Isotope distributions

This exposition is based on:

- R. Martin Smith: *Understanding Mass Spectra. A Basic Approach.* Wiley, 2nd edition 2004. [S04]
- Exact masses and isotopic abundances can be found for example at http://www.sisweb.com/referenc/source/exactmaa.htm or http://education.expasy.org/student_projects/isotopident/htdocs/motza.html
- IUPAC Compendium of Chemical Terminology - the Gold Book. http://goldbook.iupac.org/ [GoldBook]
- Sebastian Böcker, Zzuzsanna Lipták: Efficient Mass Decomposition. ACM Symposium on Applied Computing, 2005. [BL05]
- Christian Huber, lectures given at Saarland University, 2005. [H05]
- Wikipedia: http://en.wikipedia.org/, http://de.wikipedia.org/

Isotopes

This lecture addresses some more combinatorial aspect of mass spectrometry related to isotope distributions and mass decomposition.

Most elements occur in nature as a mixture of isotopes. *Isotopes* are atom species of the same chemical element that have different masses. They have the same number of protons and electrons, but a different number of neutrons. The main elements occurring in proteins are CHNOPS. A list of their naturally occurring isotopes is given below.

Isotope	Mass [Da]	% Abundance
^{1}H	1.007825	99.985
^{2}H	2.014102	0.015
^{12}C	12. (exact)	98.90
^{13}C	13.003355	1.10
^{14}N	14.003074	99.63
^{15}N	15.000109	0.37

Isotope	Mass [Da]	% Abundance
^{16}O	15.994915	99.76
^{17}O	16.999131	0.038
^{18}O	17.999159	0.20
^{31}P	30.973763	100.
^{32}S	31.972072	95.02
^{33}S	32.971459	0.75
^{34}S	33.967868	4.21

Isotopes (2)

Note that the lightest isotope is also the most abundant one for these elements. Here is a list of the heavy isotopes, sorted by abundance:

Isotope	Mass [Da]	% Abundance
^{34}S	33.967868	4.21
^{13}C	13.003355	1.10
^{33}S	32.971459	0.75
^{15}N	15.000109	0.37
^{18}O	17.999159	0.20
^{17}O	16.999131	0.038
^{2}H	2.014102	0.015

We see that sulfur has a big impact on the isotope distribution. But it is not always present in a peptide (only the amino acids Cystein or Methionin contain sulfur). Apart from that, ^{13}C is most abundant, followed by ^{15}N. These isotopes lead to "+1" peaks. The heavy isotopes ^{18}O and ^{34}S lead to "+2" peaks. Note that ^{17}O and ^{2}H are very rare.

Isotopes (3)

The two isotopes of hydrogen have special names: ^1H is called *protium*, and $^2H = D$ is called *deuterium* (or sometimes "heavy" hydrogen).

Note that whereas the exact masses are universal physical constants, the relative abundances are different at each place on earth and can in fact be used to trace the origin of substances. They are also being used in *isotopic labeling* techniques.

The standard unit of mass, the *unified atomic mass unit*, is defined as $1/12$ of the mass of ^{12}C and denoted by *u* or *Da*, for *Dalton*. Hence the atomic mass of ^{12}C is $12\,u$ by definition. The atomic masses of the isotopes of all the other elements are determined as ratios against this standard, leading to non-integral values for essentially all of them.

The subtle differences of masses are due to the *mass defect* (essentially, the binding energy of the nucleus). We will return to this topic later. For understanding the next few slides, the difference between nominal and exact masses is not essential.

Isotopes (4)

The *average atomic mass* (also called the *average atomic weight* or just *atomic weight*) of an element is defined as the weighted average of the masses of all its naturally occurring stable isotopes.

For example, the average atomic mass of carbon is calculated as

$$\frac{(98.9\% * 12.0 + 1.1\% * 13.003355)}{100\%} \doteq 12.011$$

For most purposes such as weighing out bulk chemicals only the average molecular mass is relevant since what one is weighing is a statistical distribution of varying isotopic compositions.

The *monoisotopic mass* is the sum of the masses of the atoms in a molecule using the *principle* isotope mass of each atom instead of the isotope averaged atomic mass and is most often used in mass spectrometry. The monoisotopic mass of carbon is 12.

Isotopes (5)

According to the [GoldBook] the *principal ion* in mass spectrometry is a molecular or fragment ion which is made up of the *most abundant* isotopes of each of its atomic constituents.

Sometimes compounds are used that have been artificially isotopically enriched in one or more positions, for example $CH_3{}^{13}CH_3$ or CH_2D_2. In these cases the principal ion may be defined by treating the heavy isotopes as new atomic species. Thus, in the above two examples, the principal ions would have masses 31 (not 30) and 18 (not 16), respectively.

In the same vein, the *monoisotopic mass spectrum* is defined as a spectrum containing only ions made up of the principal isotopes of atoms making up the original molecule.

You will see that the monoisotopic mass is sometimes defined using the *lightest* isotope. In most cases the distinction between ”principle” and ”lightest” isotope is non-existent, but there *is* a difference for some elements, for example iron and argon.

Isotopic distributions

The mass spectral peak representing the monoisotopic mass is not always the most abundant isotopic peak in a spectrum – although it stems from the most abundant isotope of each atom type.

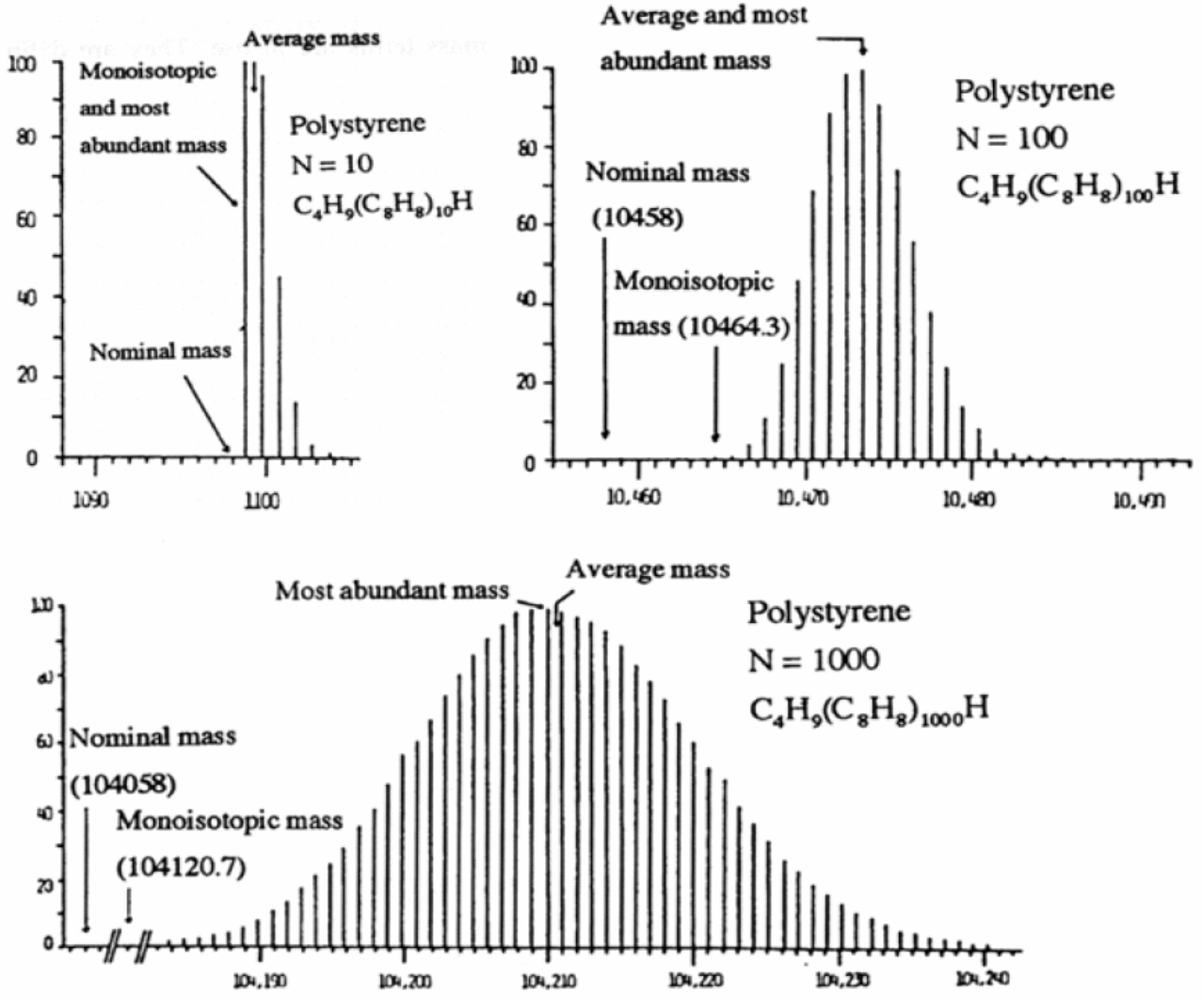
This is due to the fact that as the number of atoms in a molecule increases the probability of the entire molecule containing at least one heavy isotope increases. For example, if there are 100 carbon atoms in a molecule, each of which has an approximately 1% chance of being a heavy isotope, then the whole molecule is not unlikely to contain at least one heavy isotope.

The monoisotopic peak is sometimes not observable due to two primary reasons.

- The monoisotopic peak may not be resolved from the other isotopic peaks. In this case only the average molecular mass may be observed.
- Even if the isotopic peaks are resolved, the monoisotopic peak may be below the noise level and heavy isotopomers may dominate completely.

Isotopic distributions (2)

Terminology



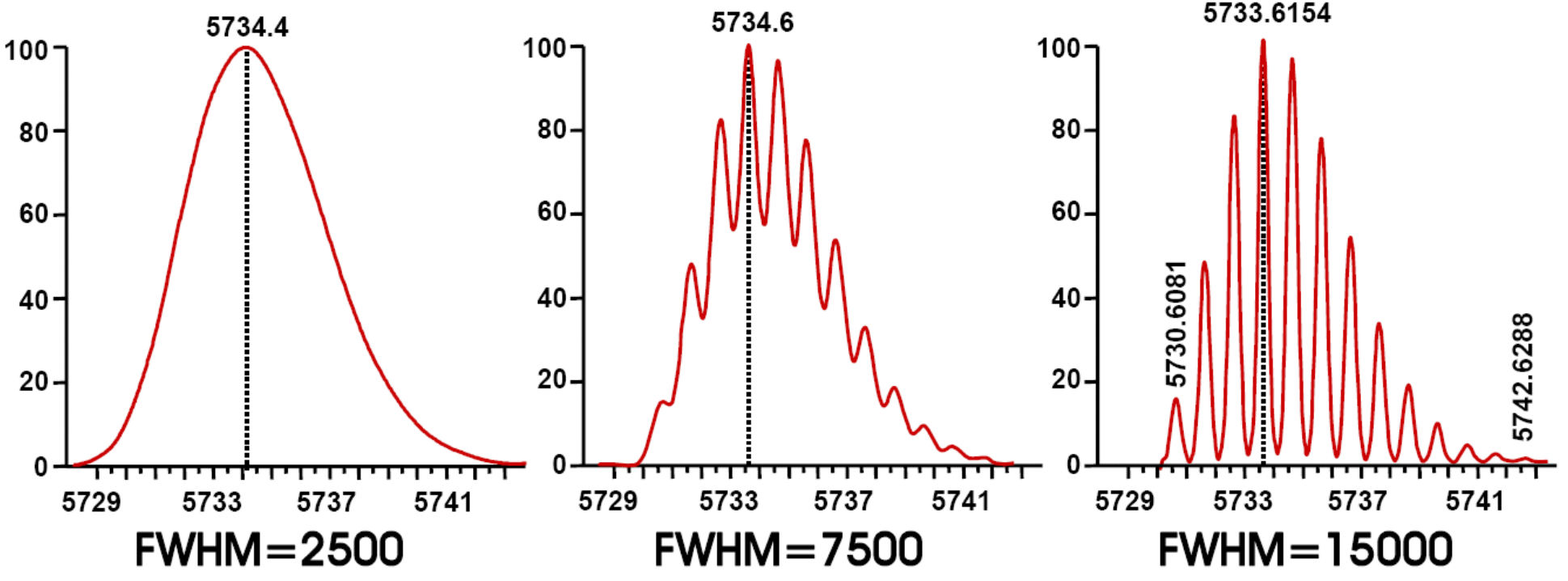
mit zunehmender Masse:

- wird die Isotopenverteilung symmetrischer
- mittlere Masse und intensivste Masse nähern sich an
- steigt die Auflösung, die zur Darstellung des Isotopenmusters notwendig ist
- ist das Signal für die monoisotopische Masse im Spektrum nicht mehr erkennbar

Isotopic distributions (3)

Example:

Isotopenmuster von $[M+H]^+$ von bovinem Insulin $[C_{254}H_{378}N_{65}O_{75}S_6]^+$ (P01327), m_{mono}=5730.61081, $m_{average}$=5734.58948



Isotopic distributions (4)

To summarize: Learn to distinguish the following concepts!

- nominal mass
- monoisotopic mass
- most abundant mass
- average mass

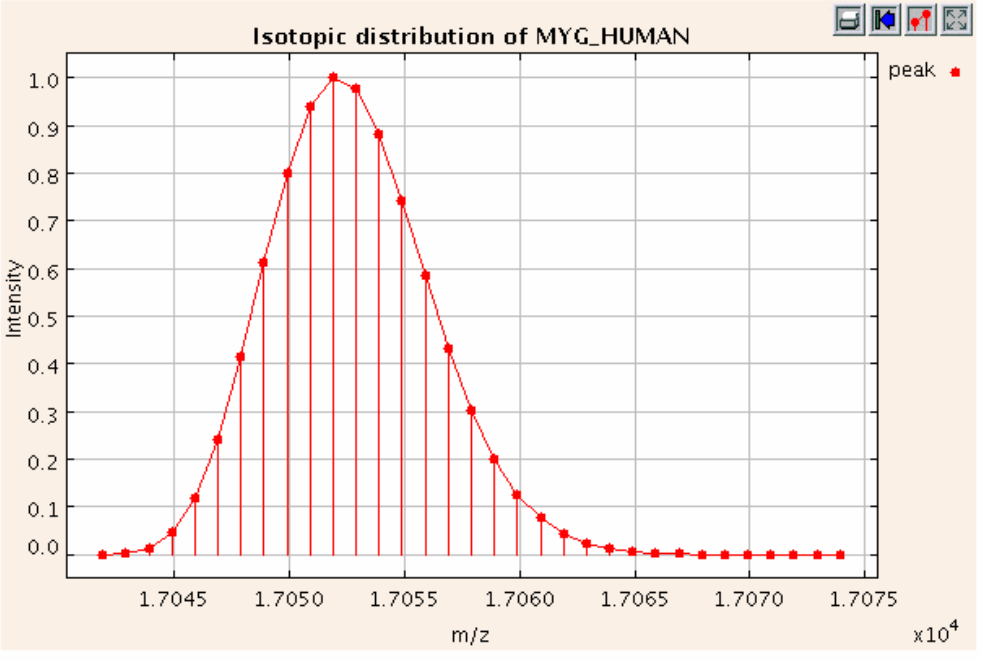
Isotopic distributions (5)

Example: Isotopic distribution of human myoglobin

The name of the query is: MYG_HUMAN
The type of composition you've chosen is: Protein One-letter code [M]
You have entered:
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKP
LAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

Element	Mass
C_{769}	9236.228
H_{1215}	1224.647
N_{209}	2927.400
O_{221}	3535.867
S_4	128.260
Average mass:	17052.403

The monoisotopic mass is: 17041.914
Monoisotopic combination:
$^{12}C_{769}$
$^{1}H_{1215}$
$^{14}N_{209}$
$^{16}O_{221}$
$^{32}S_4$



Most likely isotope combination:
$^{12}C_{761}$
$^{13}C_{8}$
$^{1}H_{1215}$
$^{14}N_{209}$
$^{16}O_{221}$
$^{32}S_4$
Exact mass is 17049.941
Probability of combination is 2.559%
The most likely combination is 27.52% of those masses rounding to 17041 amu.

Download the data

Screen shot: http://education.expasy.org/student_projects/isotopident/

Isotopic distributions (6)

A basic computational task is:

- Given an ion whose atomic composition is known, how can we compute its isotopic distribution?

We will ignore the mass defects for a moment. It is convenient to number the peaks by their number of additional mass units, starting with zero for the lowest isotopic peak. We can call this the *isotopic rank*.

Let E be a chemical element. Let $\pi_E[i]$ denote the probability of the isotope of E having i additional mass units. Thus the relative intensities of the isotopic peaks for a single atom of element E are $(\pi_E[0], \pi_E[1], \pi_E[2], ..., \pi_E[k_E])$. Here k_E denotes the isotopic rank of the heaviest isotope occurring in nature. We have $\pi_E[\ell] = 0$ for $\ell > k$.

For example carbon has $\pi_C[0] = 98.9\% = 0.989$ (isotope ^{12}C) and $\pi_C[1] = 1.1\% = 0.011$ (isotope ^{13}C).

Isotopic distributions (7)

The probability that a molecule composed out of one atom of element E and one atom of element E' has a total of n additional neutrons is

$$\pi_{EE'}[n] = \sum_{i=0}^{n} \pi_E[i]\,\pi_{E'}[n-i]\,.$$

Note that $\pi_{EE'}[\ell] = 0$ for $\ell > k_E + k'_E$.

This type of composition is very common in mathematics and known as a *convolution* operation, denoted by the operator $*$.

Using the convolution operator, we can rewrite the above equation as

$$\pi_{EE'} = \pi_E * \pi_{E'}\,.$$

For example, a hypothetical molecule composed out of one carbon and one nitrogen would have $\pi_{CN} = \pi_C * \pi_N$,

$$\begin{aligned}
\pi_{CN}[0] &= \pi_C[0]\pi_N[0]\,,\\
\pi_{CN}[1] &= \pi_C[0]\pi_N[1] + \pi_C[1]\pi_N[0]\,,\\
\pi_{CN}[2] &= \pi_C[0]\pi_N[2] + \pi_C[1]\pi_N[1] + \pi_C[2]\pi_N[0] \quad \left(= \pi_C[1]\pi_N[1]\right).
\end{aligned}$$

Isotopic distributions (8)

Clearly the same type of formula applies if we compose a larger molecule out of smaller molecules or single atoms. Molecules have isotopic distributions just like elements.

For simplicity, let us define *convolution powers*. Let $\rho^1 := \rho$ and $\rho^n := \rho^{n-1} * \rho$, for any isotopic distribution ρ. Moreover, it is natural to define ρ^0 by $\rho^0[0] = 1$, $\rho^0[\ell] = 0$ for $\ell > 0$. This way, ρ^0 will act as neutral element with respect to the convolution operator $*$, as expected.

Then the isotopic distribution of a molecule with the chemical formula $E^1_{n_1} \cdots E^\ell_{n_\ell}$, composed out of the elements $E^1, \ldots, E^\ell$, can be calculated as

$$\pi_{E^1_{n_1} \cdots E^\ell_{n_\ell}} = \pi^{n_1}_{E_1} * \ldots * \pi^{n_\ell}_{E_\ell} .$$

Isotopic distributions (9)

This immediately leads to an algorithm. for computing the isotopic distribution of a molecule.

Now let us estimate the running time for computing $\pi_{E_1}^{n_1} * ... * \pi_{E_l}^{n_l}$.

- The number of convolution operations is $n_1 + ... + n_l - 1$, which is linear in the number of atoms.

- Each convolution operation involves a summation for each $\pi[i]$. If the highest isotopic rank for E is k_E, then the highest isotopic rank for E_n is $k_E n$. Again, this in linear in the number of atoms.

We can improve on both of these factors.

Do you see how?

Isotopic distributions (10)

Bounding the range of isotopes

For practical cases, it is possible to restrict the summation range in the convolution operation.

In principle it is possible to form a molecule solely out of the heaviest isotopes, and this determines the summation range needed in the convolution calculations. However, the abundance of such an isotopomer is vanishingly small. In fact, it will soon fall below the inverse of the Avogadro number (6.0221415×10^{23}), so we will hardly ever see a single molecule of this type.

For simplicity, we let us first consider a single element E and assume that $k_E = 1$. (For example, E could be carbon or nitrogen.) In this case the isotopic distribution is a binomial with parameter $p := \pi_E[1]$,

$$\pi_{E^n}[k] = \binom{n}{k} p^k (1-p)^{n-k} .$$

The mean of this binomial distribution is pn. Large deviations can be bounded as follows.

Isotopic distributions (11)

We use the upper bound for binomial coefficients

$$\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k .$$

For $k \geq 3pn$ (that is, k three times larger than expected) we get

$$\binom{n}{k} p^k \leq \left(\frac{nep}{2pn}\right)^k = \left(\frac{e}{3}\right)^k .$$

and hence

$$\sum_{\ell \geq 3pn}^{n} \binom{n}{k} p^k (1-p)^{n-k} \leq \sum_{\ell \geq 3pn}^{n} \left(\frac{e}{3}\right)^k = O\left(\left(\frac{e}{3}\right)^{3pn}\right) = o(1) .$$

While $3pn$ is still linear in n, it is much smaller – in practice one can usually restrict the calculations to less than 10 isotopic variants for peptides.

Fortunately, if it turns out that the chosen range was too small, we can detect this afterwards because the probabilities will no longer add up to 1. (*Exercise:* Explain why.) Thus we even have an 'a posteriori error estimate'.

Isotopic distributions (12)

More generally, a peptide is composed out of the elements C, H, N, O, S. For each of these elements the lightest isotope has a natural abundance above 95% and the highest isotopic rank is at most 2. Again we can bound the sum of the abundances of heavy isotopic variants by a binomial distribution:

$$\sum_{j \geq i} \pi_{E_{n_1}^1 \cdots E_{n_\ell}^\ell}[j] \;\leq\; \sum_{j \geq i/2} \binom{n}{j} 0.05^j 0.95^{n-j} \,.$$

(In order to get i additional mass units, at least $i/2$ of the atoms must be 'heavy'.)

Isotopic distributions (13)

Computing convolution powers by iterated squaring

There is another trick which can be used to save the number of convolutions needed to calculate the n-th convolution power π^n of an elemental isotope distribution.

Observe that just like for any associative operation $*$, the convolution powers satisfy

$$\pi^{2n} = \pi^n * \pi^n .$$

In general, n is not a power of two, so let $(b_j, b_{j-1}, \ldots, b_0)$ be the bits of the binary representation of n, that is,

$$n = \sum_{\ell=0}^{j} b_\ell 2^\ell \;=\; b_j 2^j + b_{j-1} 2^{j-1} + \cdots + b_0 2^0 .$$

Then we can compute π^n as follows:

$$\pi^n = \pi^{\sum_j 2^j b_j} = \prod_j \pi^{2^j b_j} \;=\; \pi^{b_j 2^j} * \pi^{b_{j-1} 2^{j-1}} * \cdots * \pi^{b_0 2^0} ,$$

where the $\prod$ is of course meant with respect to $*$. The total number of convolutions needed for this calculation is only $O(\log n)$.

Isotopic distributions (14)

To summarize: The first $k+1$ abundances $\pi_{E_1^{n_1}\cdots E_\ell^{n_\ell}}[i]$, $i = 0, \ldots, k$, of the isotopic distribution of a molecule $E_1^{n_1} \cdots E_\ell^{n_\ell}$ can be computed in $O(\ell k \log n)$ time and $O(\ell k)$ space, where $n = n_1 + \ldots + n_\ell$.

(*Exercise:* To test your understanding, check how these resource bounds follow from what has been said above.)

Mass decomposition

A related question is:

- Given a peak mass, what can we say about the elemental composition of the ion that generated it?

In most cases, one cannot obtain much information about the chemical structural from just a single peak. The best we can hope for is to obtain the chemical formula with isotopic information attached to it. In this sense, the total mass of an ion is decomposed into the masses of its constituents, hence the term mass decomposition.

Mass decomposition (2)

This is formalized by the concept of a *compomer* [BL05].

We are given an *alphabet* Σ of size $|\Sigma| = k$, where each *letter* has a *mass* a_i, $i = 1, \ldots, k$. These letters can represent atom types, isotopes, or amino acids, or nucleotides. We assume that all masses are different, because otherwise we could never distinguish them anyway. Thus we can identify each letter with its mass, i. e., $\Sigma = \{a_1, \ldots, a_k\} \subset \mathbb{N}$. This is sometimes called a *weighted alphabet*.

The *mass of a string* $s = s_1 \ldots s_n \in \Sigma^*$ is defined as the sum of the masses of its letters, i. e., $\text{mass}(s) = \sum_{i=1}^{|s|} s_i$.

Formally, a *compomer* is an integer vector $c = (c_1, \ldots, c_k) \in (\mathbb{N}_0)^k$. Each c_i represents the number of occurrences of letter a_i. The *mass* of a compomer is $\text{mass}(c) := \sum_{i=1}^{k} c_i a_i$, as opposed to its *length*, $|c| := \sum_{i=1}^{k} c_i$.

In short: A compomer tells us how many instances of an atomic species are present in a molecule. We want to find all compomers whose mass is equal to the observed mass.

Mass decomposition (3)

There a many more strings (molecules) than compomers, but the compomers are still many.

For a string $s = s_1 \ldots s_n \in \Sigma^*$, we define comp($s$) := $(c_1, \ldots, c_k)$, where $c_i := \#\{j \mid s_j = a_i\}$. Then comp($s$) is the compomer associated with s, and vice versa.

One can prove (*exercise*):

1. The number of strings associated with a compomer $c = (c_1, \ldots, c_k)$ is $\binom{|c|}{c_1,\ldots,c_k} = \frac{|c|!}{c_1! \cdots c_k!}$.
2. Given an integer n, the number of compomers $c = (c_1, \ldots, c_k)$ with $|c| = n$ is $\binom{n+k-1}{k-1}$.

Thus a simple enumeration will not suffice for larger instances.

Mass decomposition (4)

Using dynamic programming, we can solve the following problems efficiently (Σ: weighted alphabet, M: mass):

1. *Existence problem:* Decide whether a compomers c with mass(c) = M exists.
2. *One Witness problem:* Output a compomer c with mass(c) = M, if one exists.
3. *All witnesses problem:* Compute all compomers c with mass(c) = M.

Mass decomposition (5)

The dynamic programming algorithm is a variation of the classical algorithm originally introduced for the 'Coin Change Problem', originally due to Gilmore and Gomory.

Given a query mass M, a two-dimensional Boolean table B of size kM is constructed such that

$$B[i, m] = 1 \iff m \text{ is decomposable over } \{a_1, \ldots, a_i\}\,.$$

The table can be computed with the following recursion:

$$B[1, m] = 1 \iff m \mod a_1 = 0$$

and for $i > 0$,

$$B[i, m] = \begin{cases} B[i-1, m] & m < a_i\,, \\ B[i-1, m] \vee B[i, m - a_i] & \text{otherwise}\,. \end{cases}$$

The table is constructed up to mass M, and then a straight-forward backtracking algorithm computes all witnesses of M.

Mass decomposition (6)

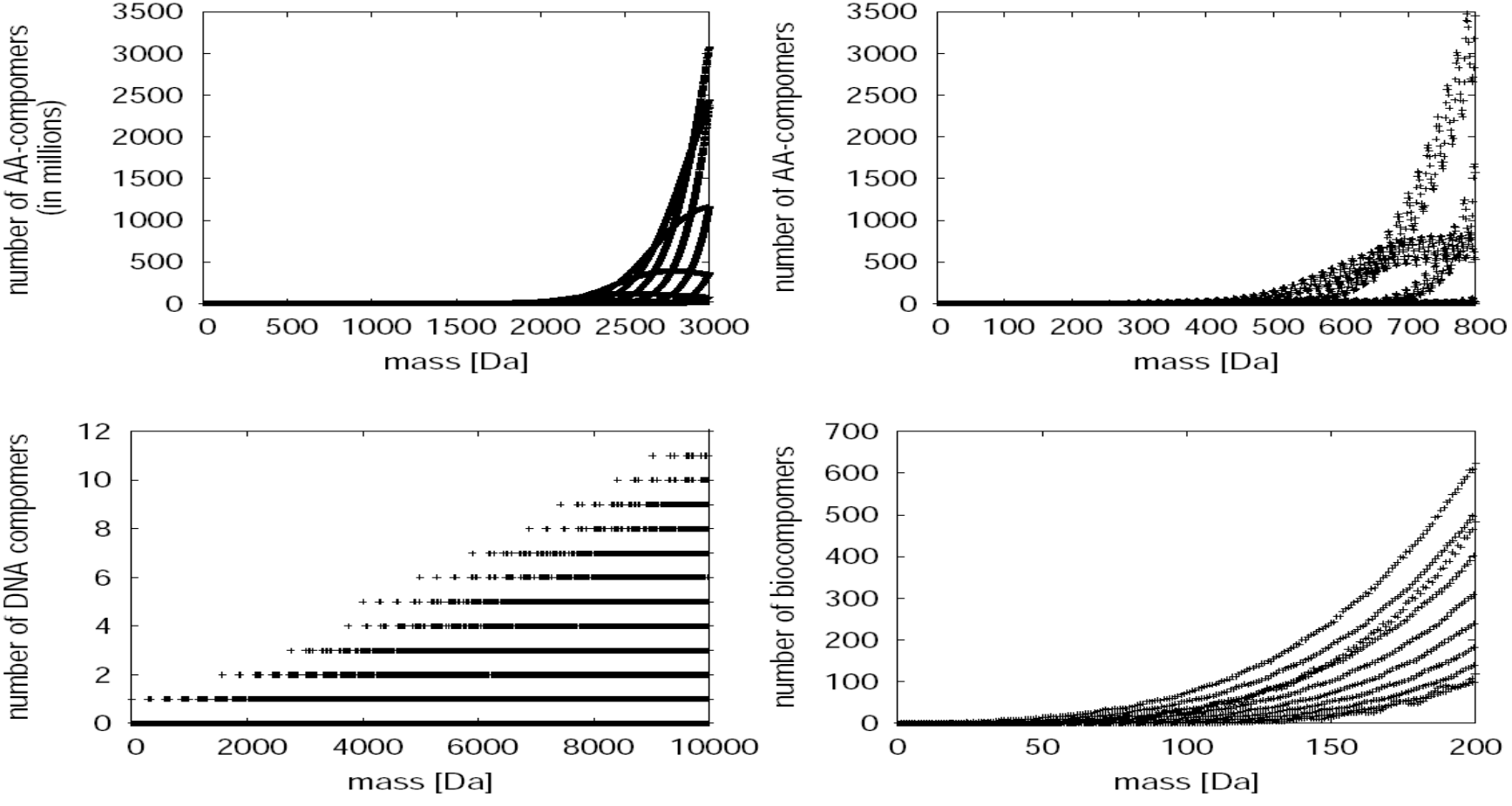
For the Existence and One Witness Problems, it suffices to construct a one-dimensional Boolean table A of size M, using the recursion $A[0] = 1$, $A[m] = 0$ for $1 \leq m < a_1$; and for $m \geq a_1$, $A[m] = 1$ if there exists an i with $1 \leq i \leq k$ such that $A[m - a_i] = 1$, and 0 otherwise. The construction time is $O(kM)$ and one witness c can be produced by backtracking in time proportional to $|c|$, which can be in the worst case $\frac{1}{a_1}M$. Of course, both of these problems can also be solved using the full table B.

A variant computes $\gamma(M)$, the number of decompositions of M, in the last row, where the entries are integers, using the recursion $C[i, m] = C[i - 1, m] + C[i, m - a_i]$.

The running time for solving the All Witnesses Problem is $O(kM)$ for the table construction, and $O(\gamma(M)\frac{1}{a_1}M)$ for the computation of the witnesses (where $\gamma(M)$ is the size of the output set), while storage space is ($O(kM)$.

Mass decomposition (7)

The number of compomers is $O(M^{\Sigma})$. (*Exercise*: why?) Depending on the mass resolution, the results can be useful for M up to, say, 1000 Da, but in general one has to take further criteria into account. (Figure from [BL].)



Mass decomposition (8)

Example output from http://bibiserv.techfak.uni-bielefeld.de/decomp/

```
# imsdecomp 1.3
# Copyright 2007,2008 Informatics for Mass Spectrometry group
#                     at Bielefeld University
#
# http://BiBiServ.TechFak.Uni-Bielefeld.DE/decomp/
#
# precision: 4e-05
# allowed error: 0.1 Da
# mass mode: mono
# modifiers: none
# fixed modifications: none
# variable modifications: none
# alphabet (character, mass, integer mass):
#       H  1.007825      25196
#       C        12     300000
#       N 14.003074     350077
#       O 15.994915     399873
#       P 30.973761     774344
#       S 31.972071     799302
# constraints: none
# chemical plausibility check: off
#
# Shown in parentheses after each decomposition:
# - actual mass
# - deviation from actual mass
```

```
#
# mass 218.03 has 1626 decompositions (showing the best 100):
H2 C6 N8 O2 (218.03007; +7.1384e-05)
H8 C7 N1 O7 (218.03008; +7.6606e-05)
H13 C2 N5 O1 P1 S2 (218.02991; -8.7014e-05)
H139 O1 P2 (218.03012; +0.000117068)
H19 C1 N1 O3 P3 S1 (218.02985; -0.000151322)
H16 N3 O4 S3 (218.03029; +0.000293122)
H5 C2 N9 O2 P1 (218.03038; +0.00038199)
H11 C3 N2 O7 P1 (218.03039; +0.000387212)
H10 C6 N4 O1 S2 (218.0296; -0.00039762)
H16 C5 O3 P2 S1 (218.02954; -0.000461928)
H16 C3 N3 P4 (218.02947; -0.000531458)
H21 C2 N1 P1 S4 (218.02944; -0.000556018)
H8 N7 O5 S1 (218.03076; +0.000762126)
H14 C1 O10 S1 (218.03077; +0.000767348)
H18 C5 O1 P4 (218.03081; +0.000811196)
H13 C7 N2 P3 (218.02916; -0.000842064)
H18 C6 S4 (218.02913; -0.000866624)
H12 C8 N1 O2 S2 (218.03095; +0.000945034)
H9 C1 N5 O6 P1 (218.02904; -0.000955442)
H3 N12 O1 P1 (218.02904; -0.000960664)
H21 C1 N1 O1 P5 (218.03112; +0.001121802)
H10 C11 N1 P2 (218.02885; -0.00115267)
H137 O3 S1 (218.02884; -0.001156056)
H15 C4 N2 O2 P1 S2 (218.03126; +0.00125564)
H6 C5 N4 O6 (218.02873; -0.001266048)
C4 N11 O1 (218.02873; -0.00127127)
H4 C8 N5 O3 (218.03141; +0.001414038)
H17 C1 N1 O5 P1 S2 (218.02858; -0.001424446)
```

H11 N8 P1 S2 (218.02857; -0.001429668)
H7 C15 P1 (218.02854; -0.001463276)
H135 C3 N1 S1 (218.03152; +0.00152403)
H134 C2 N2 P1 (218.02846; -0.001536192)
H18 N3 O2 P2 S2 (218.03157; +0.001566246)
H12 C1 N7 S3 (218.03163; +0.001630554)
H18 C2 O5 S3 (218.03164; +0.001635776)
H7 C4 N6 O3 P1 (218.03172; +0.001724644)
H14 C5 O5 S2 (218.02826; -0.001735052)
H8 C4 N7 S2 (218.02826; -0.001740274)
H14 C3 N3 O2 P2 S1 (218.0282; -0.001804582)
H131 C6 N1 (218.02815; -0.001846798)
H11 C12 P1 S1 (218.03191; +0.001907552)
H10 N7 O3 P2 (218.03204; +0.00203525)
H16 C1 O8 P2 (218.03204; +0.002040472)
H4 C1 N11 O1 S1 (218.0321; +0.002099558)
H10 C2 N4 O6 S1 (218.0321; +0.00210478)
H11 C7 N2 O2 P1 S1 (218.02788; -0.002115188)
H14 C8 N1 P2 S1 (218.03222; +0.002218158)
H13 N1 O10 P1 (218.02771; -0.002292874)
H8 C11 N1 O2 S1 (218.02757; -0.002425794)
H22 C3 S5 (218.0325; +0.002504204)
H17 C4 N2 P3 S1 (218.03253; +0.002528764)
H10 C4 O10 (218.0274; -0.00260348)
H4 C3 N7 O5 (218.02739; -0.002608702)
H6 C10 N2 O4 (218.03276; +0.002756692)
H20 N3 P4 S1 (218.03284; +0.00283937)
H20 C2 O3 P2 S2 (218.03291; +0.0029089)
H14 C3 N4 O1 S3 (218.03297; +0.002973208)
H9 C6 N3 O4 P1 (218.03307; +0.003067298)

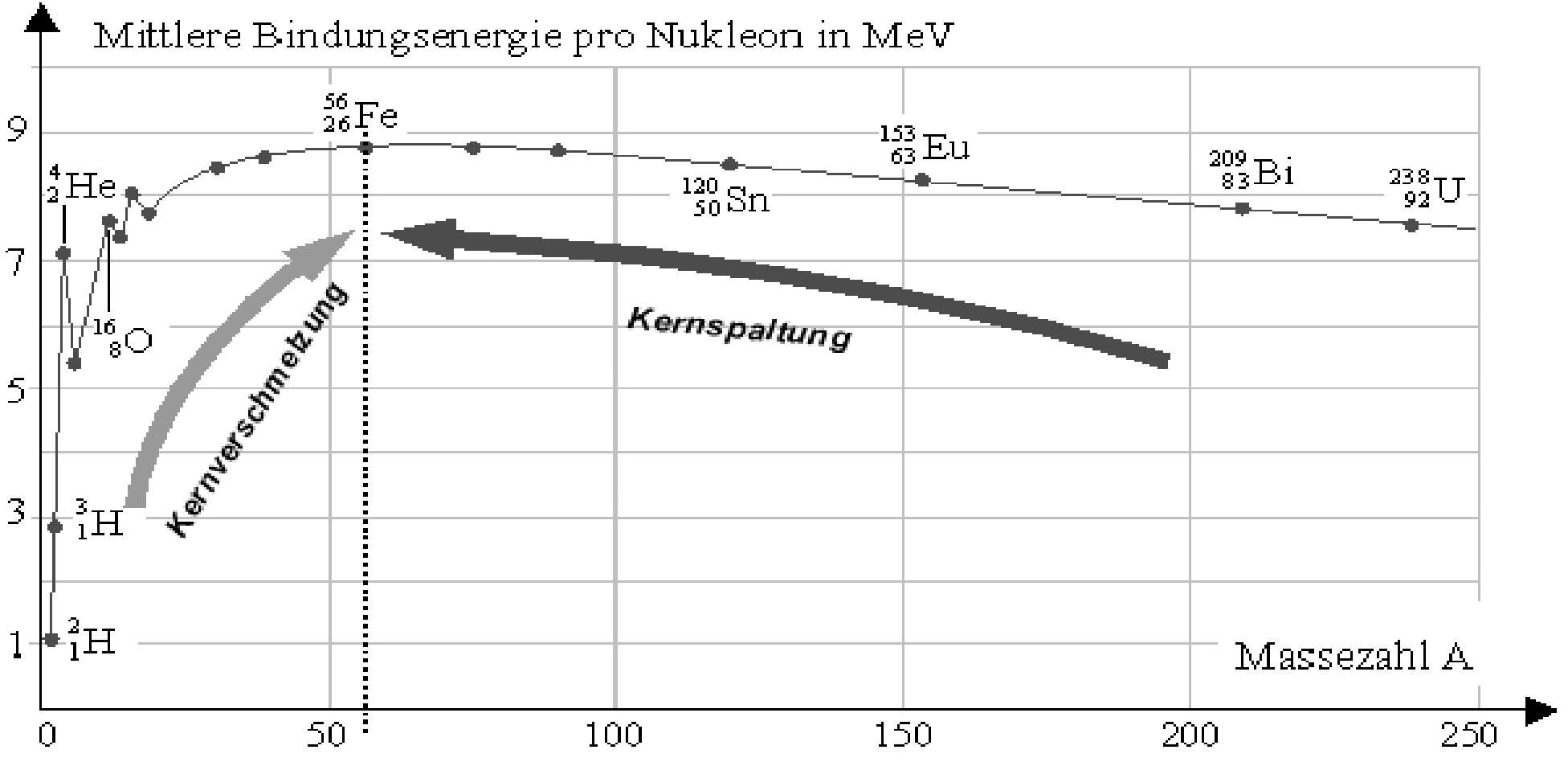
H12 C3 N3 O4 S2 (218.02692; -0.003077706)
H12 C1 N6 O1 P2 S1 (218.02685; -0.003147236)
H18 N2 O3 P4 (218.02679; -0.003211544)
H12 C2 N4 O4 P2 (218.03338; +0.003377904)
H6 C3 N8 O2 S1 (218.03344; +0.003442212)
H12 C4 N1 O7 S1 (218.03345; +0.003447434)
H9 C5 N5 O1 P1 S1 (218.02654; -0.003457842)
H15 C4 N1 O3 P3 (218.02648; -0.00352215)
H20 C1 N2 P2 S3 (218.02638; -0.00361624)
H15 N2 O7 P1 S1 (218.03376; +0.00375804)
H6 C9 N4 O1 S1 (218.02623; -0.003768448)
H12 C8 O3 P2 (218.02617; -0.003832756)
H17 C5 N1 P1 S3 (218.02607; -0.003926846)
H8 C2 N3 O9 (218.02605; -0.003946134)
H2 C1 N10 O4 (218.02605; -0.003951356)
H2 C11 N6 (218.03409; +0.004094124)
H22 C2 O1 P4 S1 (218.03418; +0.004182024)
H14 C9 S3 (218.02576; -0.004237452)
H16 C5 N1 O2 S3 (218.03432; +0.004315862)
H5 C7 N7 P1 (218.0344; +0.00440473)
H11 C8 O5 P1 (218.03441; +0.004409952)
H10 C1 N6 O3 S2 (218.02558; -0.00442036)
H16 N2 O5 P2 S1 (218.02552; -0.004484668)
H133 C3 O3 (218.02547; -0.004526884)
H19 C1 N2 O2 P1 S3 (218.03463; +0.004626468)
H8 C3 N8 P2 (218.03472; +0.004715336)
H14 C4 N1 O5 P2 (218.03472; +0.004720558)
H8 C5 N5 O3 S1 (218.03478; +0.004784866)
H13 C4 N1 O5 P1 S1 (218.0252; -0.004795274)
H7 C3 N8 P1 S1 (218.0252; -0.004800496)

```
H13 C2 N4 O2 P3 (218.02514; -0.004864804)
H18 C1 N2 O2 S4 (218.02511; -0.004889364)
H139 N1 S2 (218.03489; +0.004894858)
H17 N2 O5 P3 (218.03503; +0.005031164)
H11 C1 N6 O3 P1 S1 (218.0351; +0.005095472)
H10 C8 O5 S1 (218.02489; -0.00510588)
H4 C7 N7 S1 (218.02489; -0.005111102)
H10 C6 N3 O2 P2 (218.02482; -0.00517541)
H15 C9 P1 S2 (218.03528; +0.00527838)
H6 N6 O8 (218.02471; -0.005288788)
H21 C2 O1 P3 S2 (218.02467; -0.005333808)
H131 N5 O1 (218.03536; +0.005363862)

# done
```

Mass defect

The difference between the actual atomic mass of an isotope and the nearest integral mass is called the *mass defect*. The size of the mass defect varies over the Periodic Table. The mass defect is due to the binding energy of the nucleus:



http://de.wikipedia.org/w/index.php?title=Datei:Bindungsenergie_massenzahl.jpg

Mass defect (2)

The mass differences of light and heavy isotopes are also not exactly multiples of the atomic mass unit. We have

$$\begin{aligned}
\text{mass}\,(^{2}H) - \text{mass}\,(^{1}H) &= 1.00628 \doteq 1 \\
\text{mass}\,(^{13}C) - \text{mass}\,(^{12}C) &= 1.003355 \doteq 1 \\
\text{mass}\,(^{18}O) - \text{mass}\,(^{16}O) &= 2.004244 \doteq 2 \\
\text{mass}\,(^{15}N) - \text{mass}\,(^{14}N) &= 0.997035 \doteq 1 \\
\text{mass}\,(^{34}S) - \text{mass}\,(^{32}S) &= 1.995796 \doteq 2
\end{aligned}$$

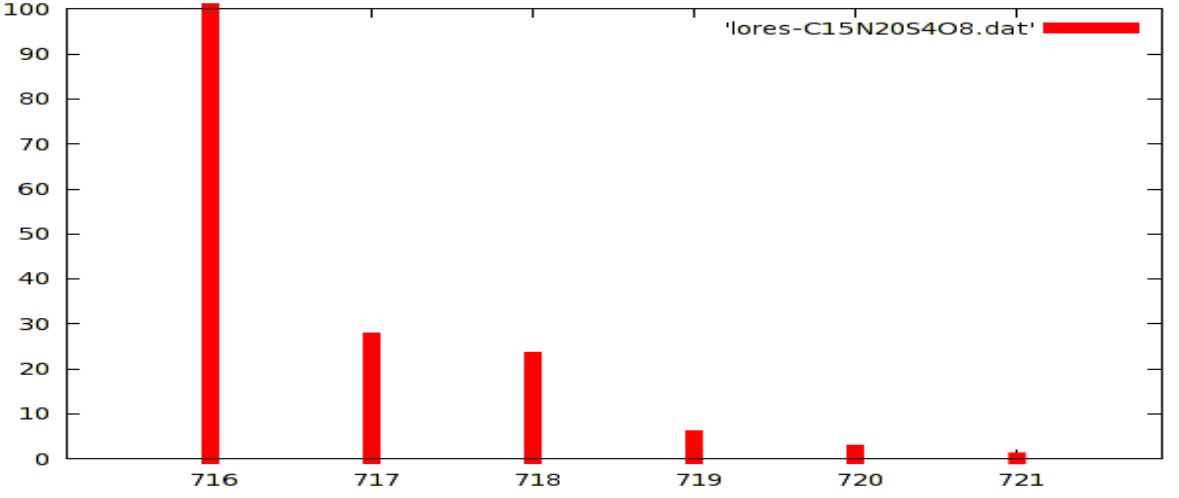
These differences (due to the mass defect) are subtle but become perceptible with very high resolution mass spectrometers. (*Exercise:* About which resolution is necessary?) This is currently an active field of research.

Mass defect (3)

High resolution and low resolution isotopic spectra for C15N20S4O8.

hires.dat		lores.dat	
715.909120	100.0	716	100.0
716.906159	7.2	717	26.8
716.908509	3.2		
716.912479	16.1		
716.913329	0.2		
717.903189	0.2	718	22.6
717.904919	17.6		
717.905549	0.2		
717.909520	1.1		
717.911870	0.5		
717.913360	1.6		
717.915830	1.1		
718.901960	1.3	719	5.0
718.904310	0.4		
718.908279	2.8		
718.910399	0.1		
718.916720	0.2		
719.900719	1.1	720	1.8
719.905319	0.2		
719.909160	0.2		
719.911629	0.2		
720.904080	0.2	721	0.2

Isotopic pattern:



Zoom on ”+2” mass peak (≈718):