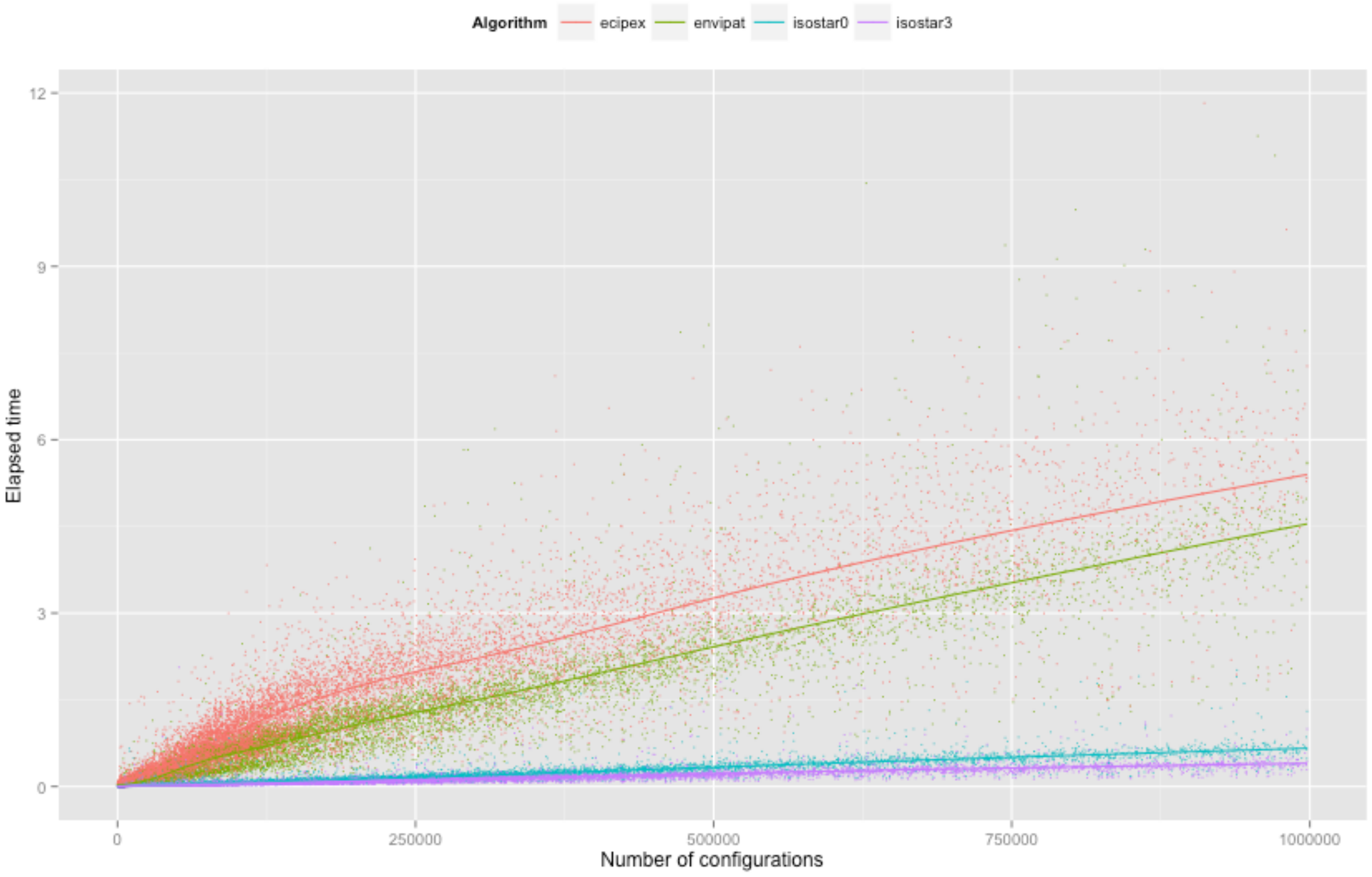
our **OPTIMAL** implementation uses

complexity

- queue for storing subsequent layers — $O(n)$
- a version of quick select for trimming — $O(n)$
- other tricks — $O(n)$

O(n) in the total number of configurations

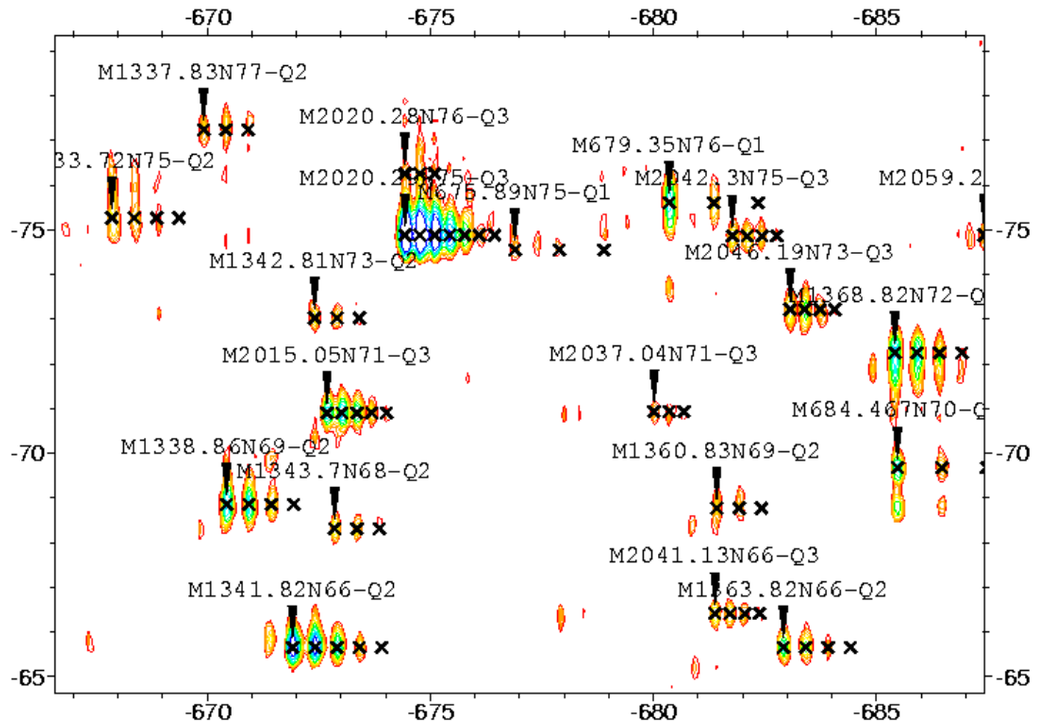
We provide **theoretical background**
and get **better run times**

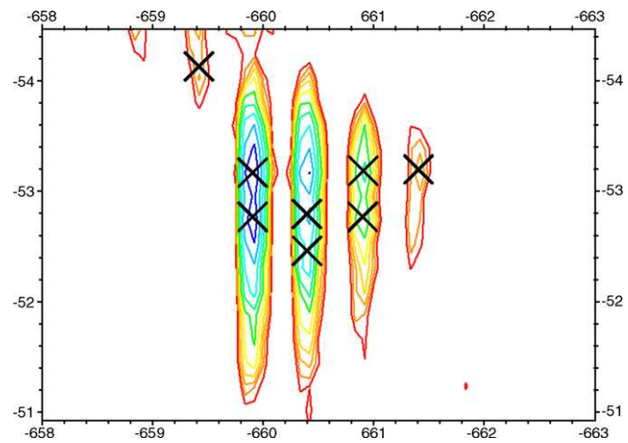


proteolytic fragmentation

LC-MS/MS

- data for colorectal cancer patients and healthy donors
- ca 1000 peptides
- **preprocessing**: spectra interpretation and retention time aligning





ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

International Journal of Mass Spectrometry 260 (2007) 20–30

Mass Spectrometry

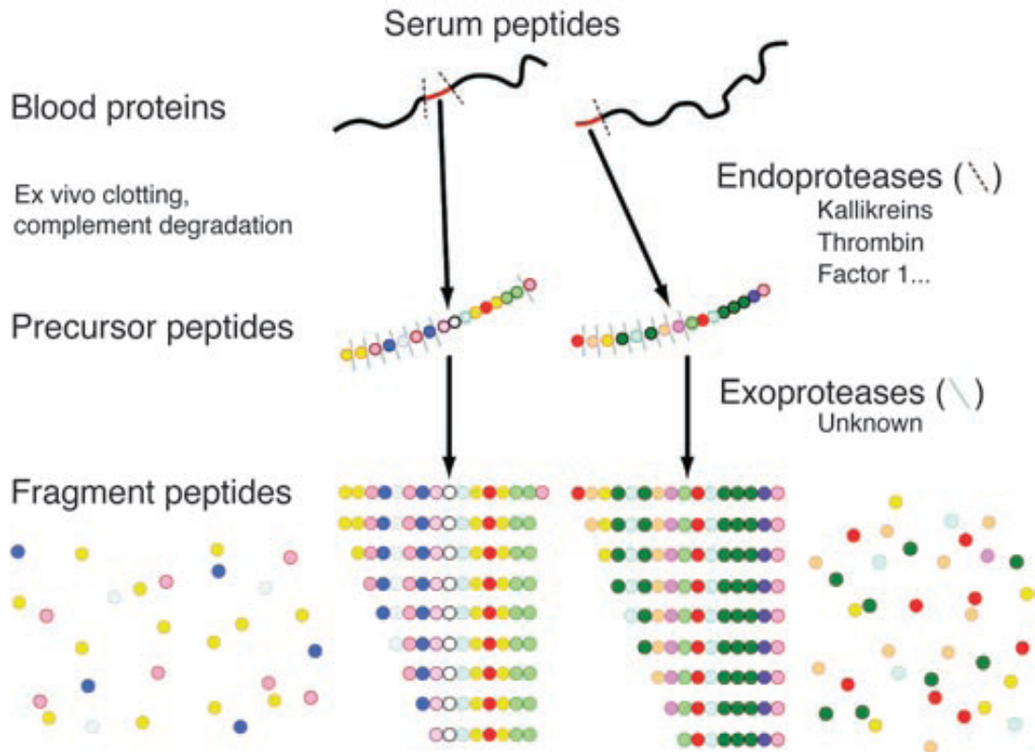
www.elsevier.com/locate/ijms

Automated reduction and interpretation of multidimensional mass spectra for analysis of complex peptide mixtures

Anna Gambin[a*], Janusz Dutkowski[a], Jakub Karczmarski[b], Bogusław Kluge[a], Krzysztof Kowalczyk[a], Jerzy Ostrowski[b], Jarosław Poznański[c], Jerzy Tiuryn[a], Magda Bakun[c], Michał Dadlez[c,d]

Exopeptidase activity

- **motivation**: **differential exoprotease activities contribute to cancer type–specific serum peptidome degradation**
- **our goal: first formal model estimated from LC-MS/MS data**

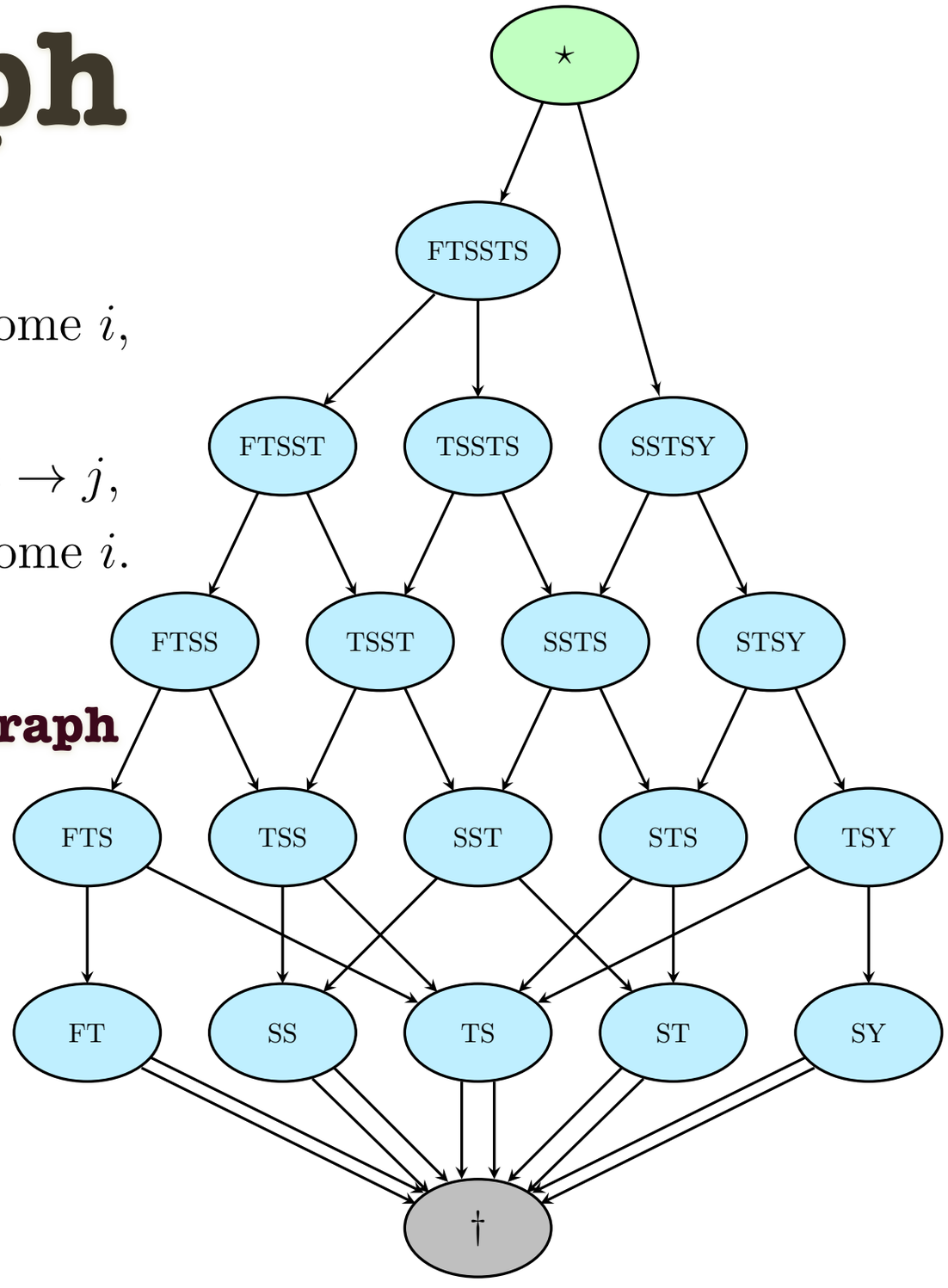


Villanueva, J., Nazarian, A., Lawlor, K., et al. 2008. A sequence-specific exopeptidase activity test (sseat) for "functional" biomarker discovery. *Mol. Cell. Proteomics* 7, 509–518.

Cleavage graph

$$Q(x,x') = \begin{cases} a_{\star i} & \text{if } x'_i = x_i + 1,\ x'_{-i} = x_{-i} \text{ for some } i, \\ a_{r(i,j)} x_i & \text{if } x'_j = x_j + 1,\ x'_i = x_i - 1, \\ & \quad \text{and } x'_{-i-j} = x_{-i-j} \text{ for some } i \to j, \\ a_{i\dagger} x_i & \text{if } x'_i = x_i - 1,\ x'_{-i} = x_{-i} \text{ for some } i. \end{cases}$$

transition intensities for Markov process describing the flow of particles through the graph i.e. the process of peptidome degradation



$$Q(x,x') = \begin{cases} a_{\star i} & \textbf{create} \\ a_{r(i,j)} x_i & \textbf{move} \\ a_{i\dagger} x_i & \textbf{annihilate/degrade} \end{cases}$$

in equilibrium

Proposition 1 (Equilibrium distribution). *The process* $(X(t))$ *has the equilibrium (stationary) distribution* π *given by:*

$$\pi(x) = \prod_{i \in \mathcal{V}} \mathrm{e}^{\lambda_i} \frac{\lambda_i^{x_i}}{x_i!},$$

where the configuration of intensities $(\lambda_i)_{i \in \mathcal{V}}$ *is the unique solution to the following system of "balance" equations:*

$$\sum_{k \to i} \lambda_k a_{r(k,i)} + a_{\star i} = \lambda_i \left(\sum_{i \to j} a_{r(i,j)} + a_{i\dagger} \right) \qquad \text{for every } i \in \mathcal{V}.$$

JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 16, Number 2, 2009

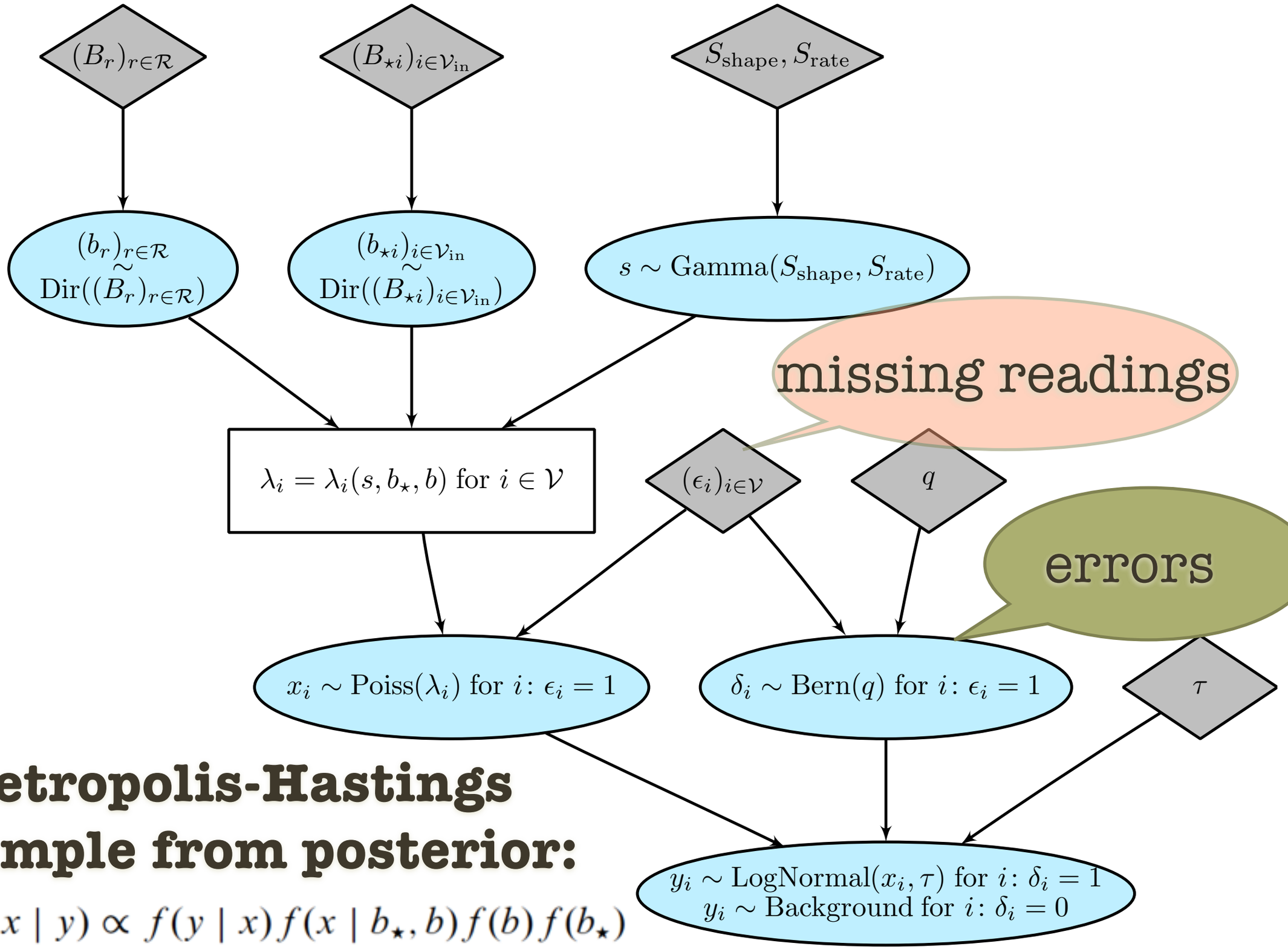
Pp. 395–406
DOI: 10.1089/cmb.2008.22TT

old as the hills, but...

Modeling Exopeptidase Activity from LC-MS Data

BOGUSŁAW KLUGE,[1] ANNA GAMBIN,[1] and WOJCIECH NIEMIRO[2]

hierarchical Bayesian model



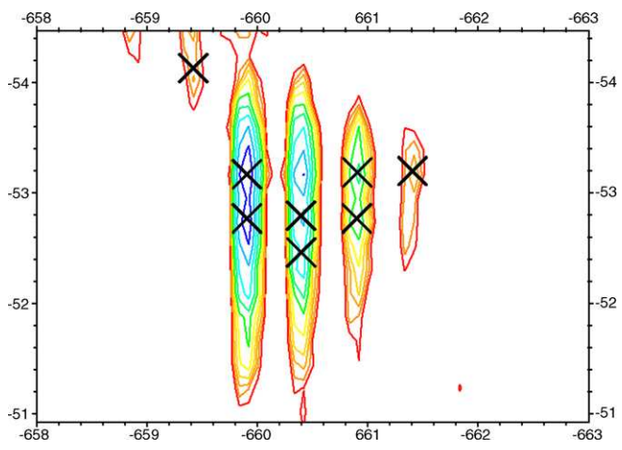
Metropolis-Hastings to sample from posterior:

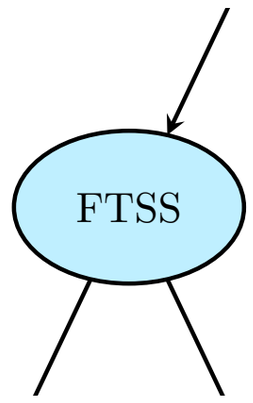
$$f(b_\star, b, x \mid y) \propto f(y \mid x)\, f(x \mid b_\star, b)\, f(b)\, f(b_\star)$$

NON TRIVIAL TASK: filling the cleavage graph with real data

- 1000 peptides: mass, charge, retention time
- 243 precursor peptides
- ca. 40 000 subsequences

- from aa sequence: calculate mass
- consider all charges
- predict retention time (random forests)

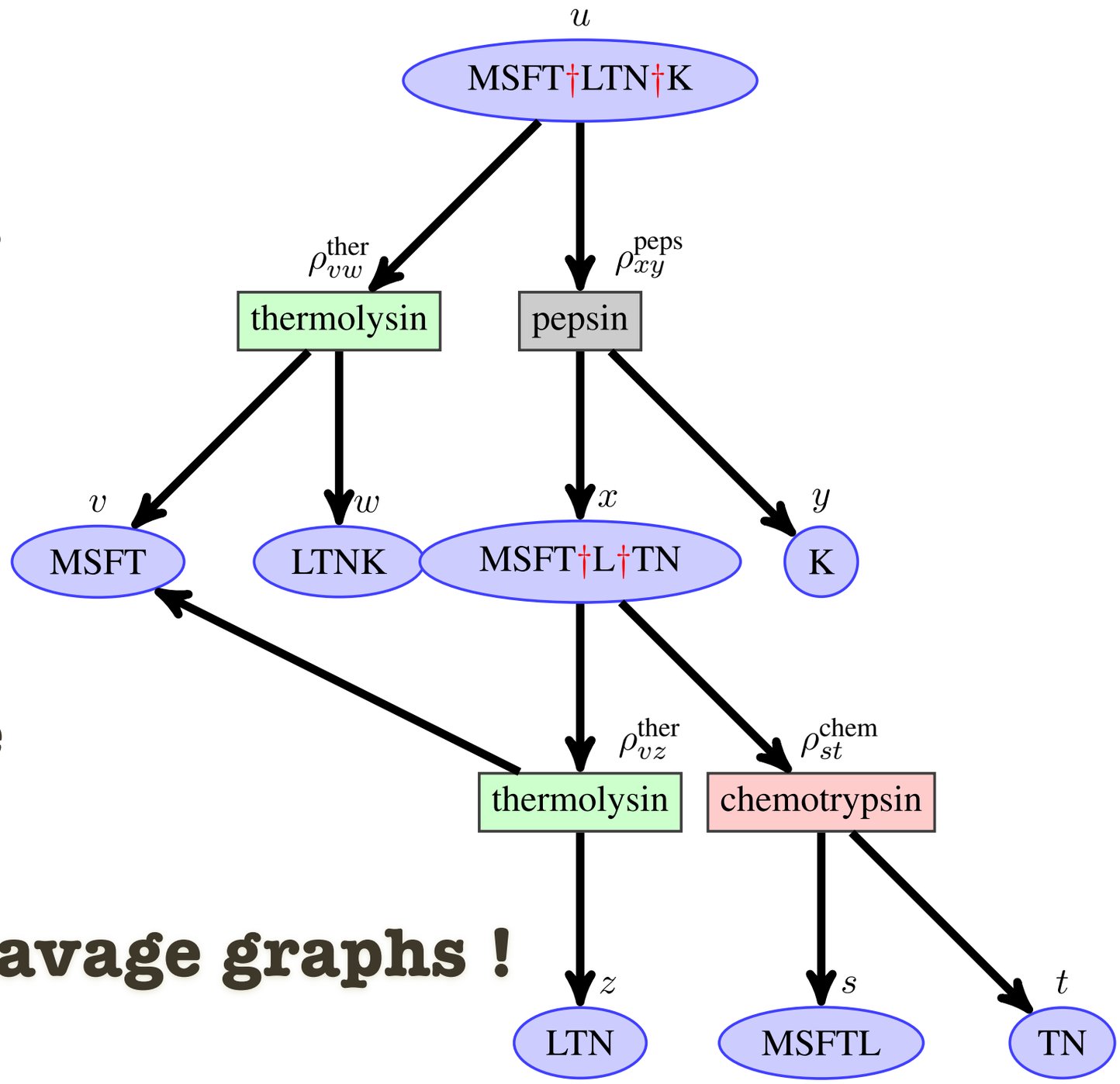




quite often: missing reads and errors !

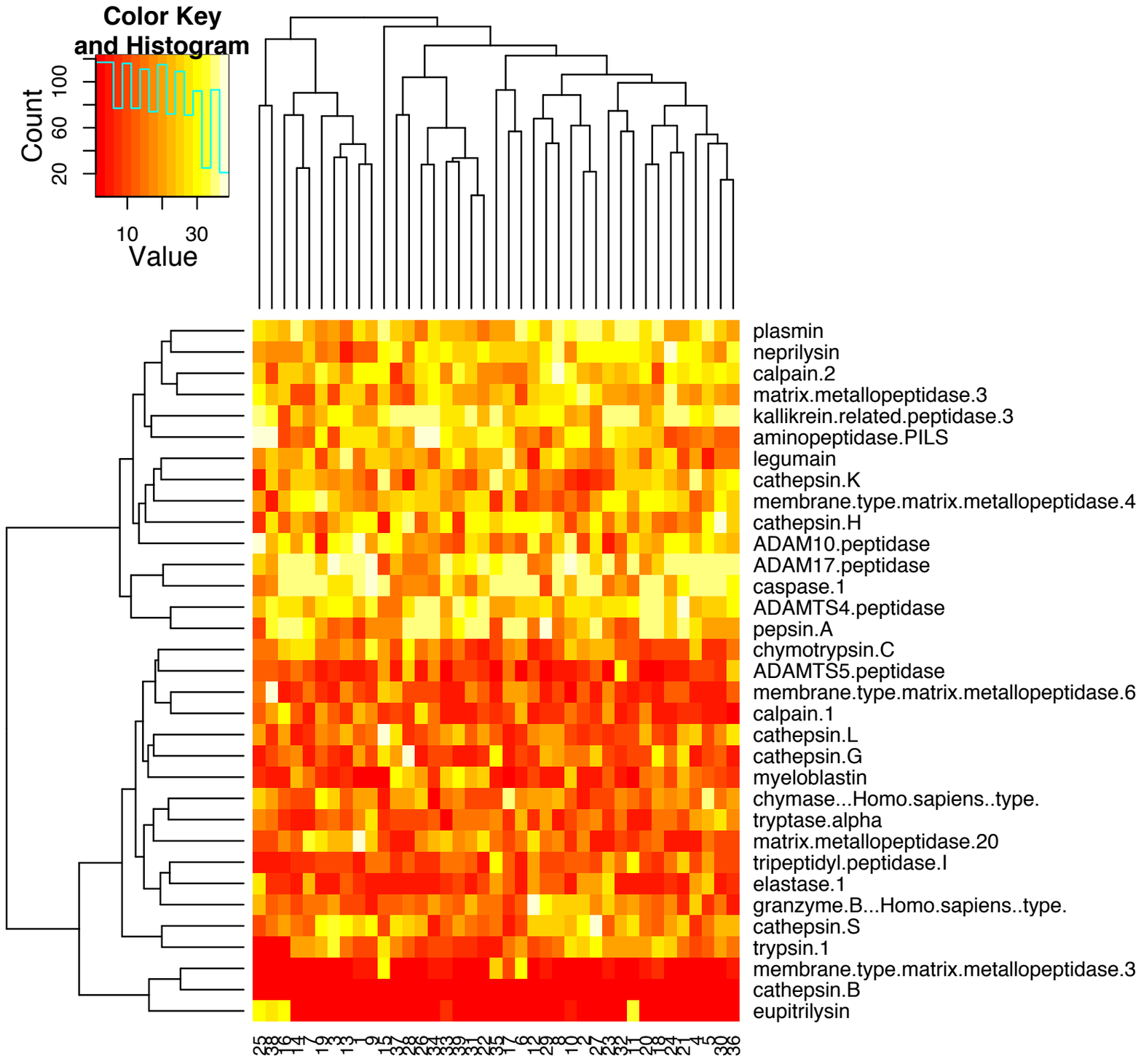
Cleavage graph for real proteolytic events

- 20 colorectal cancer patients and 20 healthy donors,
- ca 1000 peptides,
- preprocessing phase

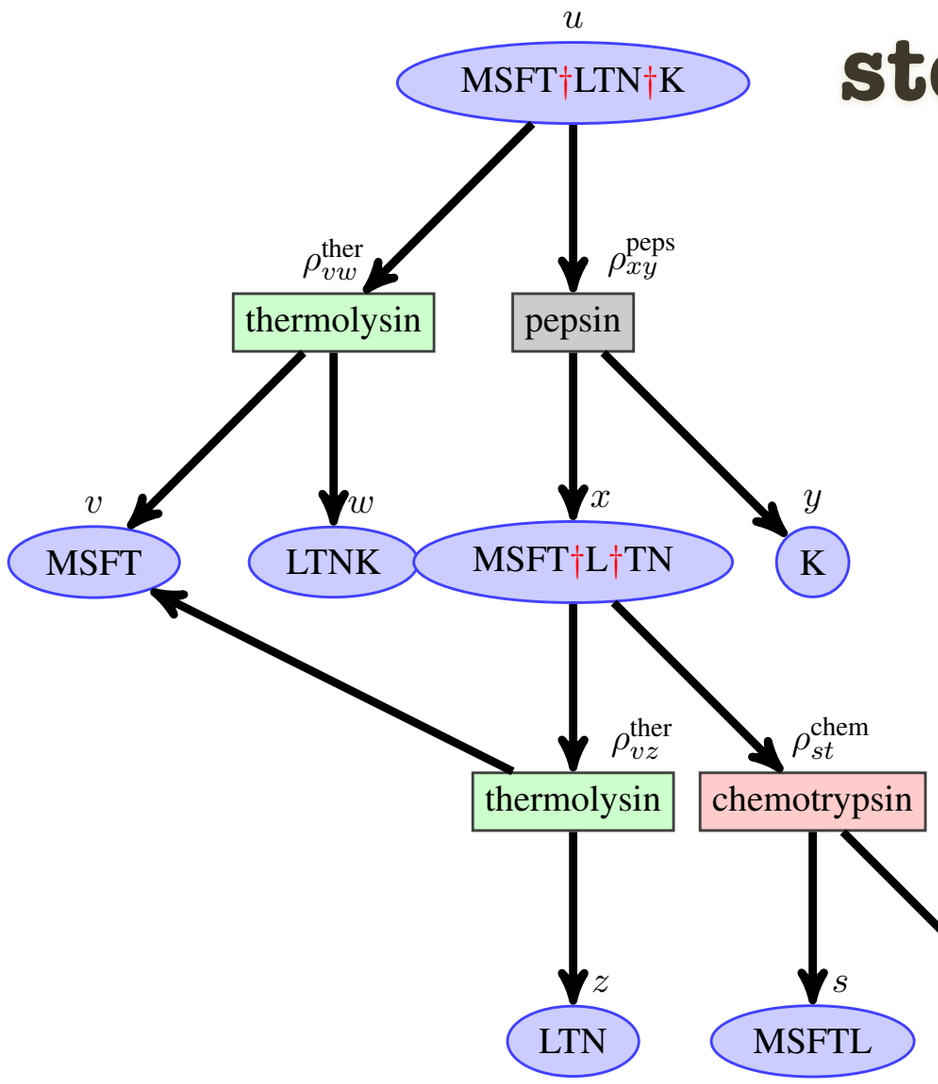


MUCH SMALLER cleavage graphs !

identified enzymes make sense !



stochastic dynamics in time



from MEROPS:

ρ_{vw} the vector of all peptidase affinity coefficients for the cleavage $v \dagger w$

$$Q_{xx'} = \begin{cases} c^{\mathrm{T}} \rho_{vw} x_u & \text{if } x' = x - \epsilon_u + \epsilon_v + \epsilon_w \text{ and } u = v \dagger w \,, \\ 0 & \text{otherwise.} \end{cases}$$

to be estimated:

peptidase cutting intensities vector

$$P(x,t) = \mathcal{P}(X(t) = x).$$

calculated from CME

no more monomolecular system -
we have reactions:
A -> B and A-> B+C (endopeptidases)

$$\begin{aligned} \frac{\partial}{\partial t} P(x,t) &= \sum_{y \neq x} (Q_{yx} P(y,t) - Q_{xy} P(x,t)) \\ &= \sum_{u=v\dagger w} c^{\mathrm{T}} \rho_{vw} \left[(x_u + 1) P(x + \epsilon_u - \epsilon_v - \epsilon_w, t) - x_u P(x,t) \right] \\ &= \sum_{u=v\dagger w} c^{\mathrm{T}} \rho_{vw} [x'_u P(x',t) - x_u P(x,t)], \end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial t}P(x,t) &= \sum_{y\neq x}\left(Q_{yx}P(y,t) - Q_{xy}P(x,t)\right)\\ &= \sum_{u=v\dagger w} c^{\mathrm{T}}\rho_{vw}\left[(x_u+1)P(x+\epsilon_u-\epsilon_v-\epsilon_w,t) - x_uP(x,t)\right]\\ &= \sum_{u=v\dagger w} c^{\mathrm{T}}\rho_{vw}[x'_uP(x',t) - x_uP(x,t)],\end{aligned}$$

where $x' = x + \epsilon_u - \epsilon_v - \epsilon_w$, i.e. x' denotes a configuration before the cleavage $v\dagger w$.

$$
\begin{aligned}
\mathrm{E}_q\left(t\right) &= \sum_x x_q P(x,t), \\
\frac{d}{dt}\mathrm{E}_q\left(t\right) &= \sum_x x_q \frac{\partial}{\partial t} P(x,t) \\
&= \sum_x x_q \sum_{u=v\dagger w} c^{\mathrm{T}} \rho_{vw} \left[(x_u+1)P(x+\epsilon_u-\epsilon_v-\epsilon_w,t) - x_u P(x,t)\right] \\
&= \sum_{u=v\dagger w} c^{\mathrm{T}} \rho_{vw} \left[\sum_x x_q (x_u+1)P(x+\epsilon_u-\epsilon_v-\epsilon_w,t) - \sum_x x_q x_u P(x,t)\right] \\
&= \sum_{u=v\dagger w} c^{\mathrm{T}} \rho_{vw} \left[\sum_x (x-\epsilon_u+\epsilon_v+\epsilon_w)_q x_u P(x,t) - \sum_x x_q x_u P(x,t)\right] \\
&= \sum_{u=v\dagger w} c^{\mathrm{T}} \rho_{vw} \sum_x (-\epsilon_u+\epsilon_v+\epsilon_w)_q x_u P(x,t).
\end{aligned}
$$

$$
\frac{d}{t}\mathrm{E}_q\left(t\right) = \sum_{u=q\dagger w} c^{\mathrm{T}} \rho_{qw} \mathrm{E}_u\left(t\right) + \sum_{u=v\dagger q} c^{\mathrm{T}} \rho_{vq} \mathrm{E}_u\left(t\right) - \sum_{q=v\dagger w} c^{\mathrm{T}} \rho_{vw} \mathrm{E}_q\left(t\right)
$$

Denote by λ_{uq} the intensity of creating q from u by a single cleavage of the form $u = q \dagger w$ or $u = v \dagger q$, i.e. $\lambda_{uq} = c^{\mathrm{T}}(\rho_{qw} + \rho_{vq})$. Let $\lambda_{qq} = -\sum_{q=v\dagger w} c^{\mathrm{T}}\rho_{vw}$, i.e. minus the intensity of consuming q in all cleavages involving this peptide. Note that the following equality holds:

$$\lambda_{qq} = -\frac{1}{2}\left[\sum_{q=v\dagger w} \lambda_{qv} + \sum_{q=z\dagger v} \lambda_{qv} - \sum_{\substack{q=v\dagger w \\ q=z\dagger v}} \lambda_{qv}\right] = -\frac{1}{2}\sum_{q\to v} \lambda_{qv}.$$

Now the equations (4.2) have the following form:

$$\left[\frac{d}{dt}\,\mathrm{E}_q\,(t) = \sum_{u \to q} \lambda_{uq}\,\mathrm{E}_u\,(t) + \lambda_{qq}\,\mathrm{E}_q\,(t)\right]_{q \in \mathcal{V}}. \tag{4.3}$$

The solution of the system of linear constant coefficient ordinary differential equations like (4.3) is given by:

$$\mathrm{E}\,(t) = \mathrm{E}\,(0)^{\mathrm{T}} \exp(\Lambda t), \tag{4.4}$$

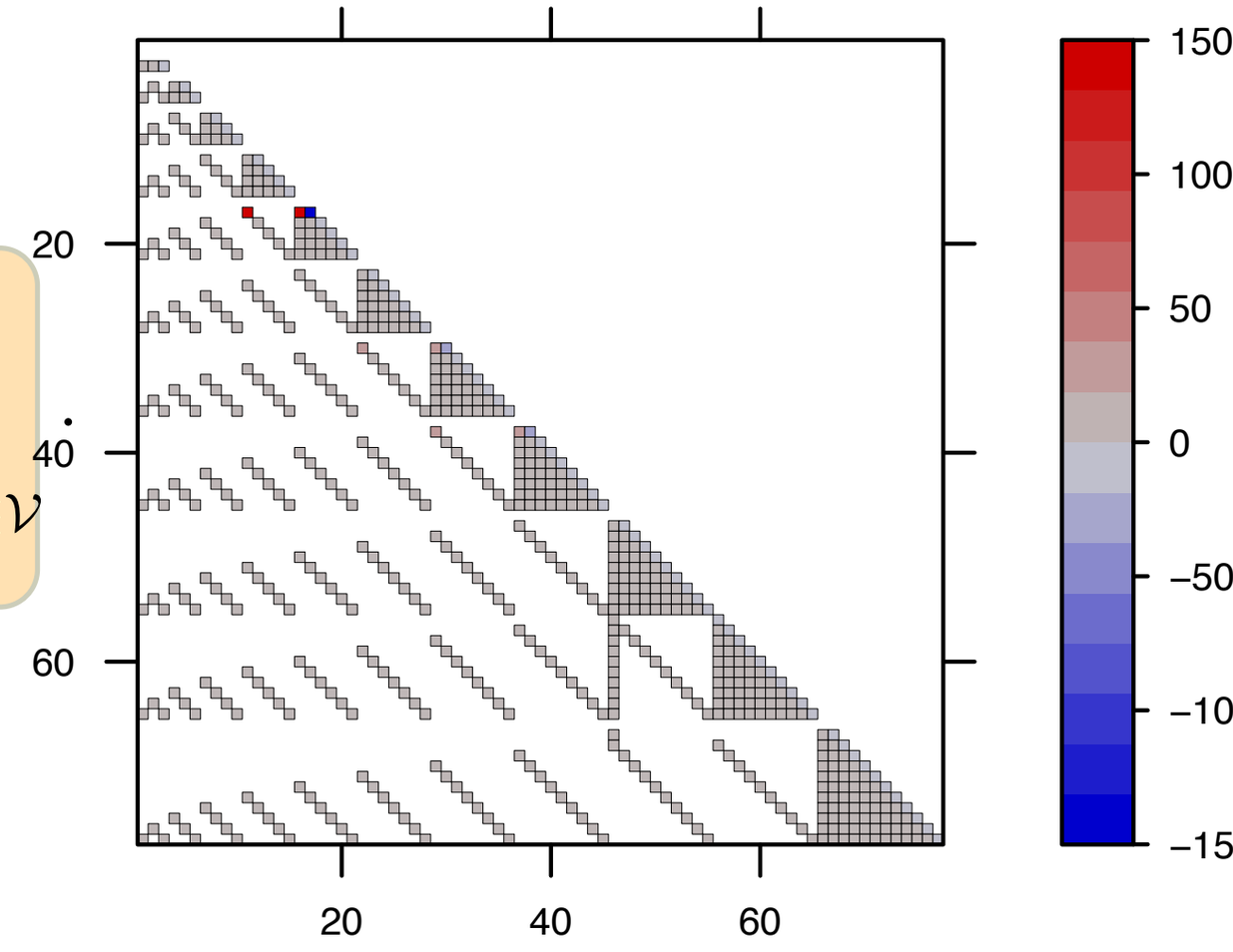
interesting moments...

$\mathrm{E}_q(t)$ the expected number of instances of peptide q at time t.

$$\mathrm{E}_q(t) = \sum_x x_q P(x,t),$$

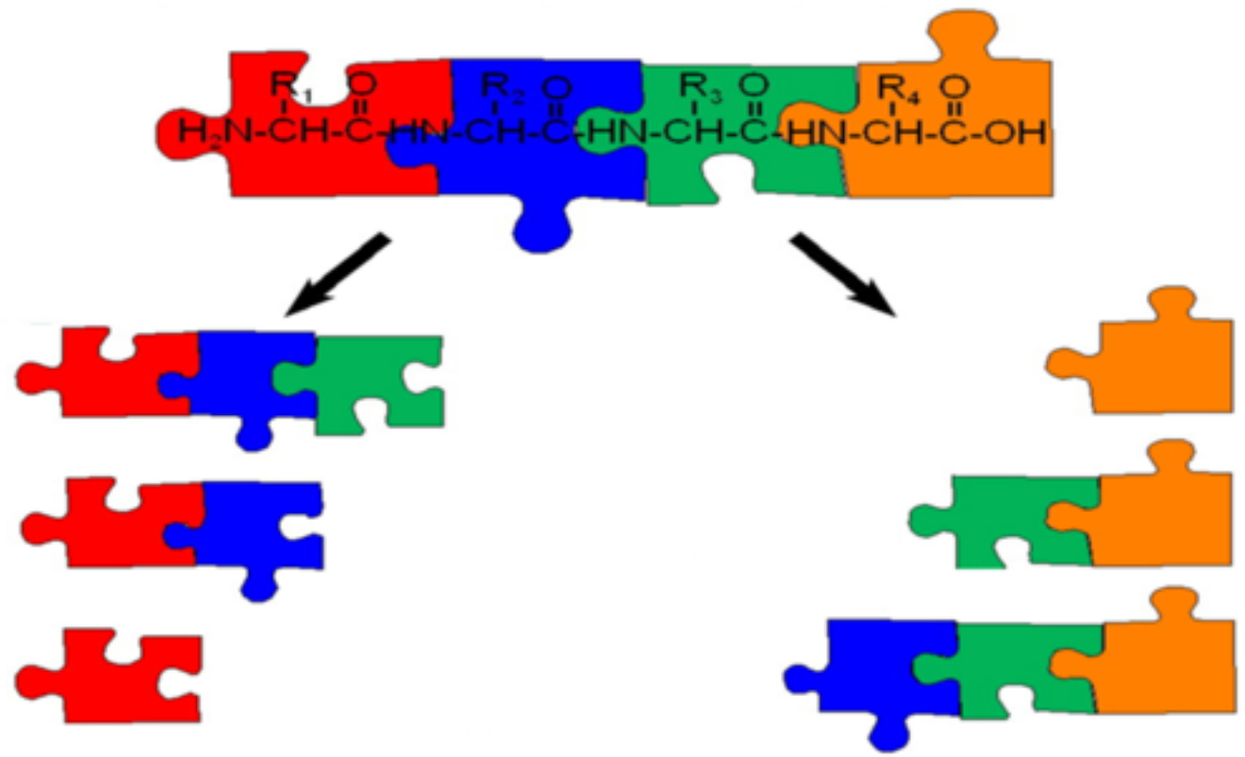
$$\left[\frac{\partial}{\partial t}\mathrm{E}_q(t) = \sum_{u \to q} \lambda_{uq}\,\mathrm{E}_u(t) + \lambda_{qq}\,\mathrm{E}_q(t)\right]_{q \in \mathcal{V}}$$

$$\mathrm{E}(t) = \mathrm{E}(0)^{\mathrm{T}} \exp(\Lambda t)$$



matrix $\Lambda = (\lambda_{vw})_{v,w \in \mathcal{V}}$ for peptide VAHRFKDLGEEN

ETD fragmentation

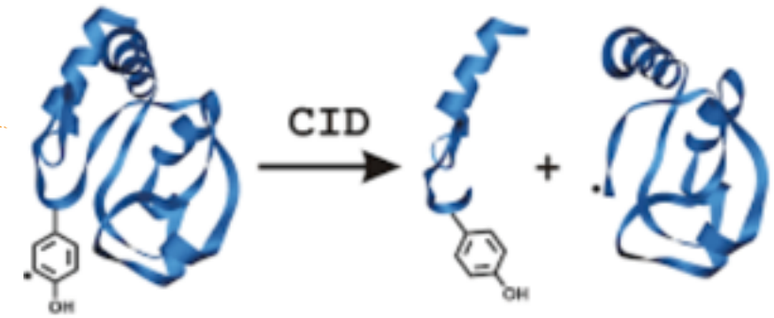


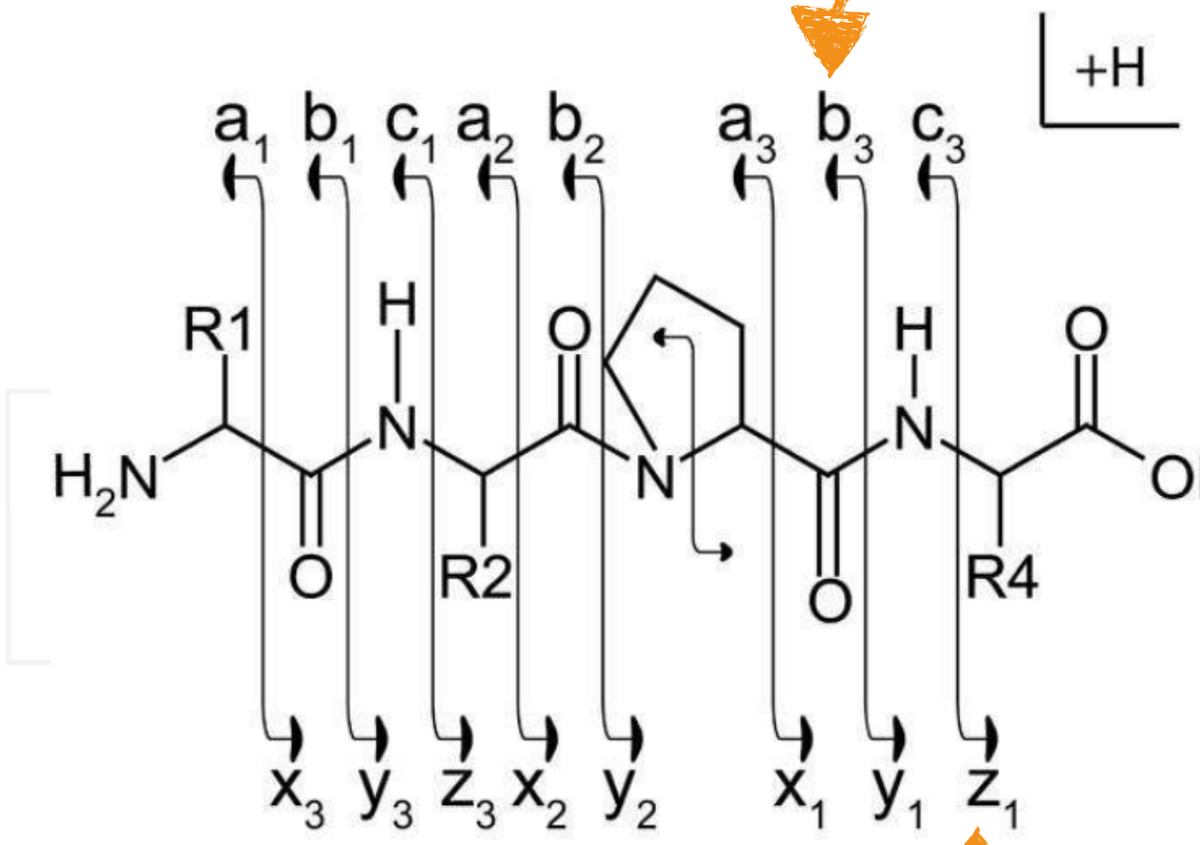
more fragments

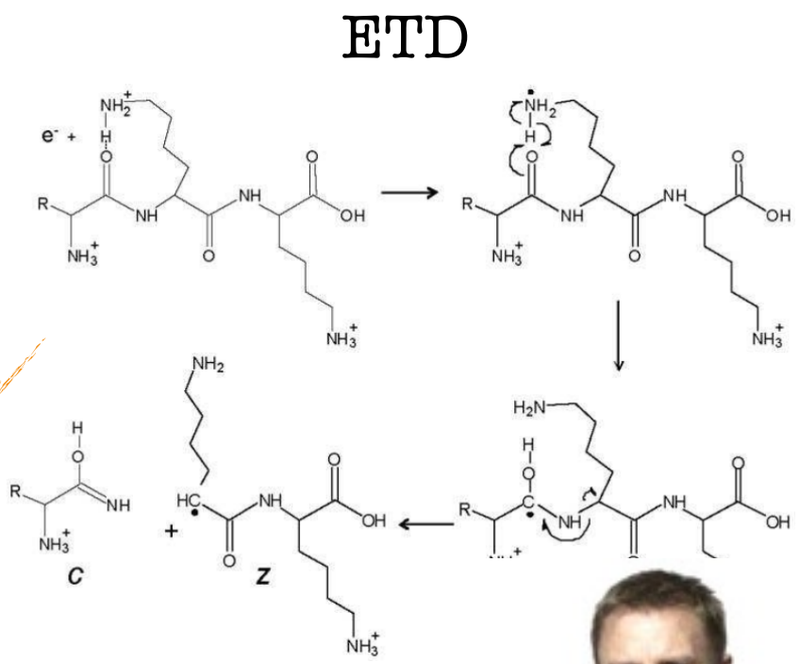
more insight into structure

more confidence in correct identification

some **bonds** get easily broken



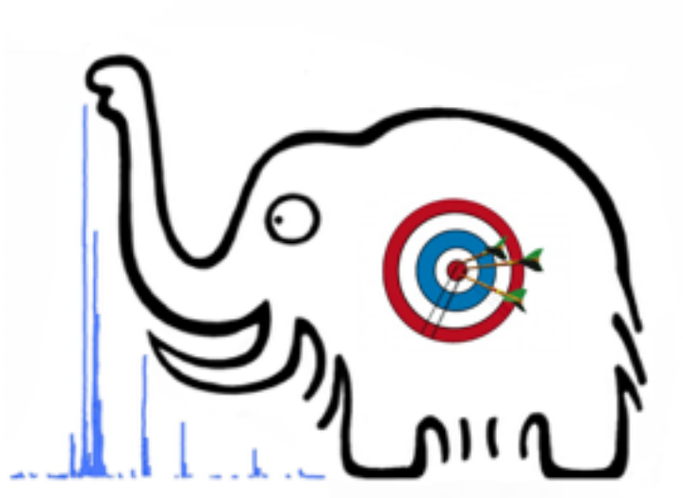




.. others **not**

the goal of masstodon

understand **fragmentation**
inside the instrument
under **different**
experimental **conditions**



solution:

use **purified** chemical **samples**
study **fragmentation pathways**
locate fragments **in data**

1. **deconvolute signals** and
2. infer **fragmentation reaction constants**

using **atomic compositions** of the **fragments** we generate isotopic spectra with

ISOSPEC
The Fine Isotopic Structure Calculator

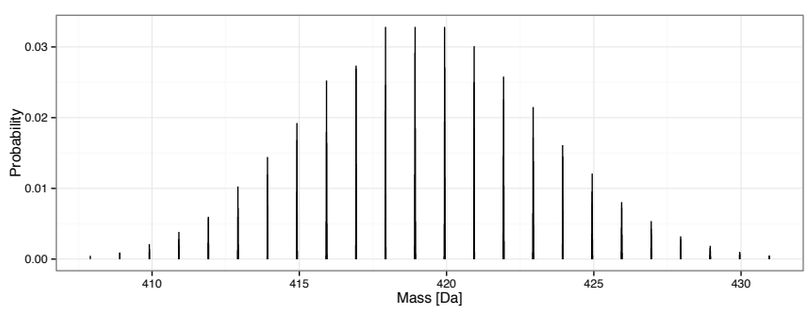
$$P(^{12}C_{c_0}\,^{13}C_{c_1}\,^{1}H_{h_0}\,^{2}H_{h_1}\,^{14}N_{n_0}\,^{15}N_{n_1}\,^{16}O_{o_0}\,^{17}O_{o_1}\,^{18}O_{o_2}\,^{32}S_{s_0}\,^{33}S_{s_1}\,^{34}S_{s_2}\,^{36}S_{s_4}) =$$

$$\binom{c}{c_0, c_1}\mathcal{P}(^{12}C)^{c_0}\mathcal{P}(^{13}C)^{c_1}\binom{h}{h_0, h_1}\mathcal{P}(^{1}H)^{h_0}\mathcal{P}(^{2}H)^{h_1}\binom{n}{n_0, n_1}\mathcal{P}(^{14}N)^{n_0}\mathcal{P}(^{15}N)^{n_1} \times$$

$$\binom{n}{o_0, o_1, o_2}\mathcal{P}(^{16}O)^{o_0}\mathcal{P}(^{17}O)^{o_1}\mathcal{P}(^{18}O)^{o_2}\binom{s}{s_0, s_1, s_2, s_4}\mathcal{P}(^{32}S)^{s_0}\mathcal{P}(^{33}S)^{s_1}\mathcal{P}(^{34}S)^{s_2}\mathcal{P}(^{36}S)^{s_4}$$

frequencies of isotopes

IUPAC
INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY



we can aggregate masses to match **data resolution**

complications

we take into account **charges**

♣ ETD

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

♢ PTR

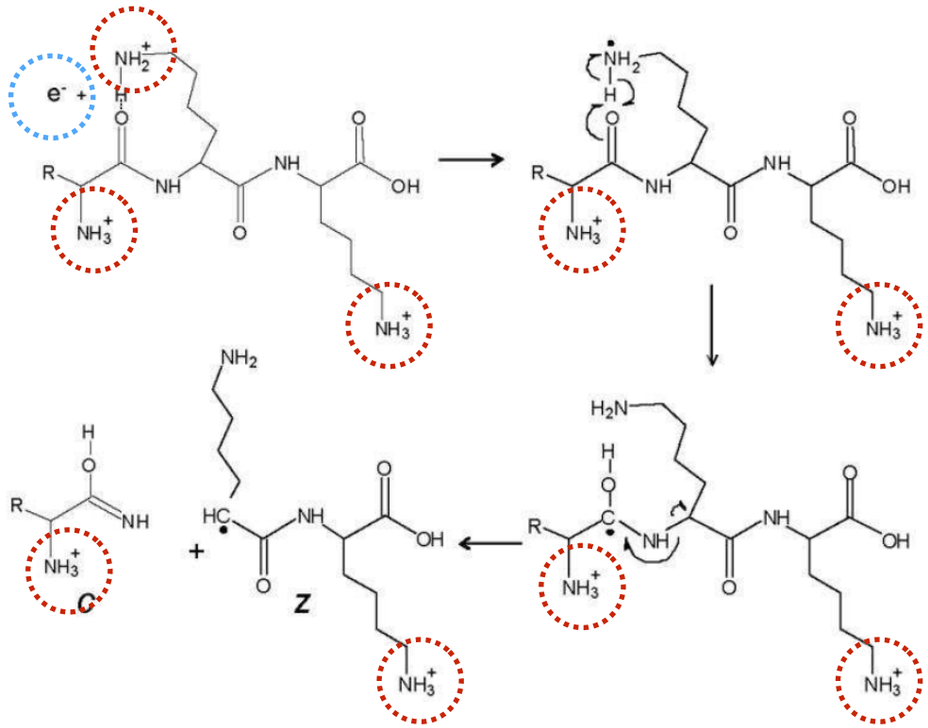
$$[M+nH]^{n+} + A^{\cdot -} \longrightarrow [M+(n-1)H]^{(n-1)+} + AH$$

♡ ETnoD

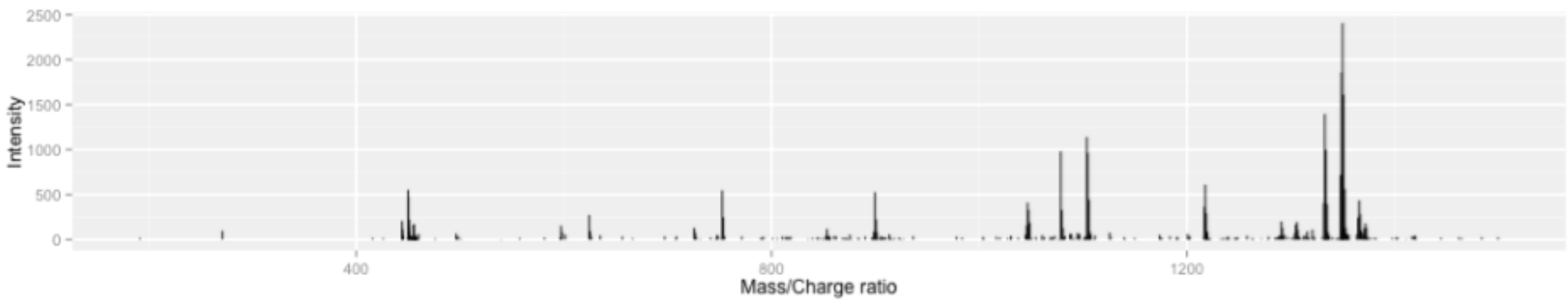
$$[M+nH]^{n+} + A^{\cdot -} \longrightarrow [M+nH]^{(n-1)+\cdot} + A$$

♠ HTR

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

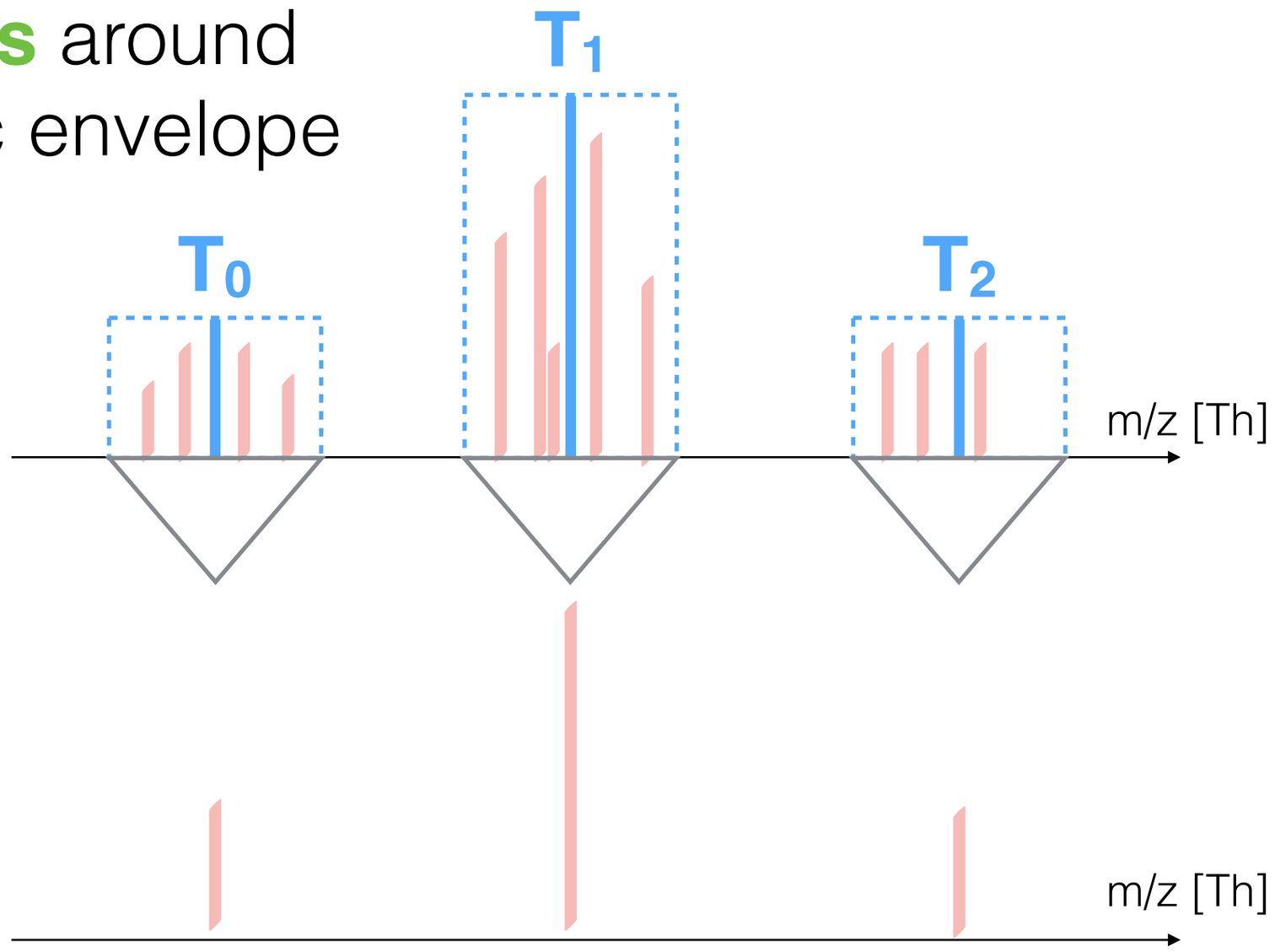


… and **imprecisions** in instrumental **mass calibration**

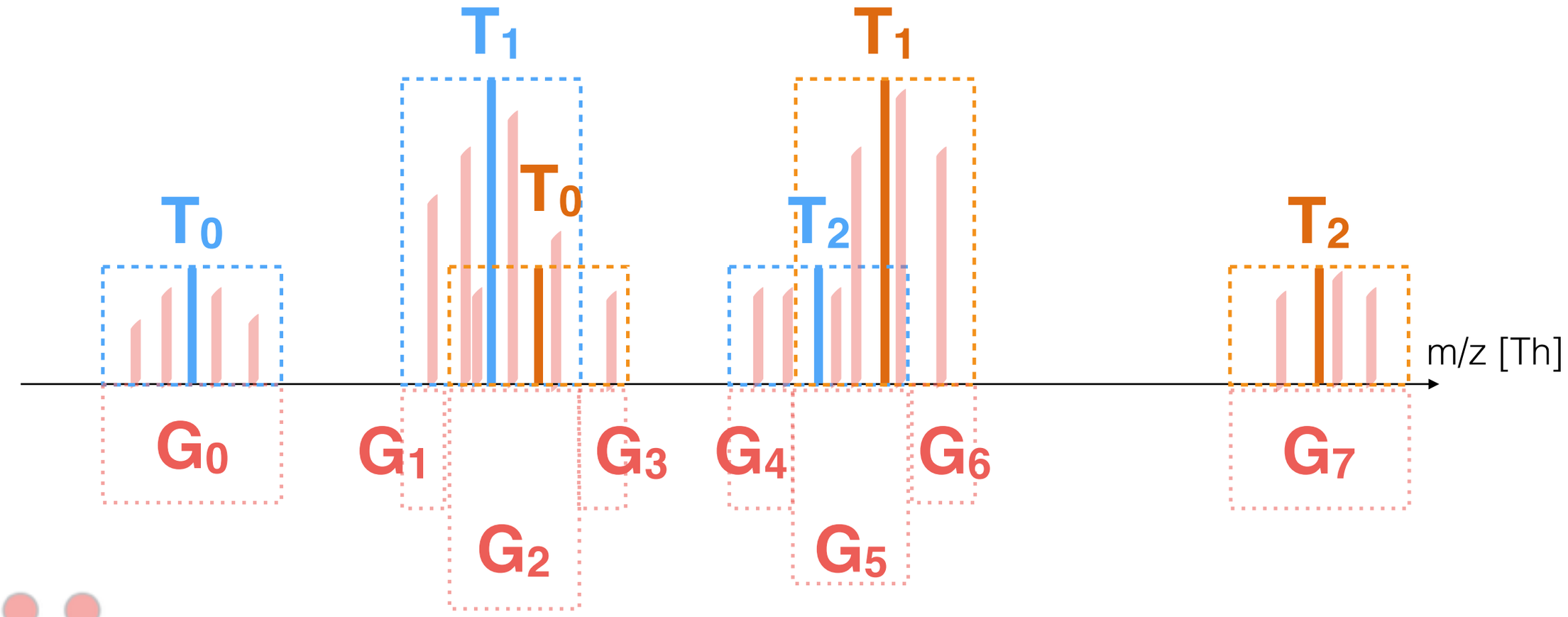


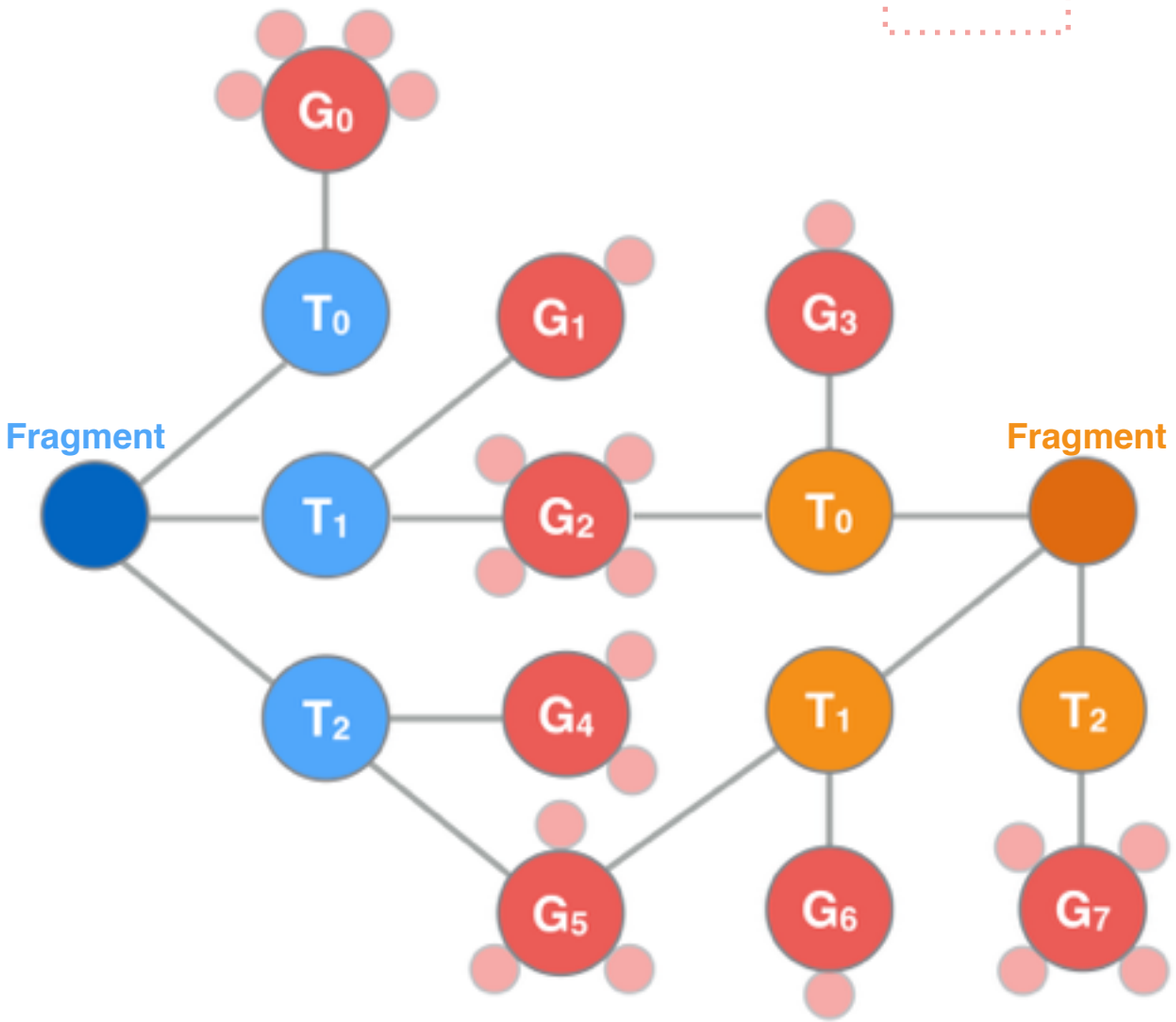
mass imprecisions

tolerance intervals around
theoretical isotopic envelope



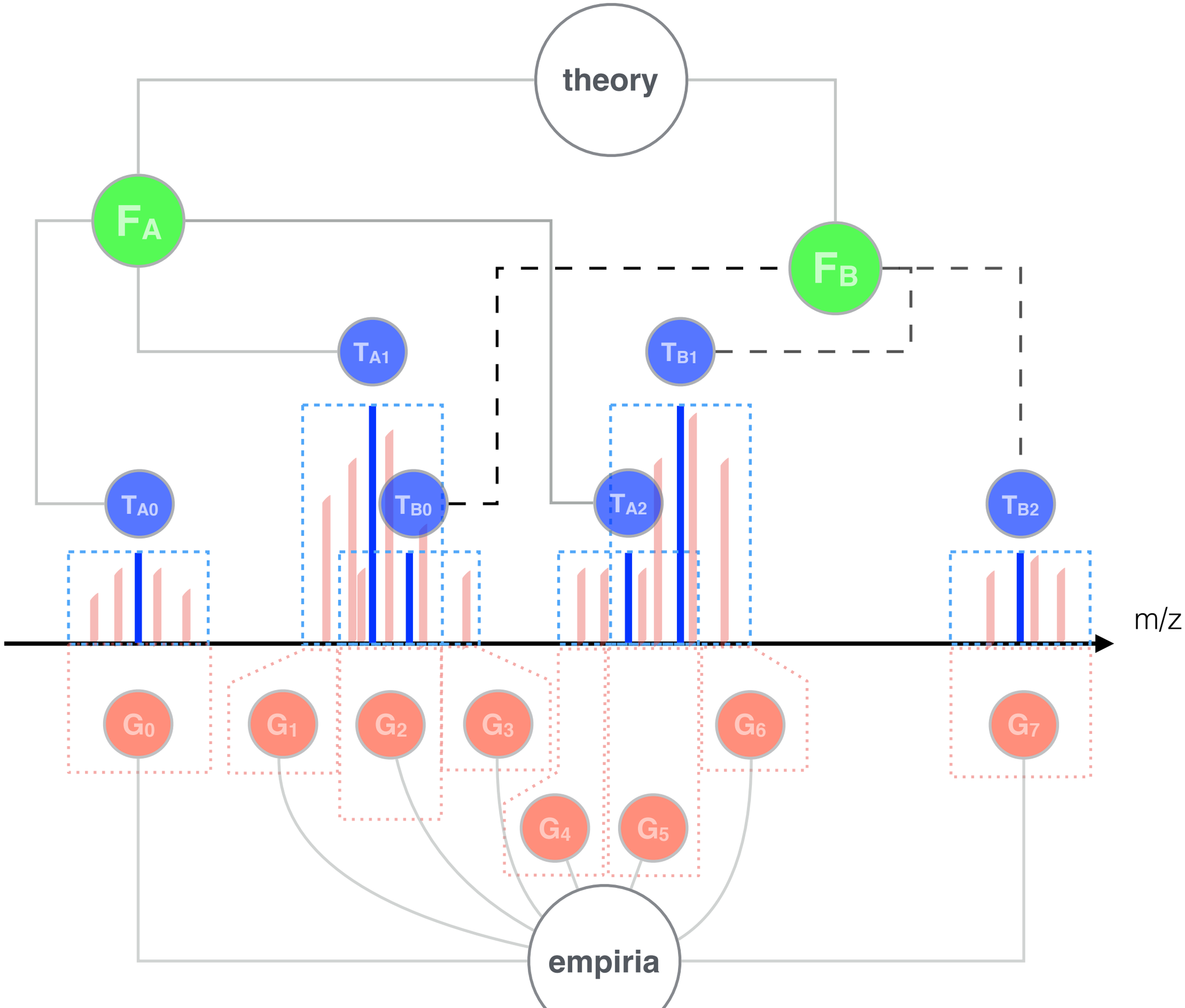
natural data **centroiding**





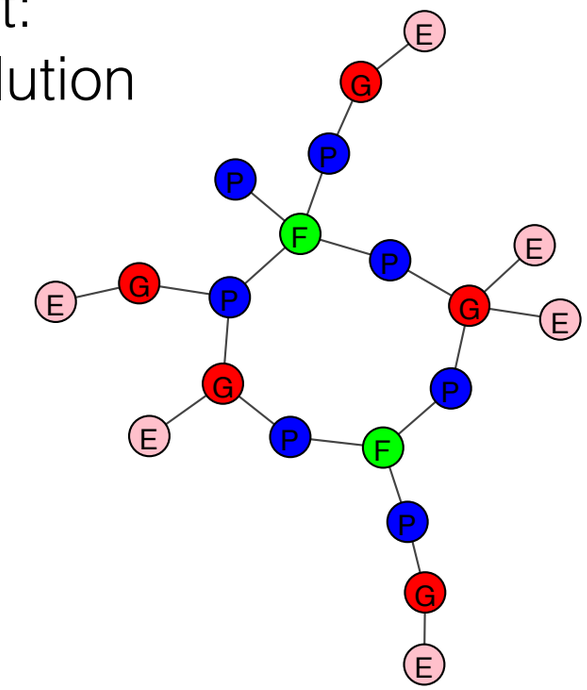
intervals may **overlap**

using **interval trees** we build up the **deconvolution graph**

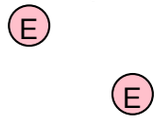

theory
F_A
F_B
T_{A1}
T_{B1}
T_{A0}
T_{B0}
T_{A2}
T_{B2}
m/z
G_0
G_1
G_2
G_3
G_4
G_5
G_6
G_7
empiria

connected components of the **deconvolution graph** provide a wealth of insight into the spectrum

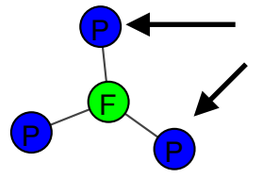
two **fragments** with empirical support: suitable for deconvolution



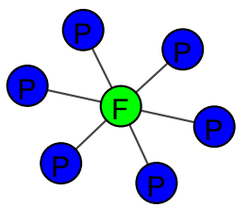
a-theoretical peaks (no **fragments** around)

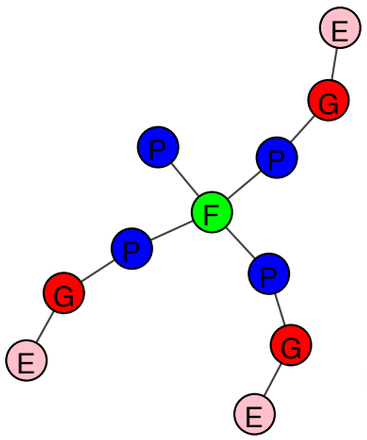


fragment with its **isotopic envelope**



fragment with no empirical support

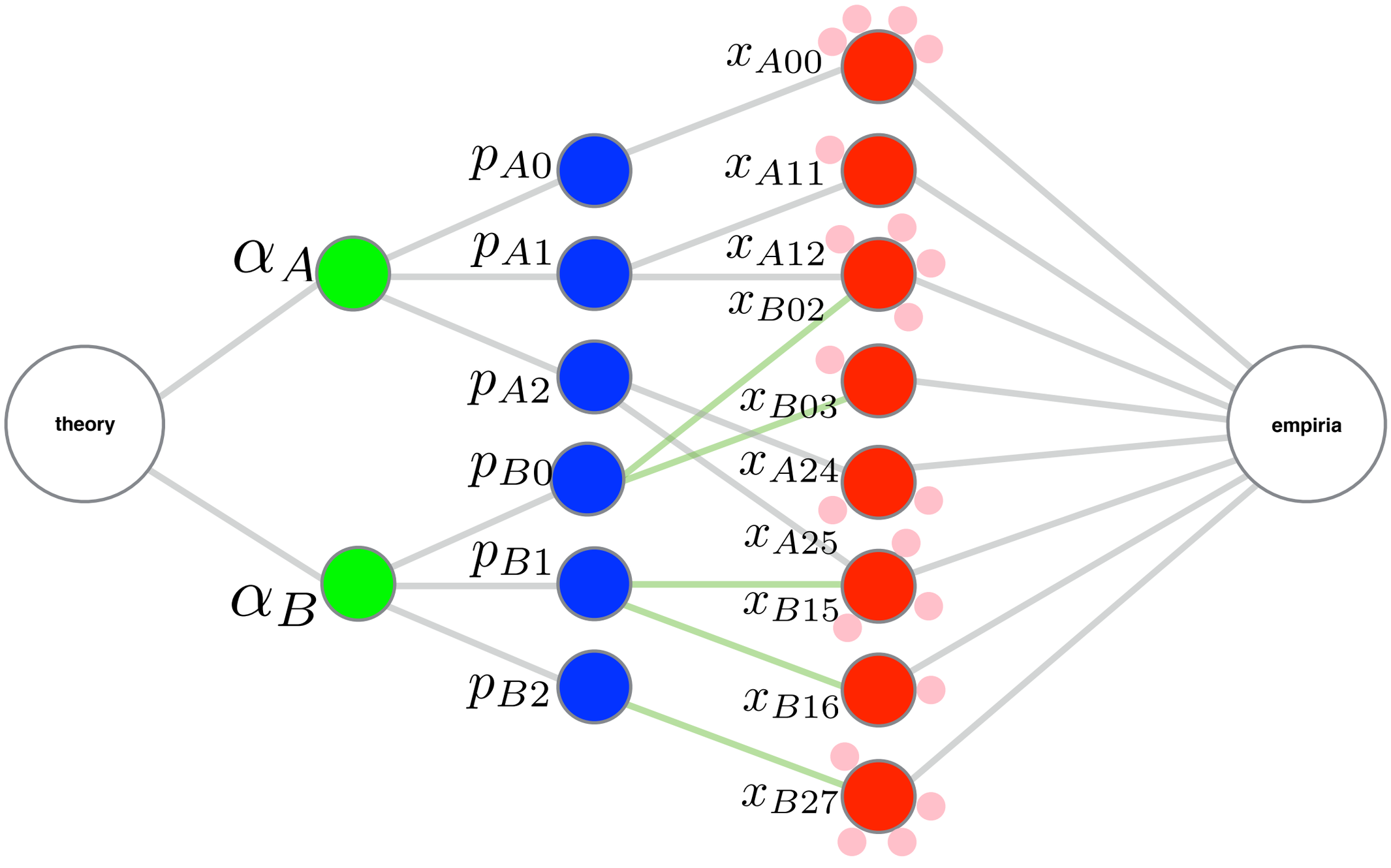




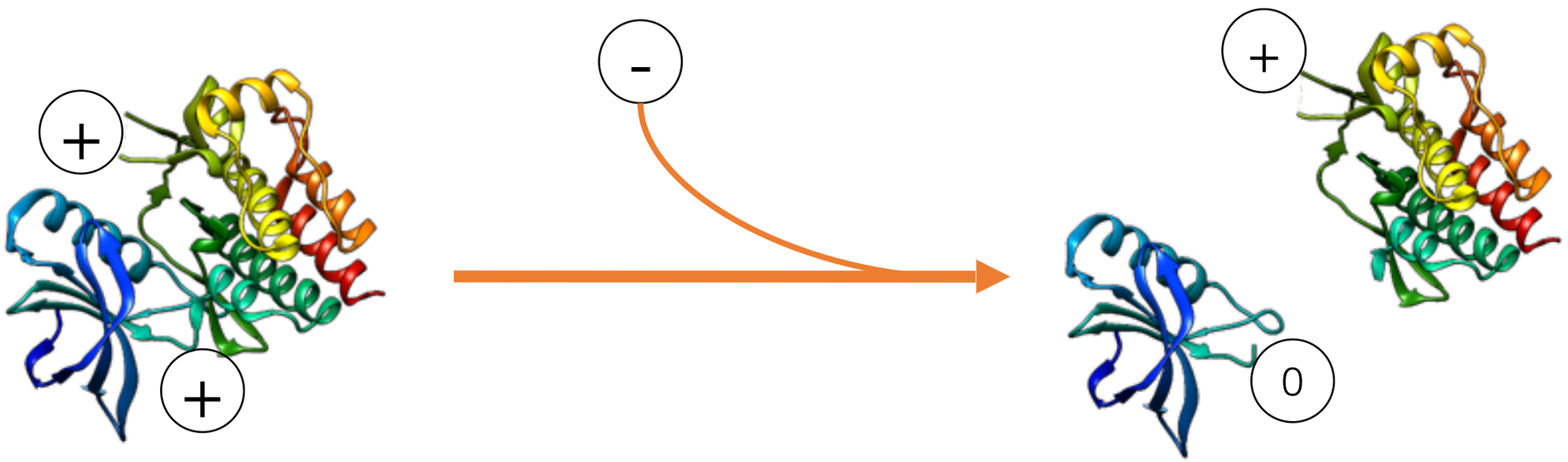
a **fragment** with empirical support: trivial case (no need for deconvolution)

more a-theoretical peaks

to perform **deconvolution** we present the problem
as a **linear programme** similar to the **max flow** problem

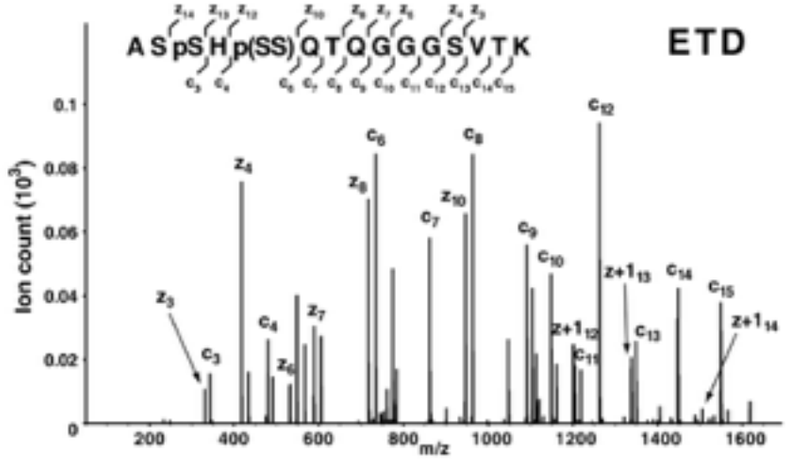


Electron Transfer Dissociation

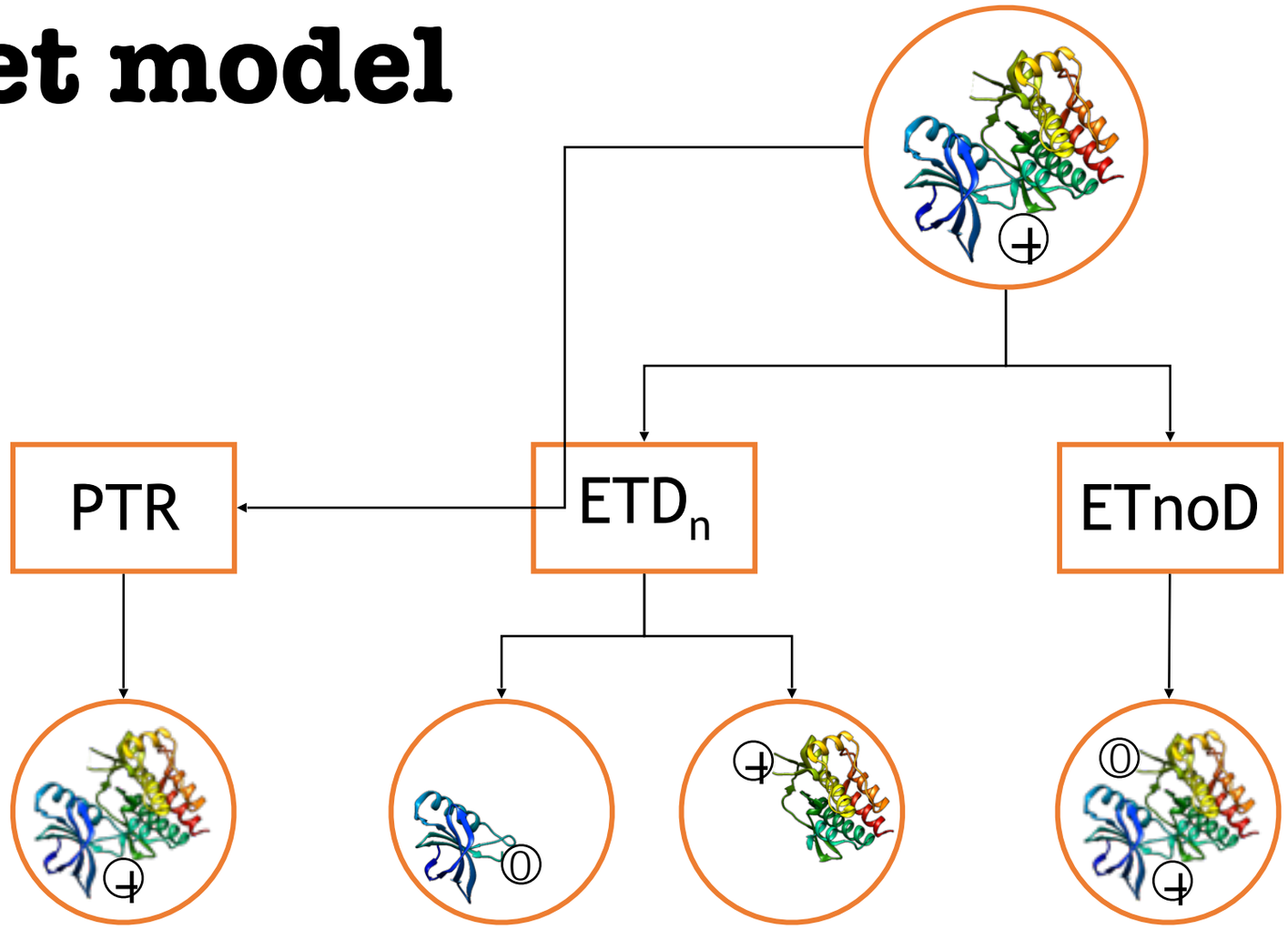


Cleavage of protein backbone by a rapid neutralization of charge

- To identify proteins
- To sequence proteins de novo
- To identify post-translational modifications



Petri Net model



- Electron Transfer Dissociation (ETD):

$$[M + nH]^{n+} \rightarrow [M_1 + n_1H]^{n1+} + [M2 + n_2H]^{n2+}$$

- Proton Transfer Reaction (PTR):

$$[M + nH]^{n+} \rightarrow [M + (n-1)H]^{(n-1)+}$$

- Electron Transfer Without Dissociation (ETnoD):

$$[M + nH]^{n+} \rightarrow [M + nH]^{(n-1)+}$$

Aminoacid Sequence

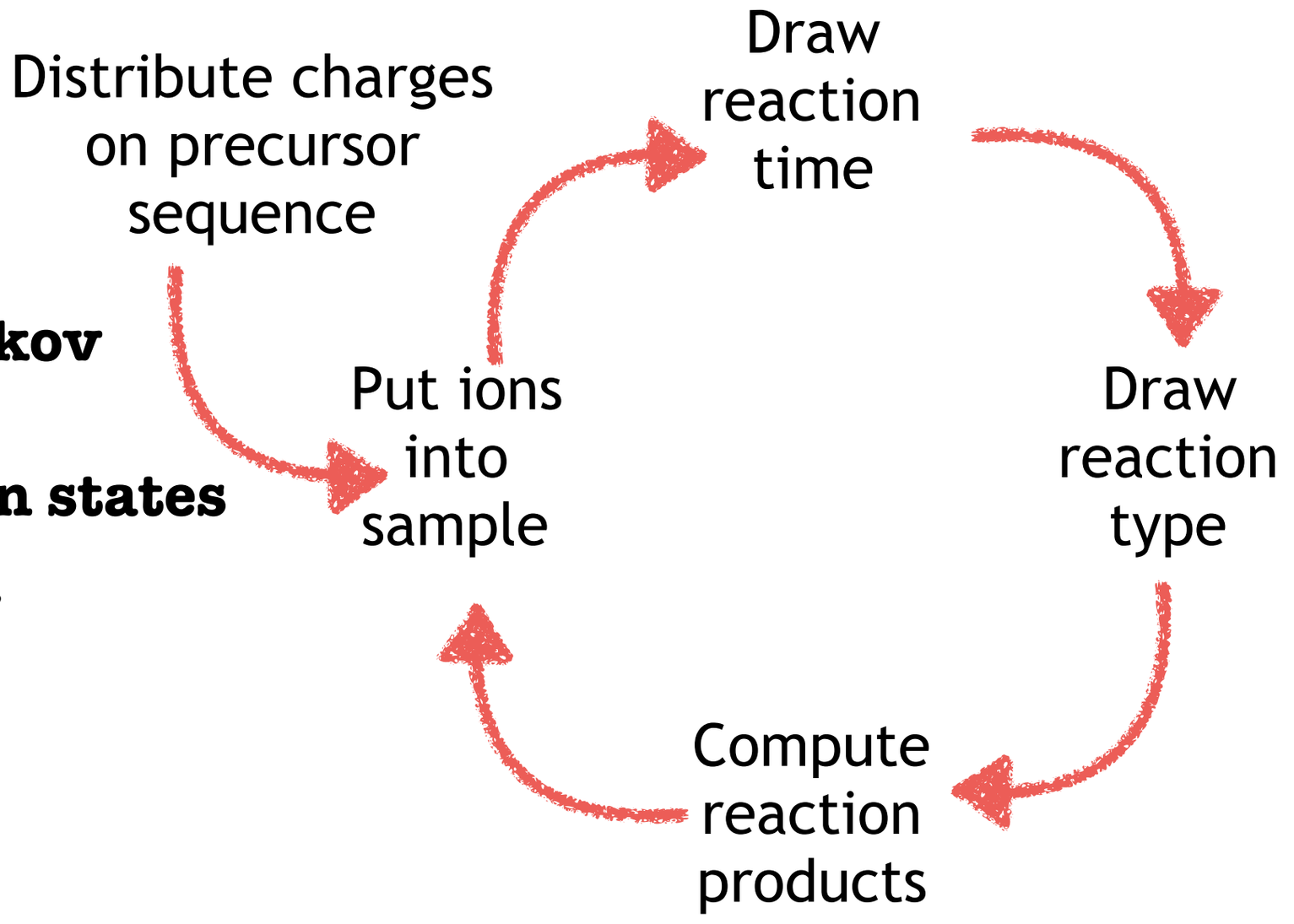
Number of protons quenched by ETnoD

(A, p, q)

Charge

Ion Space Parametrization

- **Evolution of an ion = Markov Jump Process**
- **Jump = transition between states**
- **Jump intensity = reaction intensity**



Population approach

Stochastic description of a single ion

ODE description of a big population of ions

- $x_v(t)$: average number of v molecules at time t
- P: set of input (parent) transitions
- D: set of output (daughter) transitions
- $v(r)$: substrate ion for reaction r
- $I_r = Ip_{v(r)}^2 P_r$: intensity of reaction r

$$\dot{x}_v(t) = \sum_{p \in P} I_p x_{v(p)}(t) - \sum_{d \in D} I_d x_v(t)$$

Tree-like structure

- Each ion has a unique parent (substrate) ion
- Observe that a parent has +1 charge
- No parents for the root! Can be solved analytically

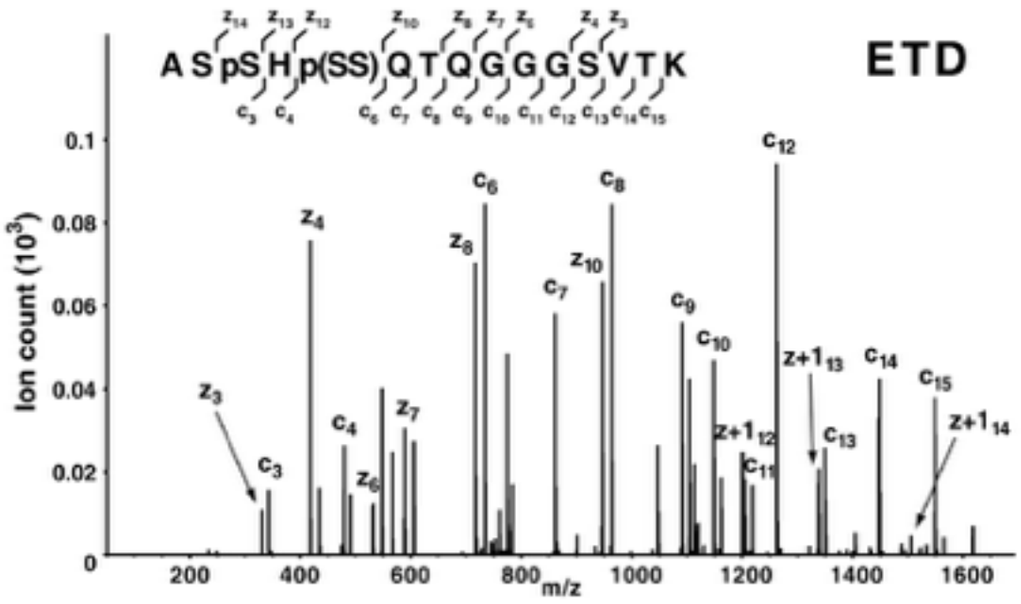
$$\dot{x}_R(t) = -Ip_R^2$$

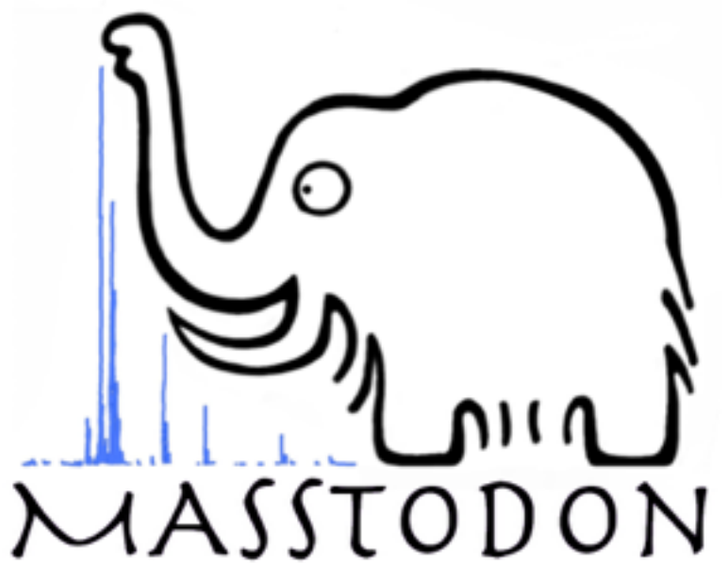
$$x_R(t) = x_0 e^{-Ip_R^2 t}$$

-

$$x_v(t) = \sum_i A_i e^{-B_i t}$$

- A_i, B_i: coefficients dependent on parent transitions
- We have derived recursive formulas for them (also previously described by Gambin & Kluge)





Sequence	Charge	Electrons	Intensity
RPKPQQ	3	0	0.25
RPKP	1	1	0.01
PQQ	1	0	0.12
…	…	…	…

Intensity estimation

- **Idea**: find a set of intensities that best predicts the observed data
  - **Solution**: minimize the discrepancy with the BFGS algorithm

$$\min_{\theta} f(\theta) = \; d(P_\theta, O) = \log\left(\sum_{(p,o)\in P_\theta \times O} (p-o)^2\right)$$

(θ: vector of intensities; P_θ: predicted ion proportions)

Many thanks to collaborators

Michal Startek

Dirk Valkenborg

Blazej Miasojedow

Mateusz Łącki

Frederik Lermyte

Frank Sobott

Michał Ciach

Piotr Dittwald