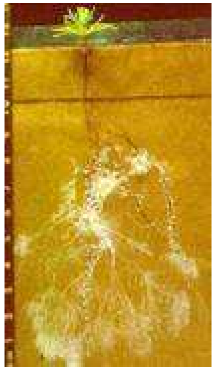


Novel method for estimating isotope incorporation using the ‘half-decimal place rule‘

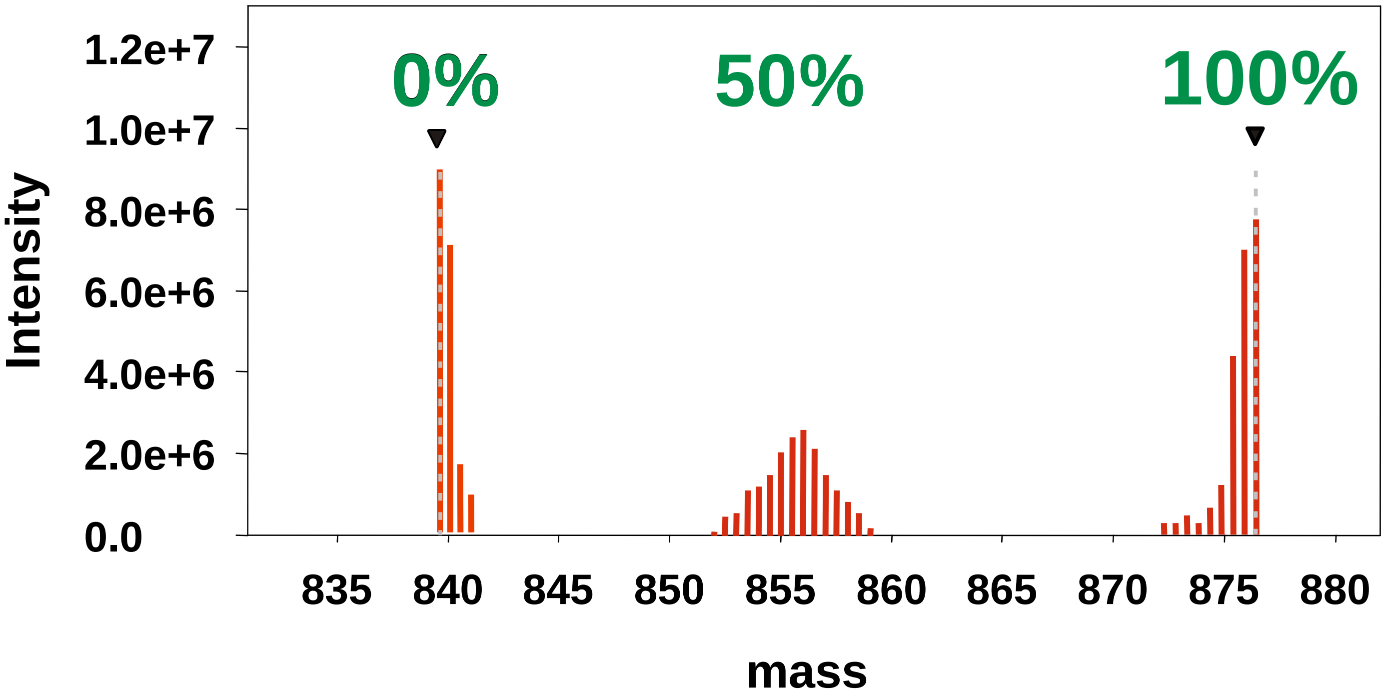
Ingo Fetzer

Department of Environmental Microbiology

userR Conference 2009, Rennes

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Problem



- Substrate fluxes in Procaryotes
- Function → Activity → Identitiy
- Interactions: Competition ↔ Mutualism

HELMHOLTZ
CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ

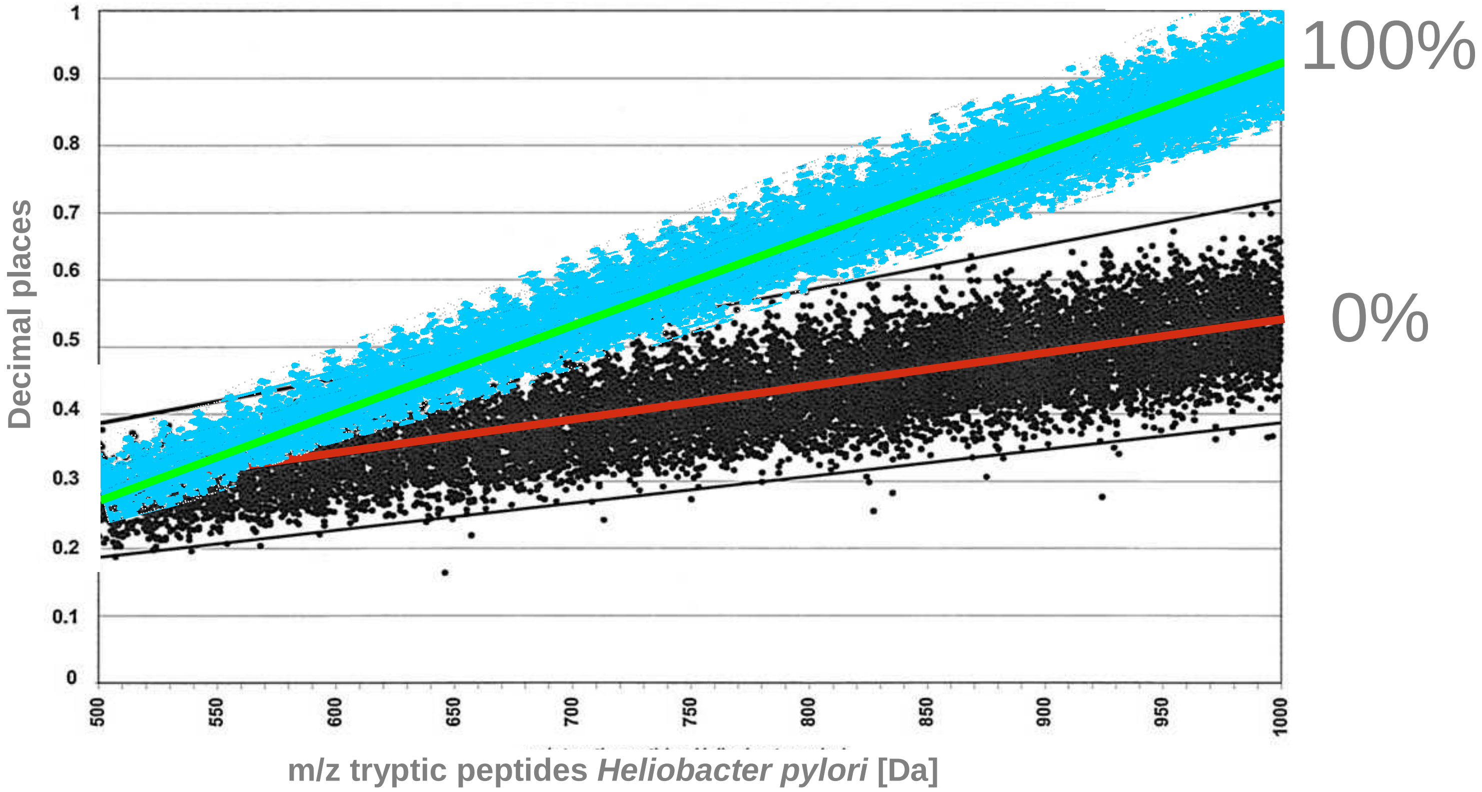
Goal

Develop an algorithm for estimating ^{13}C incorporation by using ‘half decimal place rule’

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

‘Half-decimal place rule’ (HDPR)

Mann (1995)



Schmidt et al 2003

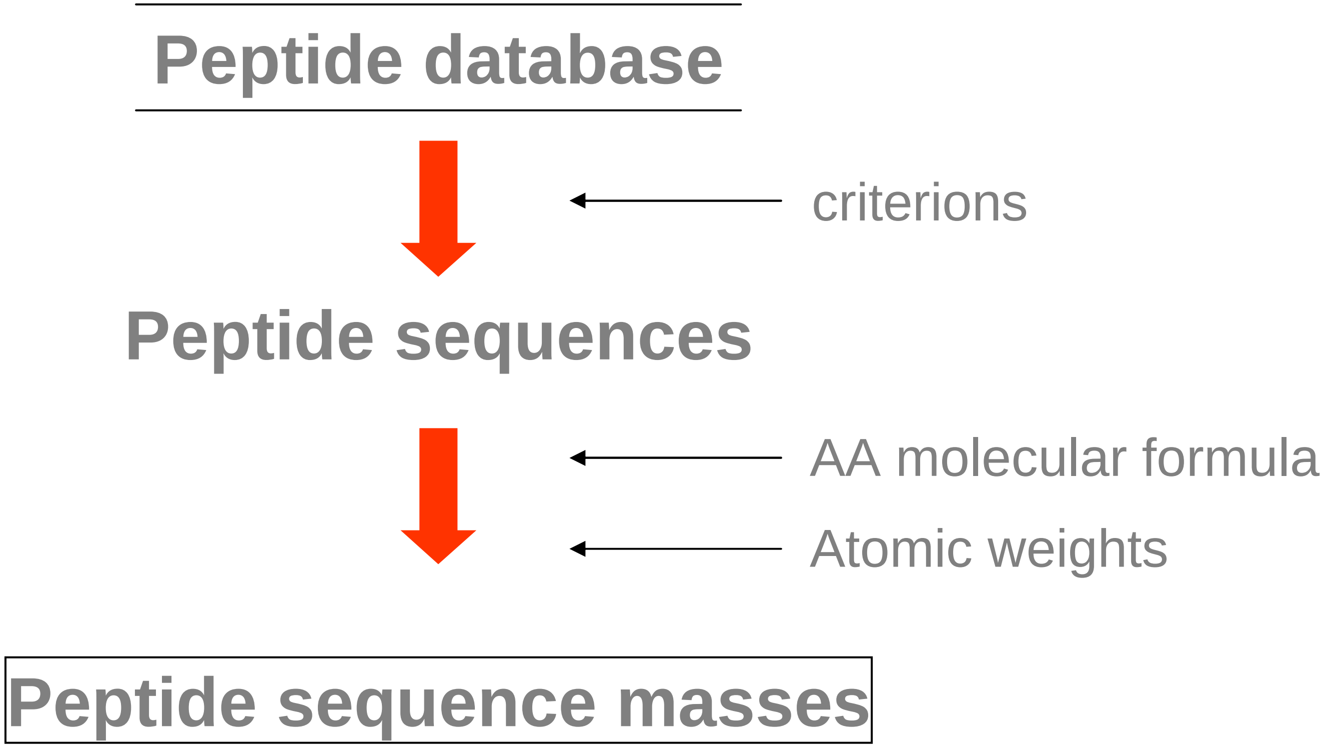
Outline

1. Peptide mass calculation for ^{12}C and ^{13}C
2. Estimation of ^{12}C and ^{13}C slopes (HDPR)
3. Estimation of relative ^{13}C incorporation rates
   (of user data)

implemented in ‘R‘
(R-project.org)

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Flowchart Script 1



Peptide database

criterions

Peptide sequences

AA molecular formula

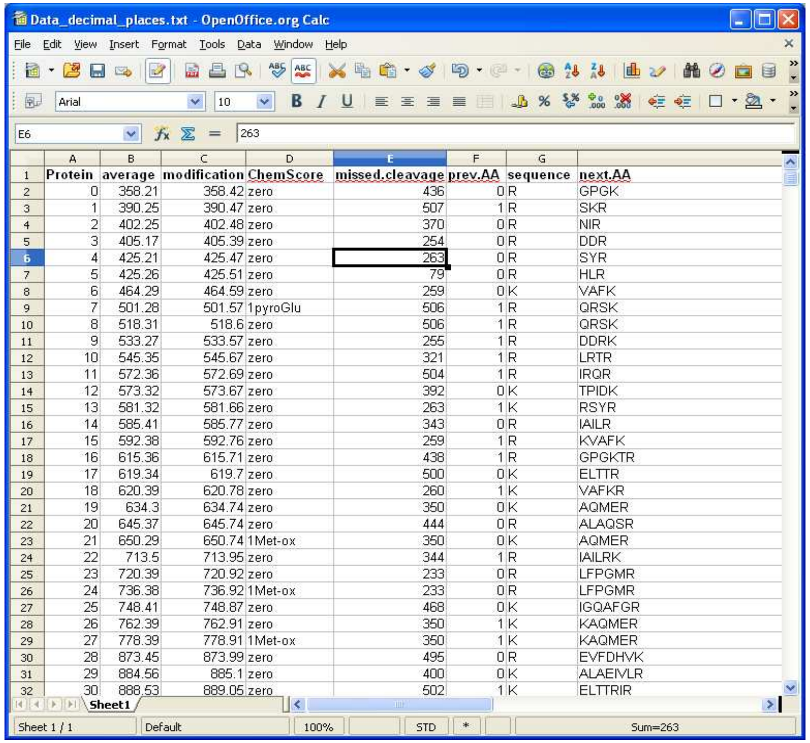
Atomic weights

Peptide sequence masses

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Peptide mass calculation for ^{12}C and ^{13}C

dataset *M. tuberculosis* H37Rv



	A	B	C	D	E	F	G	
1	**Protein**	**average**	**modification**	**ChemScore**	**missed.cleavage**	**prev.AA**	**sequence**	**next.AA**
2	0	358.21	358.42	zero	436	0	R	GPGK
3	1	390.25	390.47	zero	507	1	R	SKR
4	2	402.25	402.48	zero	370	0	R	NIR
5	3	405.17	405.39	zero	254	0	R	DDR
6	4	425.21	425.47	zero	263	0	R	SYR
7	5	425.26	425.51	zero	79	0	R	HLR
8	6	464.29	464.59	zero	259	0	K	VAFK
9	7	501.28	501.57	1pyroGlu	506	1	R	QRSK
10	8	518.31	518.6	zero	506	1	R	QRSK
11	9	533.27	533.57	zero	255	1	R	DDRK
12	10	545.35	545.67	zero	321	1	R	LRTR
13	11	572.36	572.69	zero	504	1	R	IRQR
14	12	573.32	573.67	zero	392	0	K	TPIDK
15	13	581.32	581.66	zero	263	1	K	RSYR
16	14	585.41	585.77	zero	343	0	R	IAILR
17	15	592.38	592.76	zero	259	1	R	KVAFK
18	16	615.36	615.71	zero	438	1	R	GPGKTR
19	17	619.34	619.7	zero	500	0	K	ELTTR
20	18	620.39	620.78	zero	260	1	K	VAFKR
21	19	634.3	634.74	zero	350	0	K	AQMER
22	20	645.37	645.74	zero	444	0	R	ALAQSR
23	21	650.29	650.74	1Met-ox	350	0	K	AQMER
24	22	713.5	713.95	zero	344	1	R	IAILRK
25	23	720.39	720.92	zero	233	0	R	LFPGMR
26	24	736.38	736.92	1Met-ox	233	0	R	LFPGMR
27	25	748.41	748.87	zero	468	0	K	IGQAFGR
28	26	762.39	762.91	zero	350	1	K	KAQMER
29	27	778.39	778.91	1Met-ox	350	1	K	KAQMER
30	28	873.45	873.99	zero	495	0	R	EVFDHVK
31	29	884.56	885.1	zero	400	0	K	ALAEIVLR
32	30	888.53	889.05	zero	502	1	K	ELTTRIR

Sanger Institute
(ftp://ftp.sanger.ac.uk/pub/tb/sequences/TB.pep)

Virtual digestion with MS-Digest

amino acid sequences
length 2 – 40

315,579 peptide fragments

ChemScore ≥ 10

Missing cleavage = 0

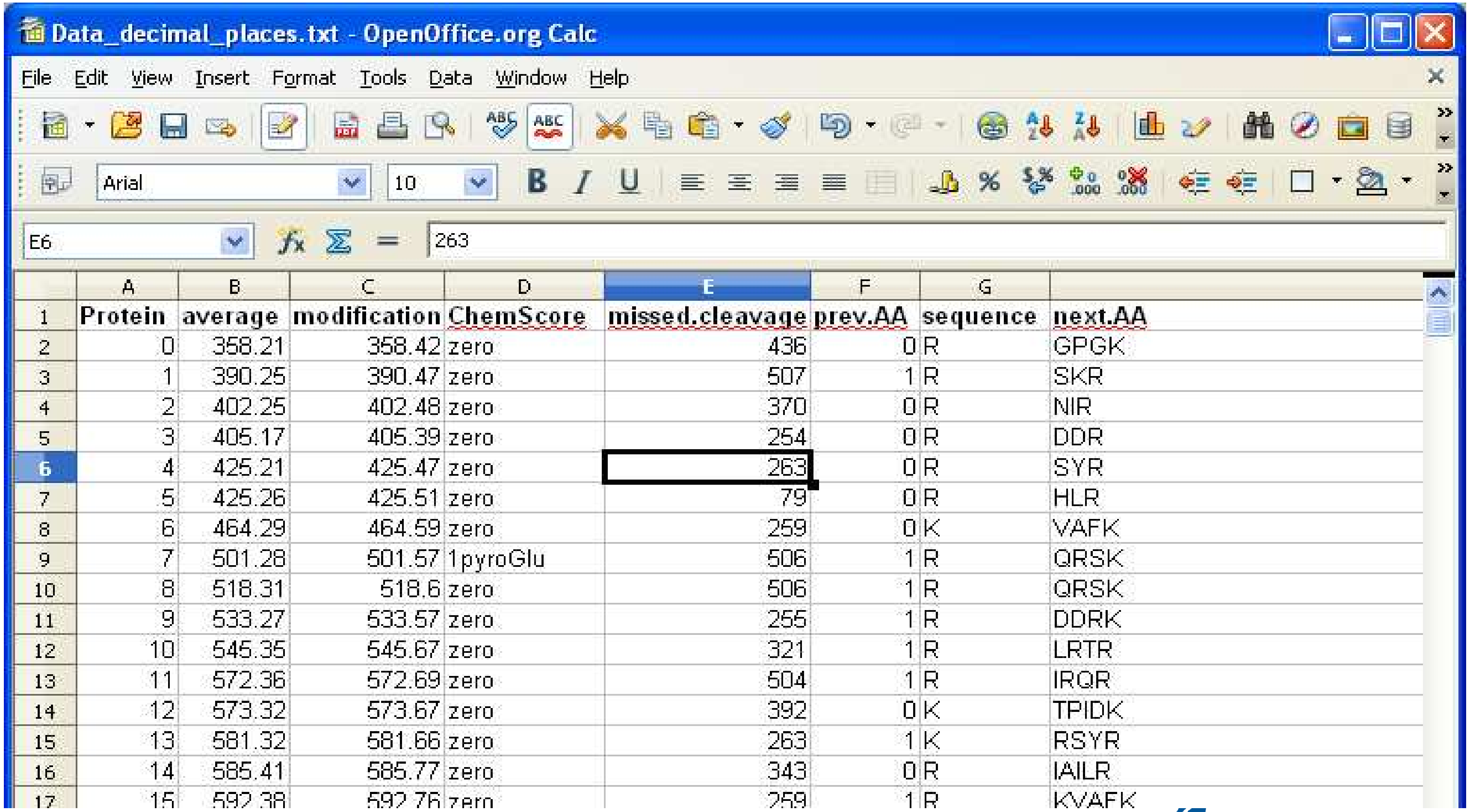
Modifications = Null

Mol. weight ≤ 5000 Da

90,637 peptide sequences

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Peptide mass calculation for ^{12}C and ^{13}C



E6 = 263

	A	B	C	D	E	F	G	
1	Protein	average	modification	ChemScore	missed.cleavage	prev.AA	sequence	next.AA
2	0	358.21	358.42	zero	436	0	R	GPGK
3	1	390.25	390.47	zero	507	1	R	SKR
4	2	402.25	402.48	zero	370	0	R	NIR
5	3	405.17	405.39	zero	254	0	R	DDR
6	4	425.21	425.47	zero	263	0	R	SYR
7	5	425.26	425.51	zero	79	0	R	HLR
8	6	464.29	464.59	zero	259	0	K	VAFK
9	7	501.28	501.57	1pyroGlu	506	1	R	QRSK
10	8	518.31	518.6	zero	506	1	R	QRSK
11	9	533.27	533.57	zero	255	1	R	DDRK
12	10	545.35	545.67	zero	321	1	R	LRTR
13	11	572.36	572.69	zero	504	1	R	IRQR
14	12	573.32	573.67	zero	392	0	K	TPIDK
15	13	581.32	581.66	zero	263	1	K	RSYR
16	14	585.41	585.77	zero	343	0	R	IAILR
17	15	592.38	592.76	zero	259	1	R	KVAFK

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Peptide mass calculation for ^{12}C and ^{13}C

Script 1:

1. Reduction of dataset (ChemScore, Modification etc.)

315,579 → 90,637

2 a. Peptide mass calculation

Sequence in DB + Molecular formula of AA → Sum of C, H, N, O for each sequence

GAG

$G=C_2H_6NO_2$ $A=C_3H_8NO_2$

$C_7H_{20}N_3O_6$

Why? Calculation of percentage ^{13}C incorporation

Peptide mass calculation for ^{12}C and ^{13}C

2b. Peptide mass calculation

Sum of

C, H, N, O of each sequence

$C_7H_{20}N_3O_6$

+ Atomic weights

^{12}C = 12.000000 Da
^{13}C = 13.003355 Da
N = 14.003074 Da
O = 15.994915 Da
H = 1.007825 Da

→ Molecular weights of sequences (with decimal residuals)

^{12}C=242.135212 Da
^{13}C=249.158697 Da

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Peptide mass calculation for ^{12}C and ^{13}C

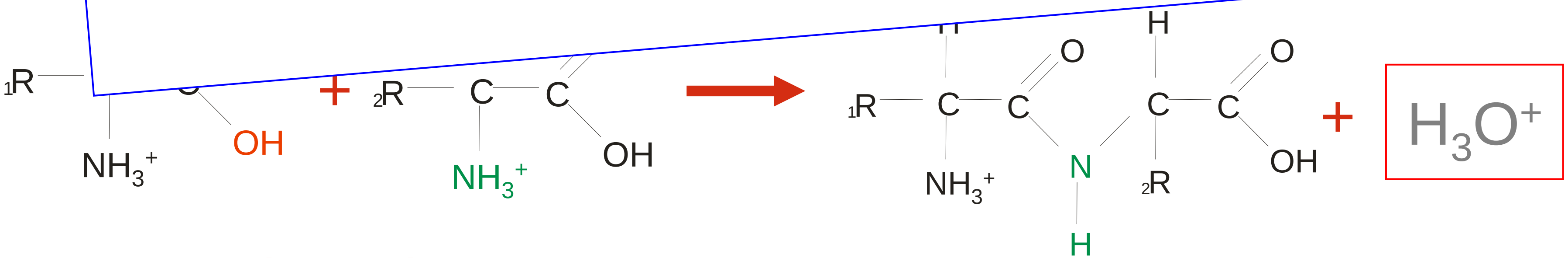
2a. Peptide mass calculation

Count sum of

C

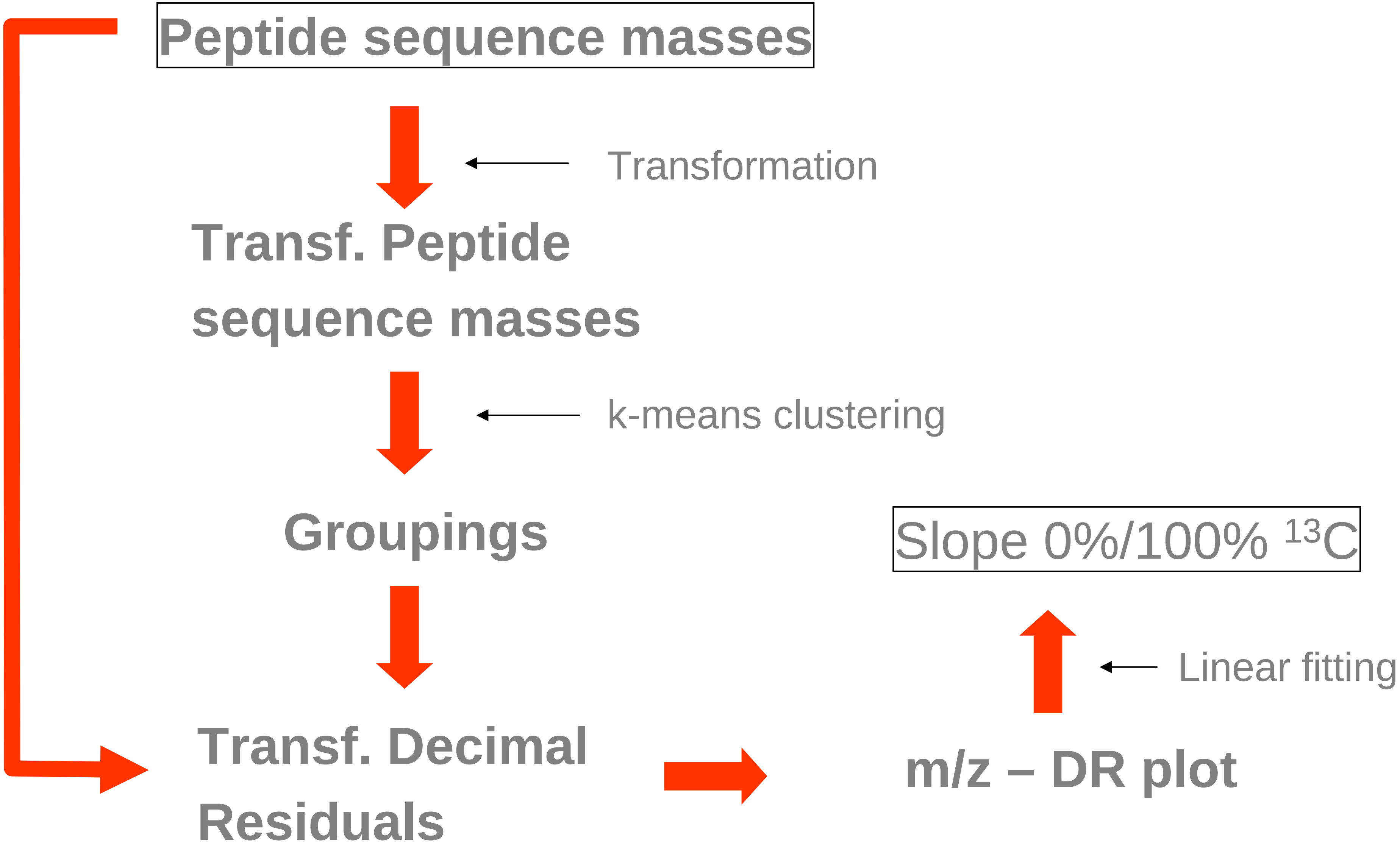
e

Molecular Masses of ^{12}C and ^{13}C peptides



$M_W(H_3O^+) = 19.01839$ Da

Flowchart Script 2



Peptide sequence masses

Transformation

Transf. Peptide sequence masses

k-means clustering

Groupings

Transf. Decimal Residuals

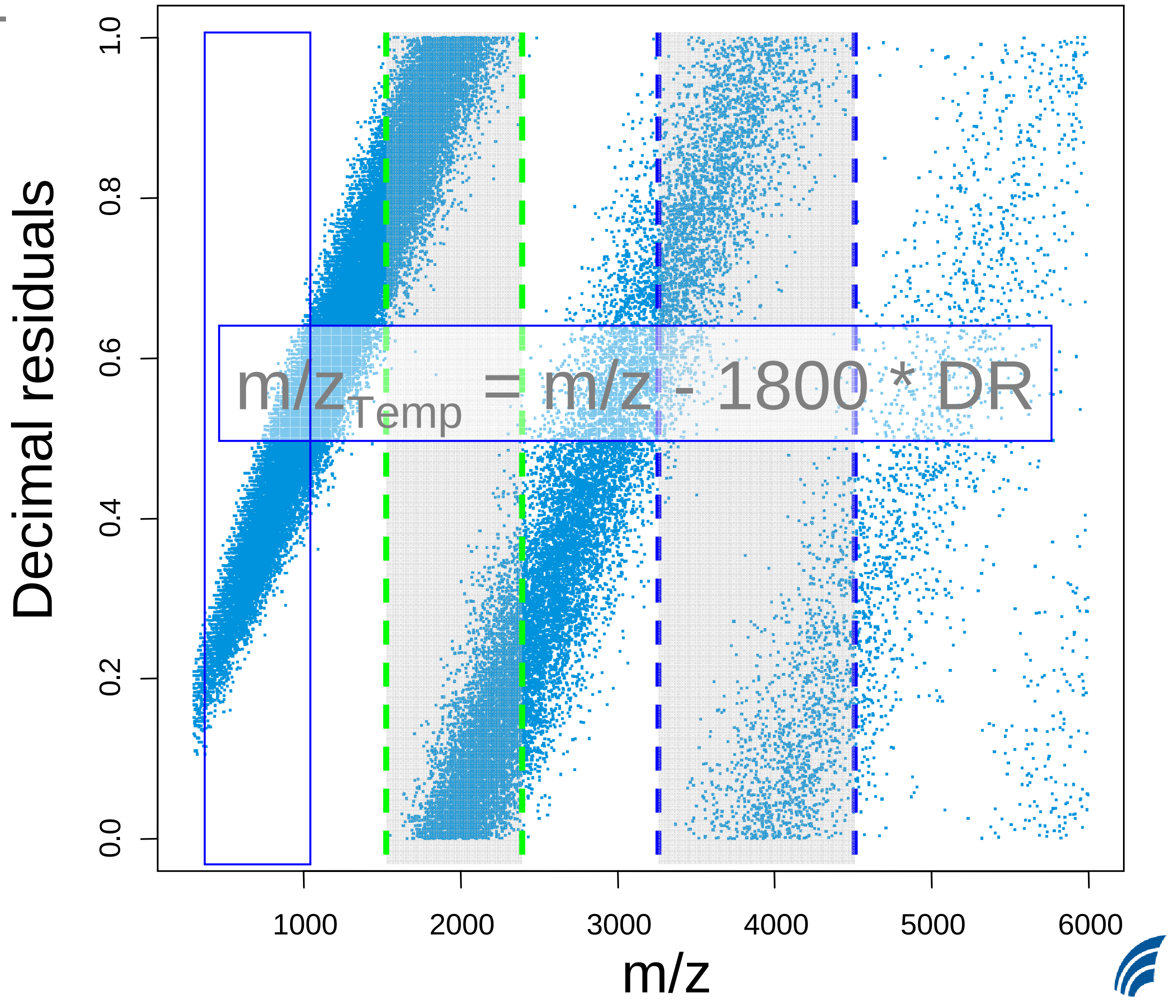
m/z – DR plot

Linear fitting

Slope 0%/100% ^{13}C

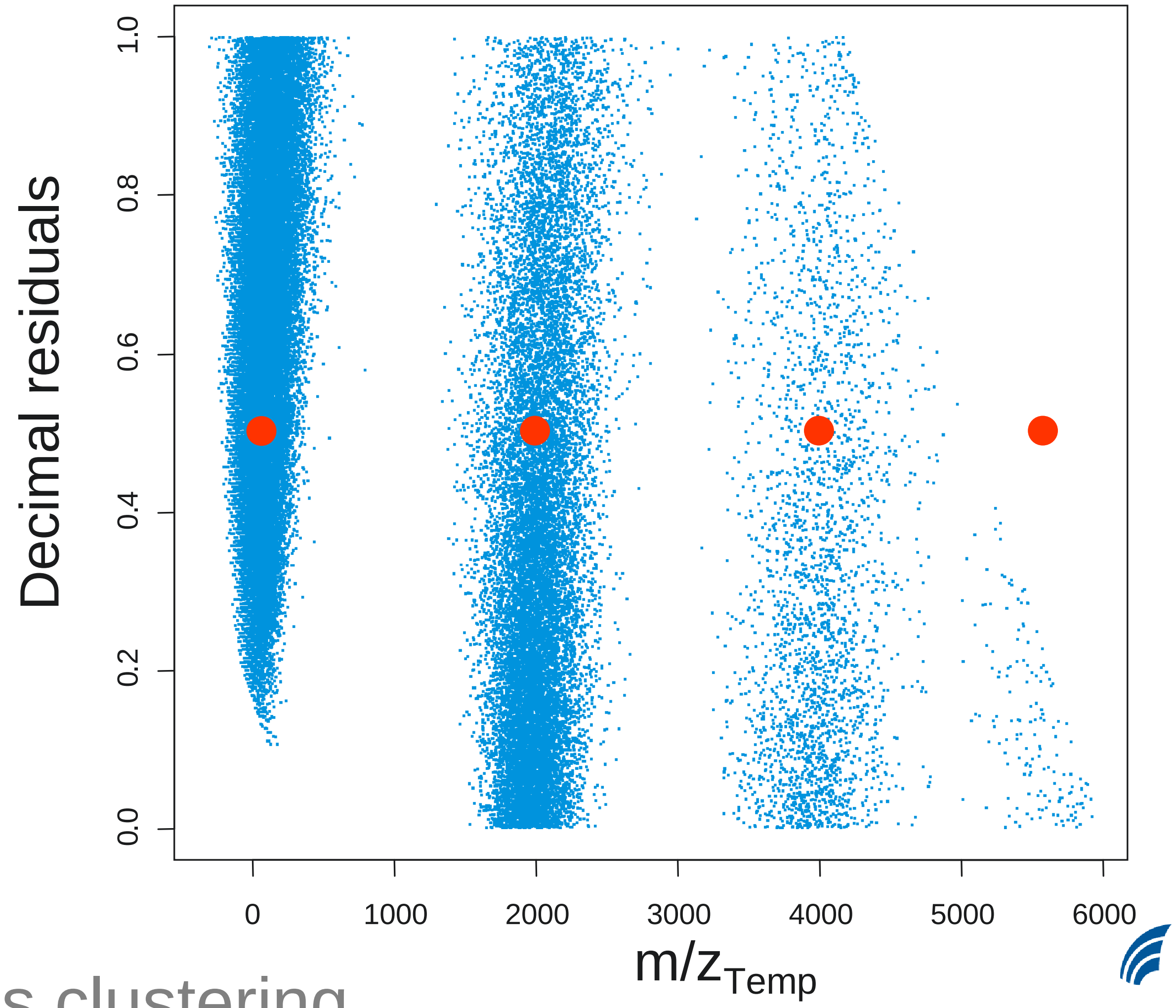
Estimation of ^{12}C and ^{13}C slopes

Script 2:



$m/z_{Temp} = m/z - 1800 * DR$

Estimation of ^{12}C and ^{13}C slopes

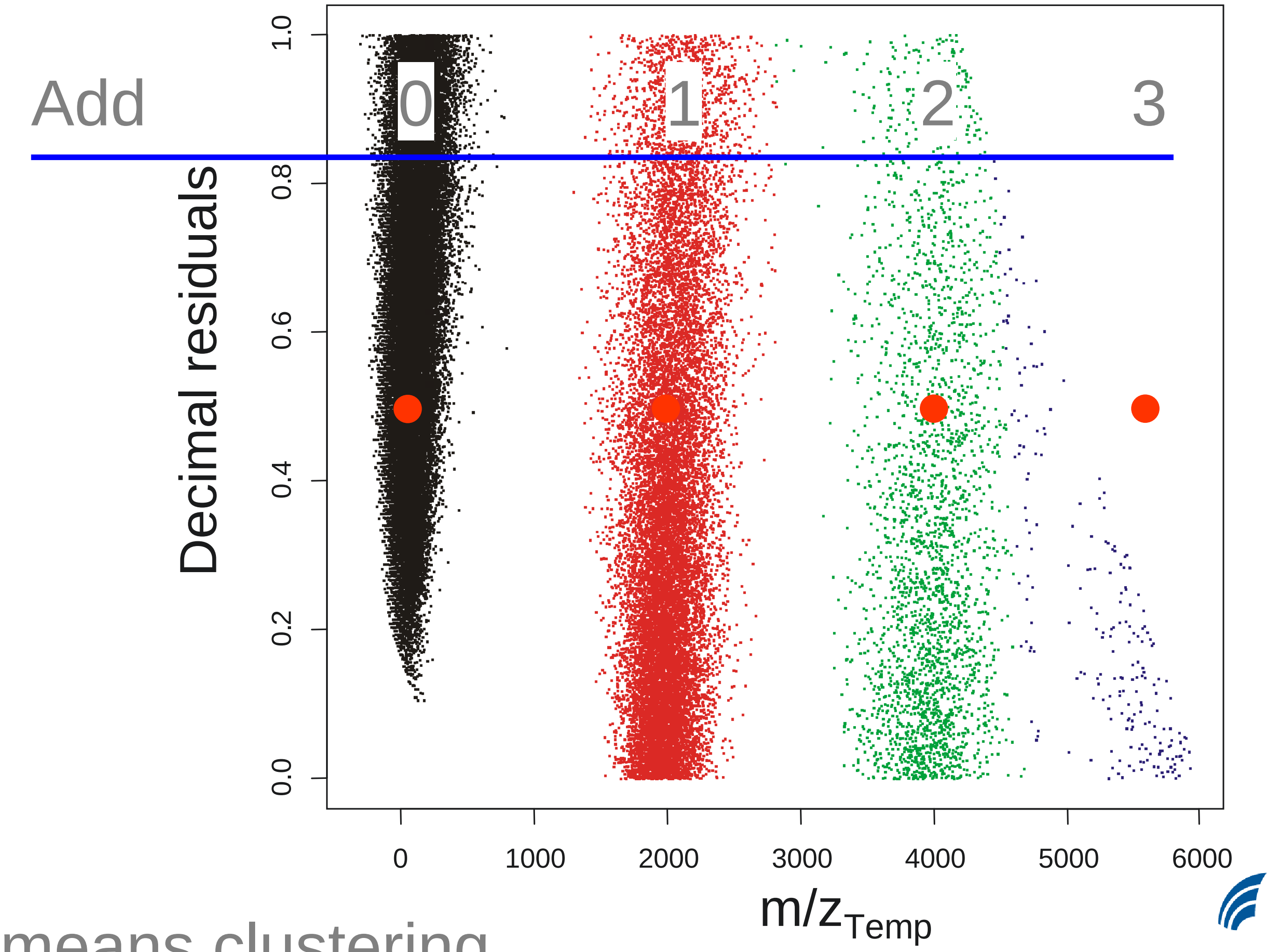


Hartigan & Wong (1979)

k-means clustering

using kmeans()

Estimation of ^{12}C and ^{13}C slopes

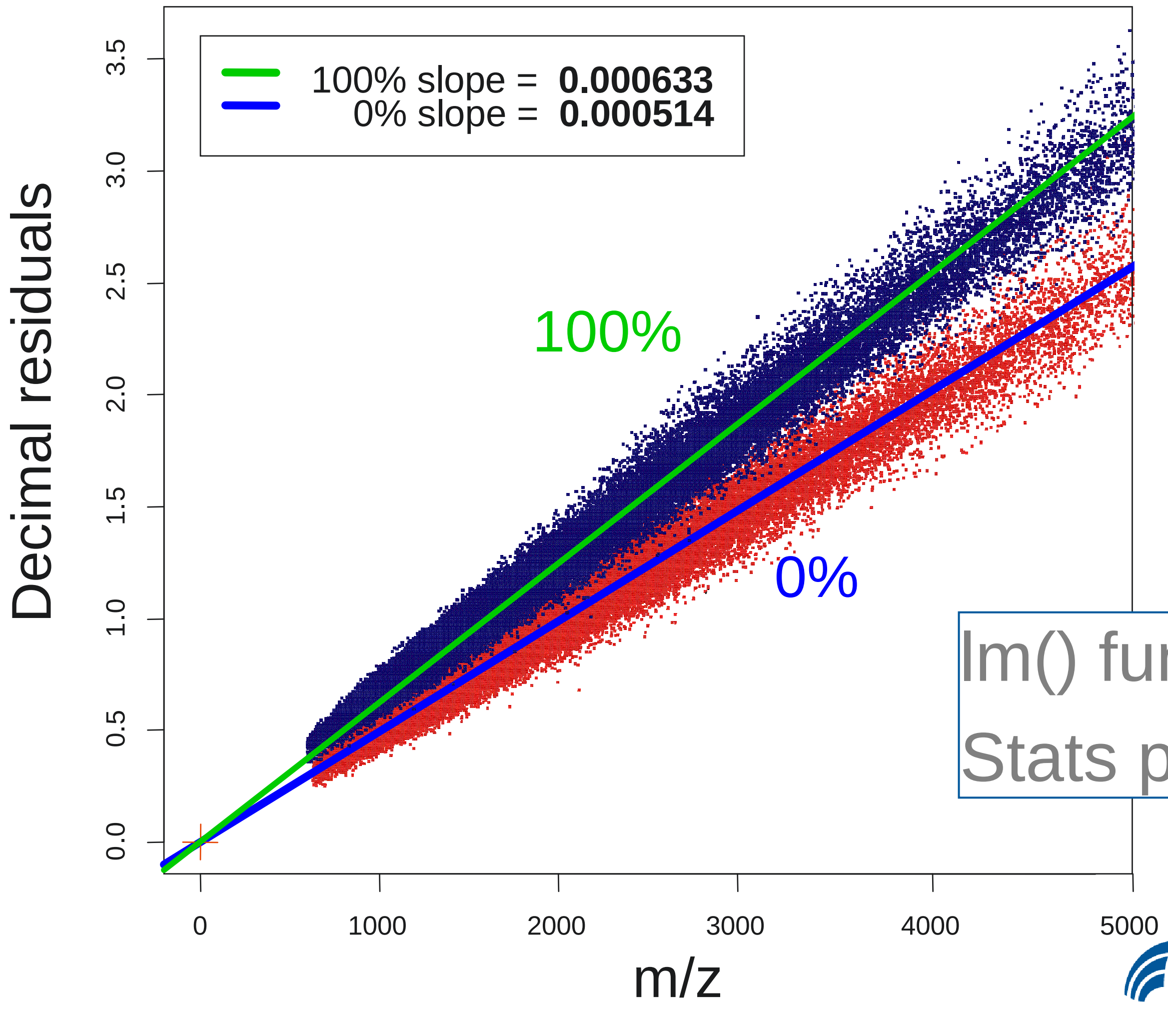


Hartigan & Wong (1979)

k-means clustering

using kmeans()

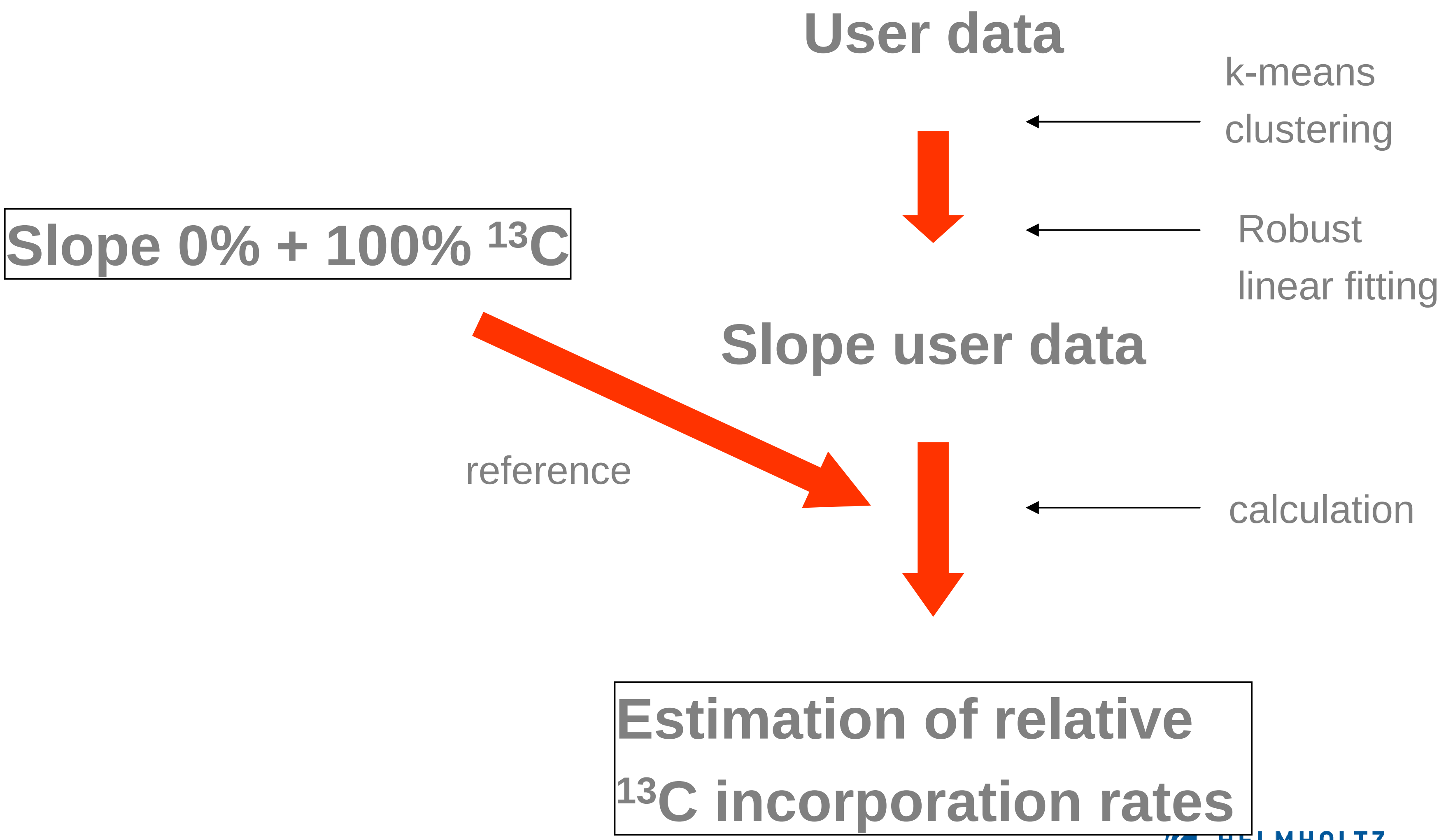
Estimation of ^{12}C and ^{13}C slopes



lm() function

Stats package

Flowchart Script 3



User data

k-means clustering

Robust linear fitting

Slope user data

Slope 0% + 100% ^{13}C

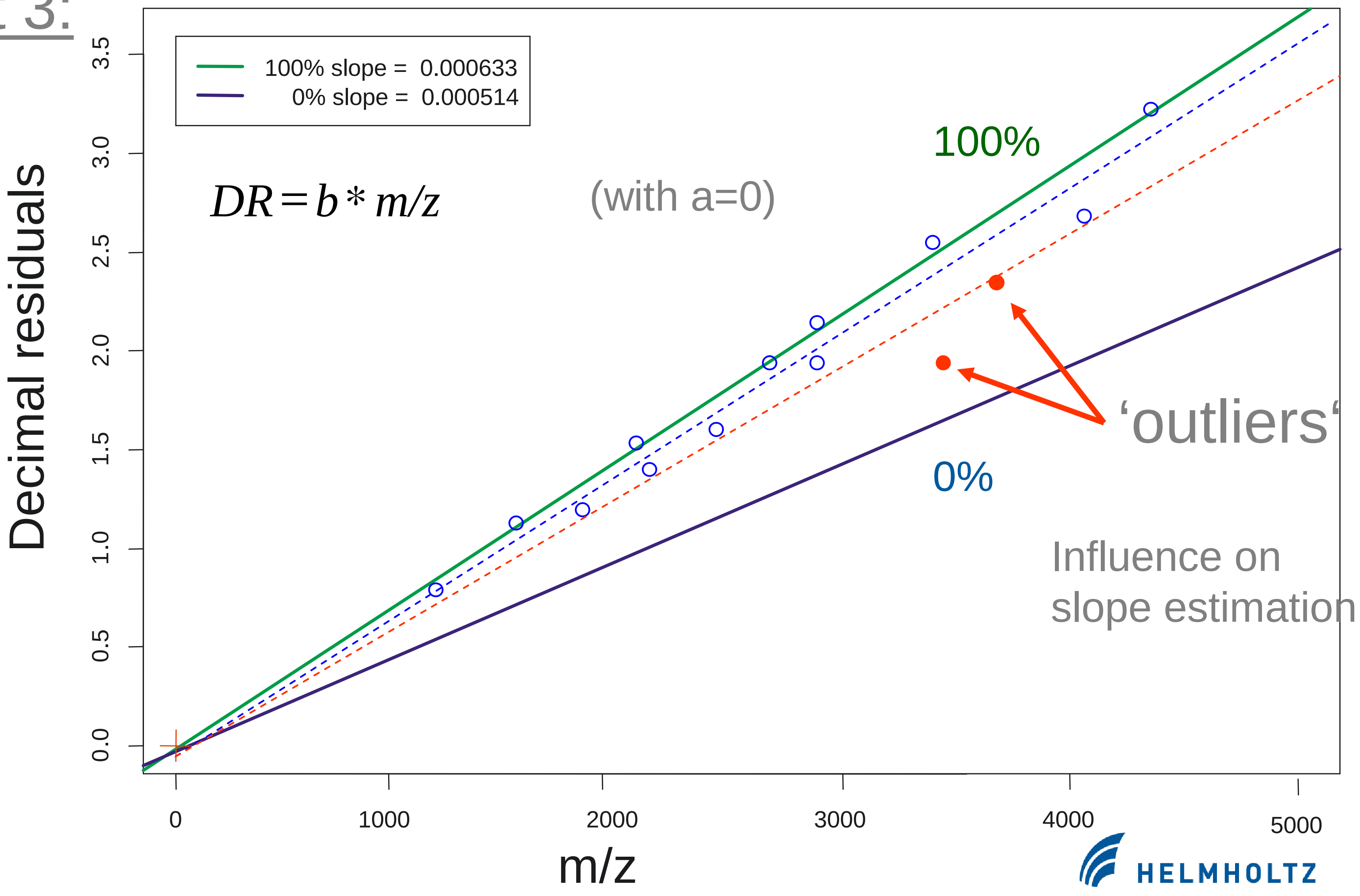
reference

calculation

Estimation of relative ^{13}C incorporation rates

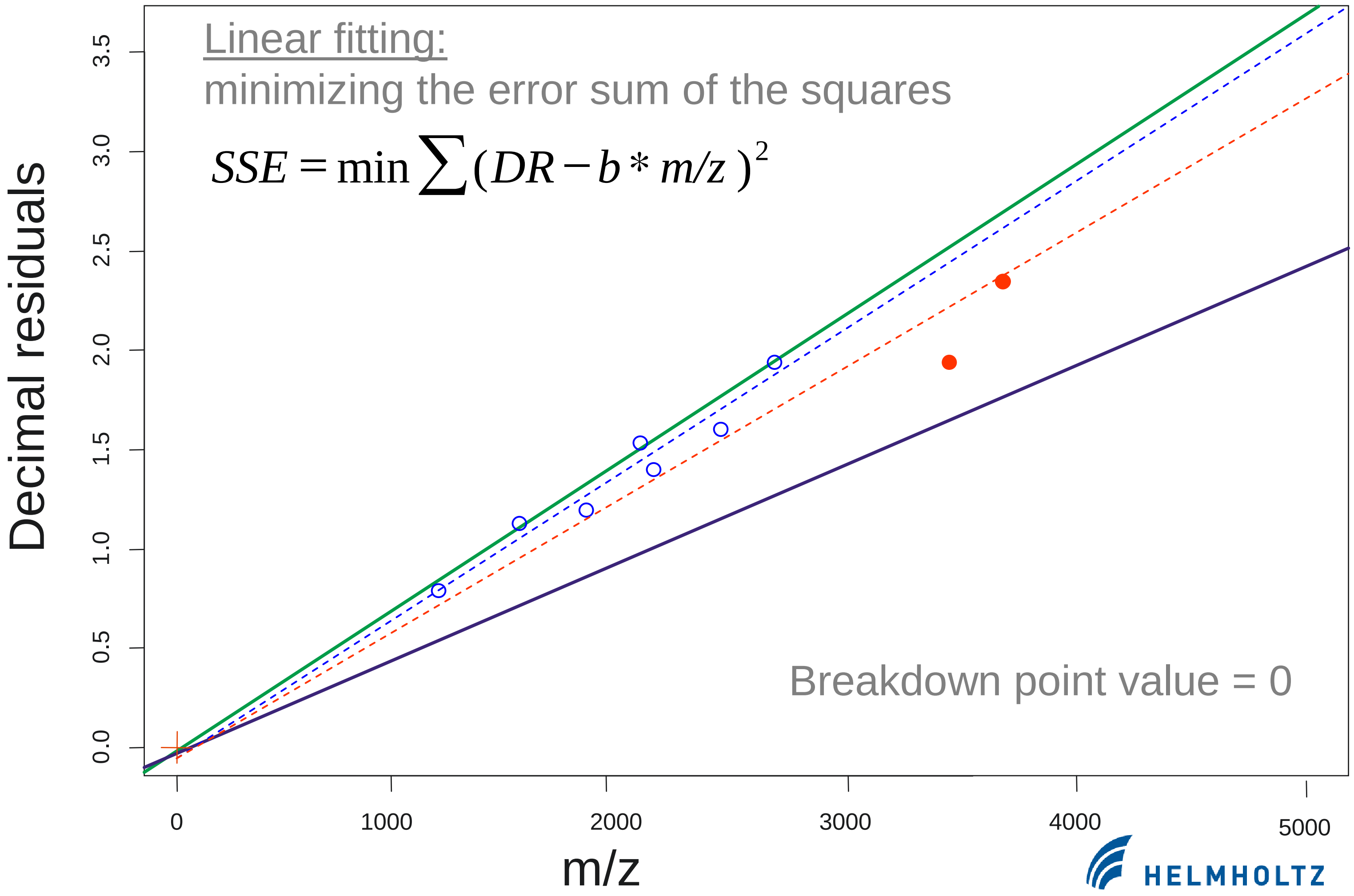
Estimation of relative ^{13}C incorporation rates

Script 3:



HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Estimation of relative ^{13}C incorporation rates



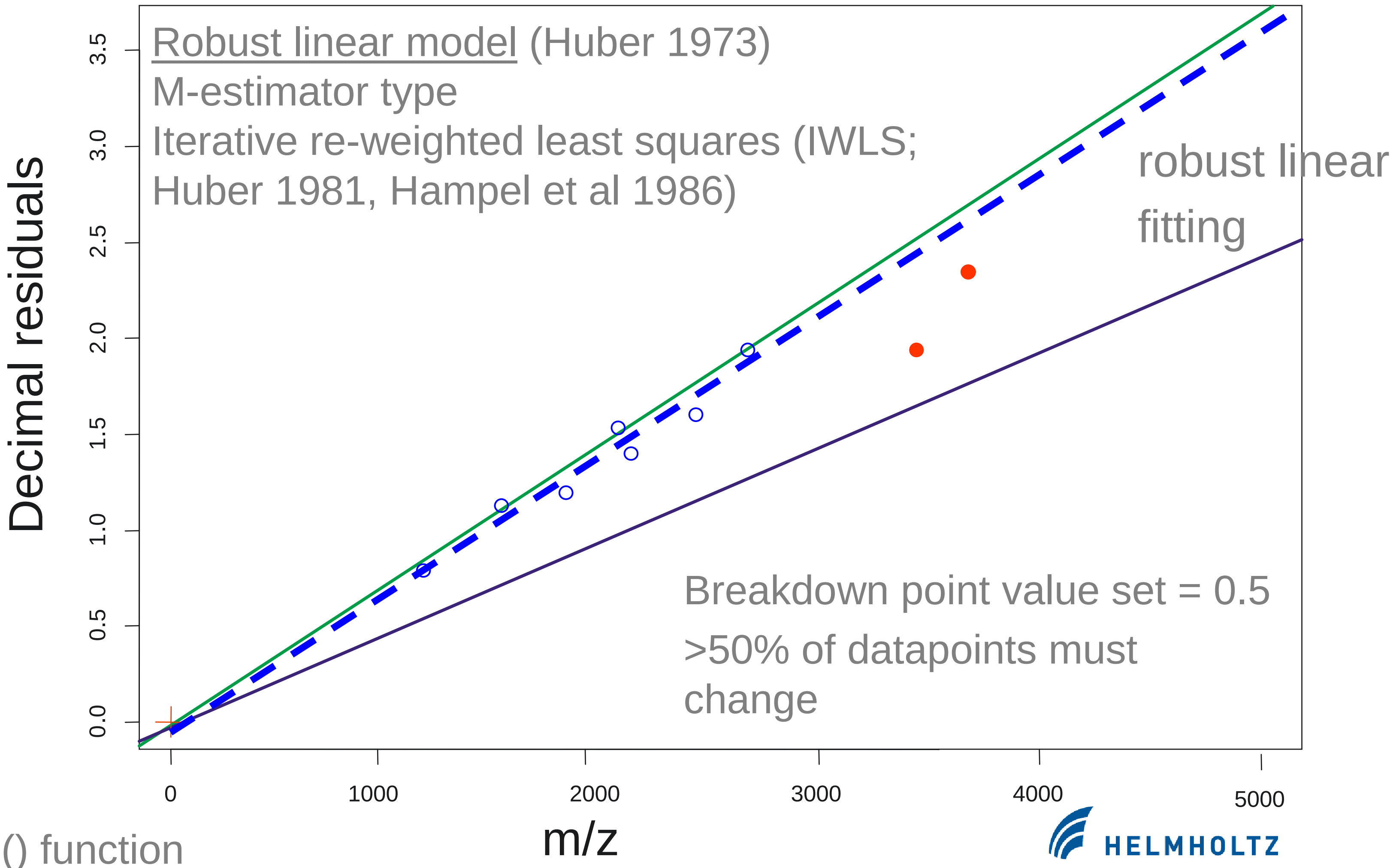
Linear fitting:

minimizing the error sum of the squares

$$SSE = \min \sum (DR - b * m/z)^2$$

Breakdown point value = 0

Estimation of relative ^{13}C incorporation rates



Robust linear model (Huber 1973)
M-estimator type
Iterative re-weighted least squares (IWLS; Huber 1981, Hampel et al 1986)

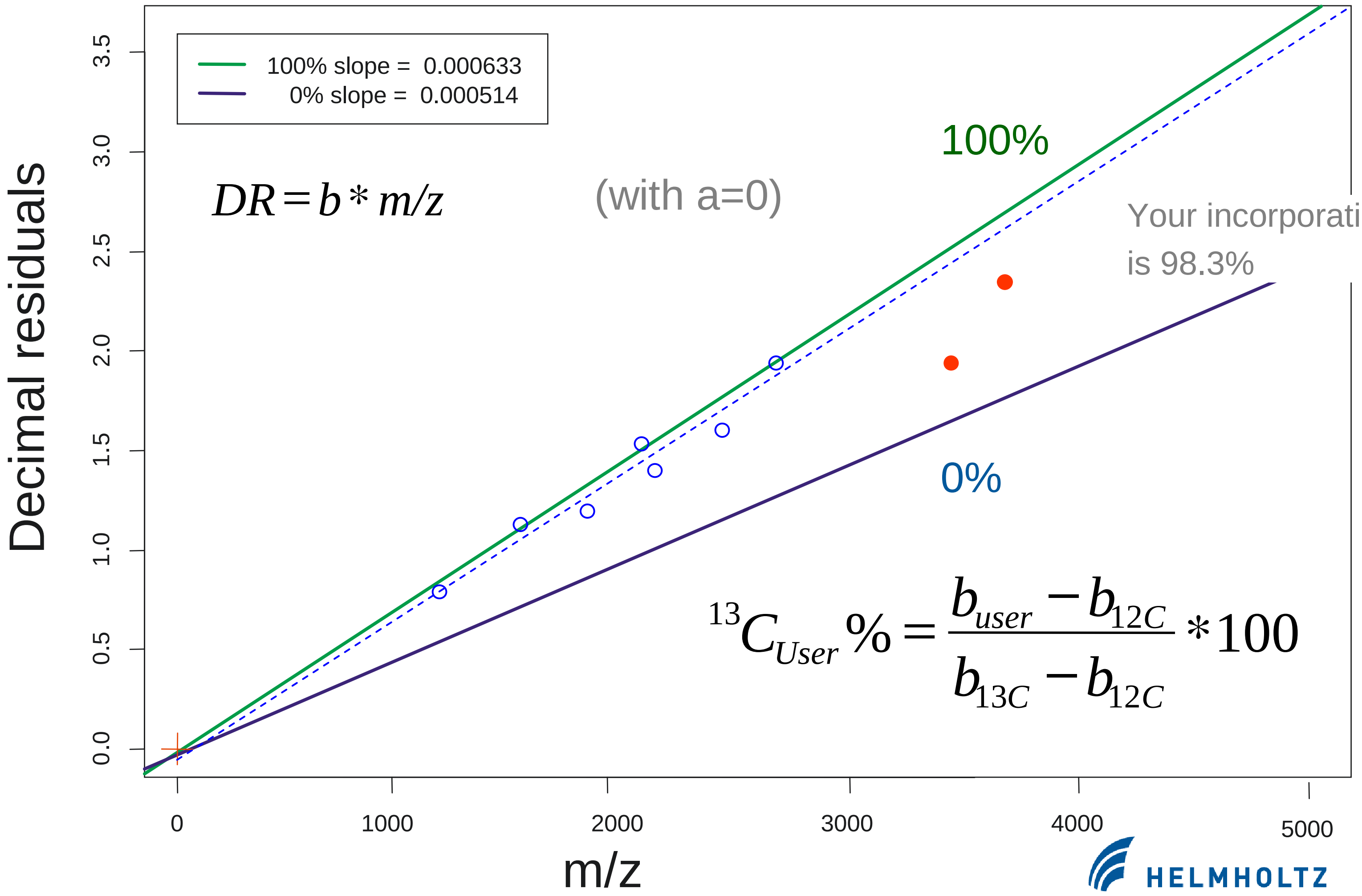
robust linear fitting

Breakdown point value set = 0.5
>50% of datapoints must change

rlm() function

MASS package (Venable & Ripley 2002)

Estimation of relative ^{13}C incorporation rates



$$DR = b * m/z$$

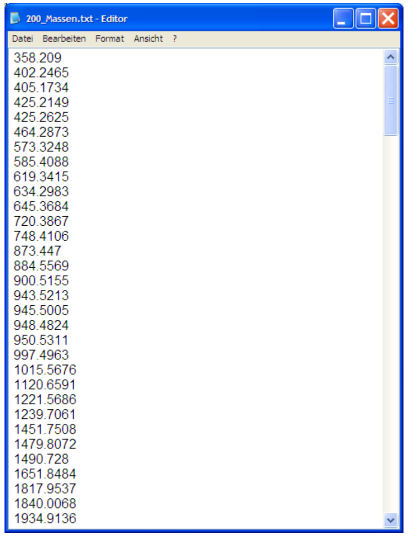
(with a=0)

Your incorporation rate is 98.3%

$$^{13}C_{User}\% = \frac{b_{user} - b_{12C}}{b_{13C} - b_{12C}} * 100$$

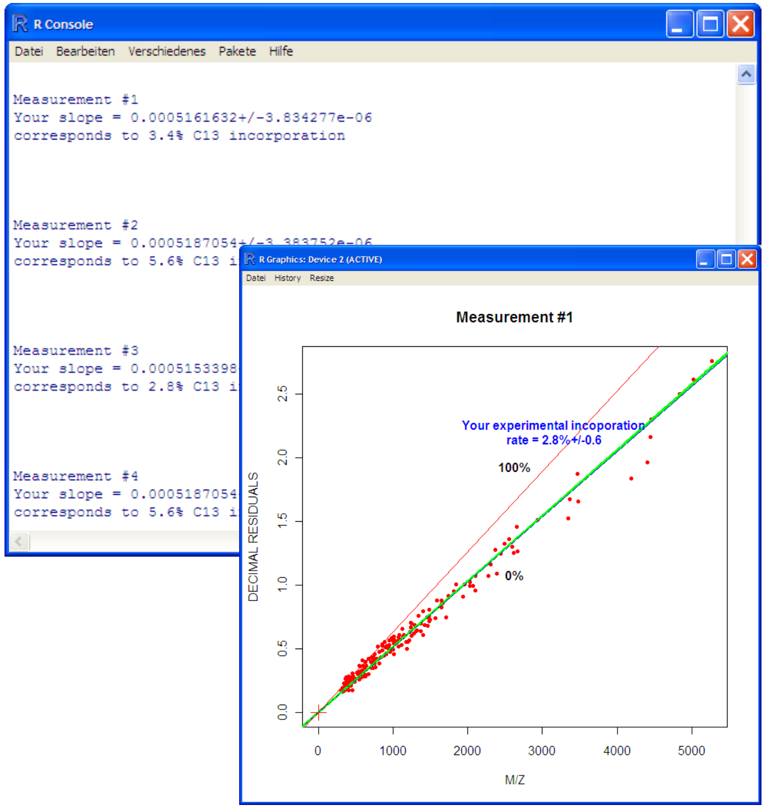
HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

User data input



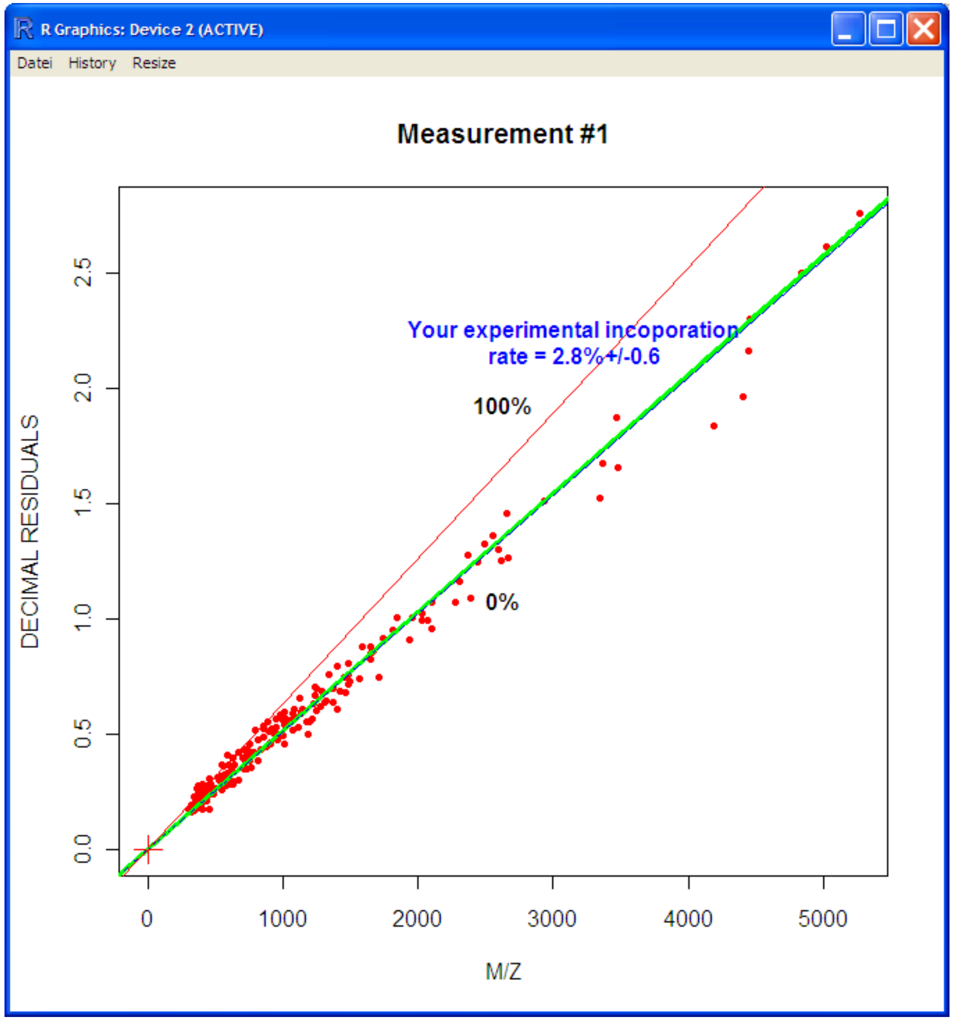
```
358.209
402.2465
405.1734
425.2149
425.2625
464.2873
573.3248
585.4088
619.3415
634.2983
645.3684
720.3867
748.4106
873.447
884.5569
900.5155
943.5213
945.5005
948.4824
950.5311
997.4963
1015.5676
1120.6591
1221.5686
1239.7061
1451.7508
1479.8072
1490.728
1651.8484
1817.9537
1840.0068
1934.9136
```

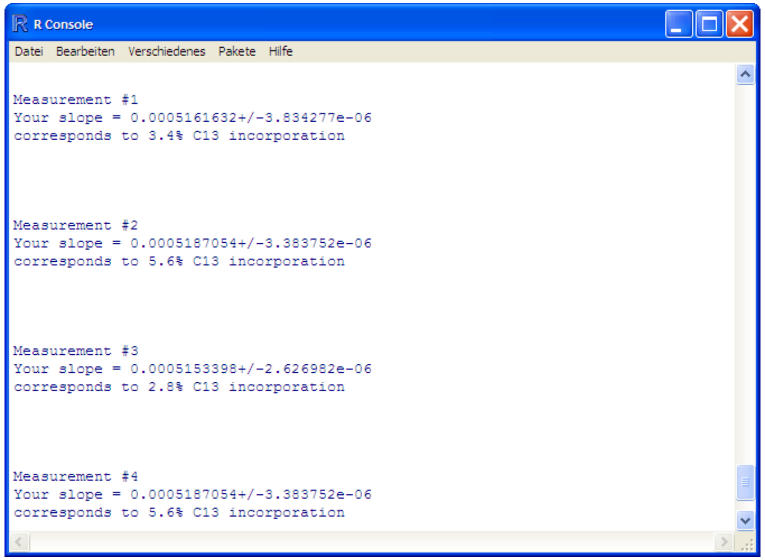
+ R Script 3



```
Measurement #1
Your slope = 0.0005161632+/-3.834277e-06
corresponds to 3.4% C13 incorporation



Measurement #2
Your slope = 0.0005187054+/-3.383752e-06
corresponds to 5.6% C13 i



Measurement #3
Your slope = 0.0005153398
corresponds to 2.8% C13 i



Measurement #4
Your slope = 0.0005187054
corresponds to 5.6% C13 i
```

User data output





```
Measurement #1
Your slope = 0.0005161632+/-3.834277e-06
corresponds to 3.4% C13 incorporation




Measurement #2
Your slope = 0.0005187054+/-3.383752e-06
corresponds to 5.6% C13 incorporation




Measurement #3
Your slope = 0.0005153398+/-2.626982e-06
corresponds to 2.8% C13 incorporation




Measurement #4
Your slope = 0.0005187054+/-3.383752e-06
corresponds to 5.6% C13 incorporation
```

Sensitivity of Method

Dataset *Pseudomonas putida*

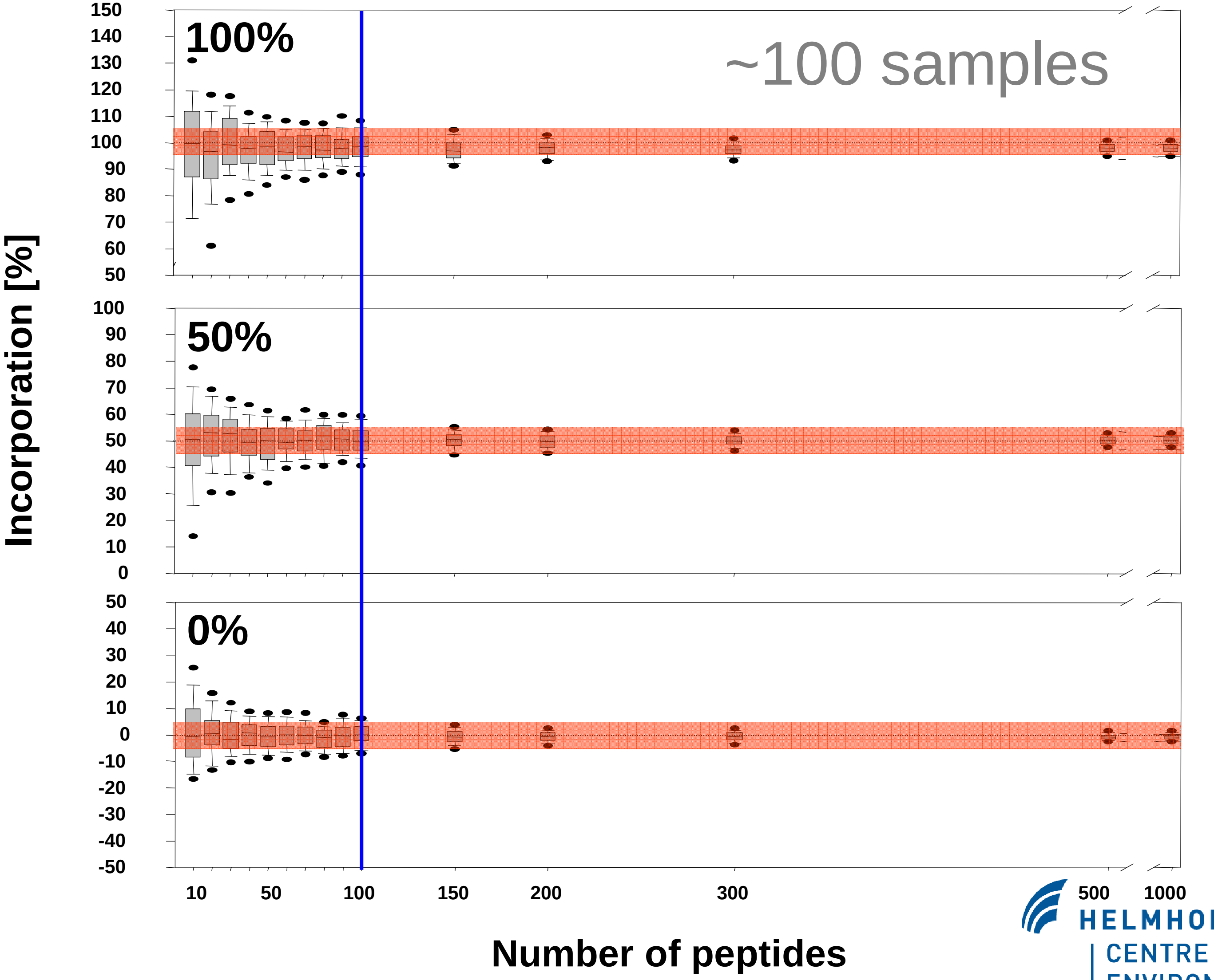
1. Calculated 50% and 100% ^{13}C incorporation

2. Randomly sampled 100 times each
   10-100 (steps by 10), 150, 200, 300, 500, 1000 sequences (0%,50% and 100%)

3. Statistics on estimated incorporation rate for 0%, 50% and 100%

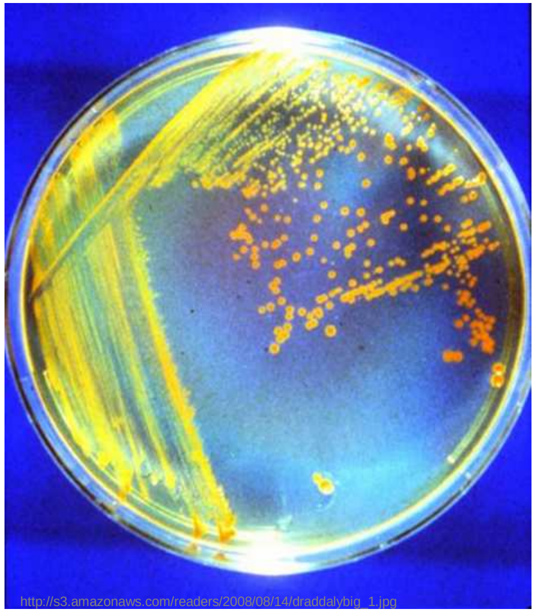
HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Sensitivity of Method



HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Conclusion


http://s3.amazonaws.com/readers/2008/08/14/draddalybig_1.jpg

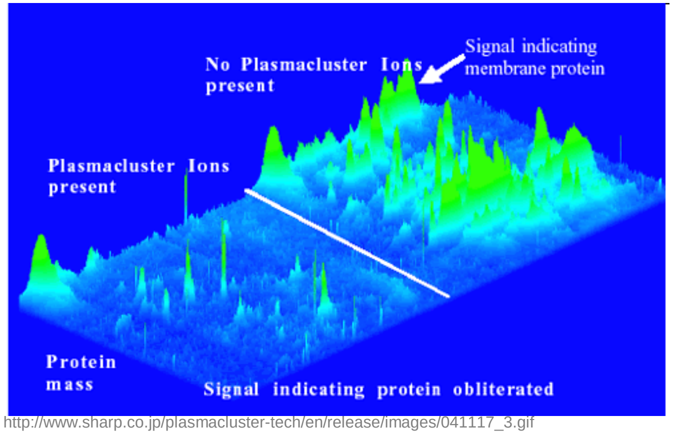
1. ‚Half-decimal place rule‘ useful for the estimation of ^{13}C incorporation rates
2. Robust linear models better suited for fitting of highly variable user data than MinSSE fitting
3. >100 measurements needed for prescision <5% incorporation estimation

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Outlook

1. Application of HPDR on DNA

2. Backcalculation to ^{12}C-peaks → function identification



http://www.sharp.co.jp/plasmacluster-tech/en/release/images/041117_3.gif

3. Include N-isotope incoorporation

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Acknowledgement

Nico Jemlich (UFZ)

Carsten Vogt (UFZ)

Martin von Bergen (UFZ)

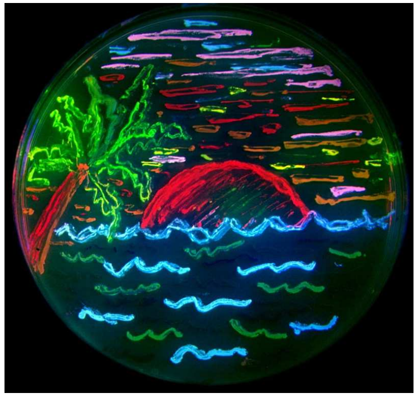
Hans-Hermann Richnow (UFZ)

Hauke Harms (UFZ)

Frank Schmidt (Uni Greifwald)

Jens Mattow (MPI Berlin)

Bernd Thiede (Uni Oslo)



http://cache.gawker.com/assets/images/gizmodo/2009/01/bactsunsuet_01.jpg

R development team

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ