

From ^{15}N metabolic labeling to targeted LC-MS; Different strategies applied to Systems Biology

Thierry Le Bihan
Nov 2009



Protein Analysis by MS

Protein



HPLC

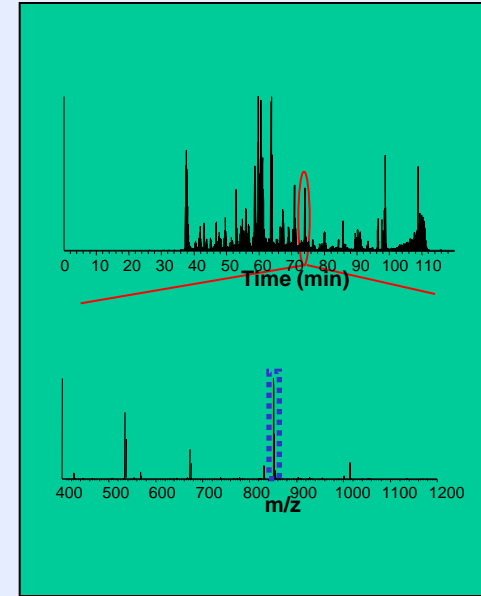


Time range chromatogram

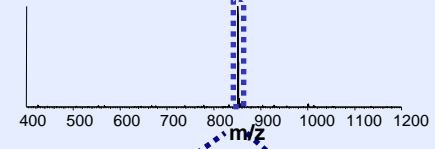
Mass range MS spectra

Mass range MSMS spectra

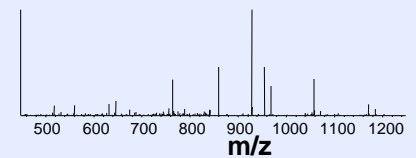
MS



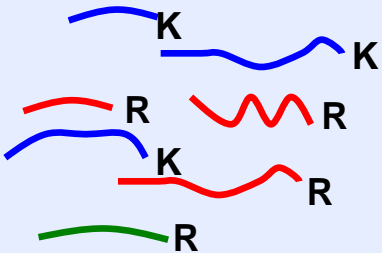
Isolation



Fragmentation



Protease digestion (trypsin)



From MSMS spectra to protein

Identification

Database

PROTEIN 1: MGHITYHGFHGRTPY
RGHVCMMNIKLTIT....

PROTEIN 2: MWQSTIFYGHTIFWVHKP
HPTIHLKHPLYITL...

PROTEIN 3: MGFWQHSTIFYGHTIFW
VHKHPTIHLKHPLYITL...

...

...

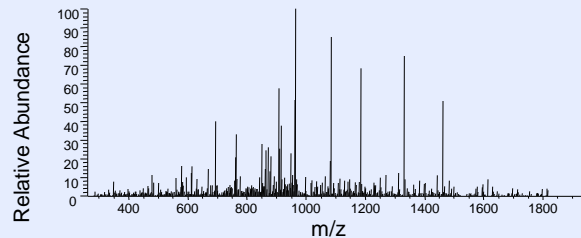
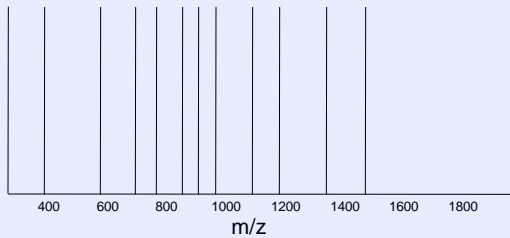
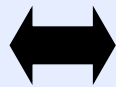


In silico digest

MGHITYHGFHGR
TGPYITK
MWQSTIFYGHTIFWVKP
PLYITTR



Generation of in-silico spectra experimental MSMS spectra



Peptide Identification by pattern match

TGHAHPLIR
AGFHPLTYWQR
FNMKLTYPHGLFSDER
WEYTIPLK

...

...



Peptides are clustered into proteins

protein model: gi|1351907|
MSL | homologues | Q7Z6L1 | x9L |
ncbi | gpm06

gi|1351907|
log(e) = -238.5

1 MEMVTFISILLIFSAYSRQVFRDTHREIARHFSDIGEEFKGLVLIAPSOVLOQDFP 60
61 DEHVFLVWELTFARCTVADESFRAGCSRLTFLDDELCKVALRETYGHADCCQKQP 120
121 ERKCFLSKEDDFFGLPKLRFDFRLCDKFRADRFKPFQVLYLVARHPQTFVAPRLIY 180
181 RMYHWQFQCCQARDFGACILRRIETWRFVRLARAFGLRQACIQPFRRALSARWFA 240
241 RLSQRFPAFVETLVDTLTVFVDRCCGDLCCADDRADLAETICDQDTISIKRE 300
301 CQDPLERKSHCIARVERKDAIFENLFFETADPAEDFQWCKVQKADAFLOSFLVYRER 360
361 RPEYAVVLLRLAREYELREKCCARDDFRACTYVYDKLELVDFQVGLIKQKCGQFKR 420
421 LGEYFQCALIVYVTRPQVDFVFLVWRFELGQVDFQKCFKFRSEMECTEYDRLI 480
481 NELCVLMEKTFVSRVETKCTCSLWRRFCFSAITFDTYVKAQKLPFFPHADICTLP 540
541 DTRQIQRQATLVELLKEFKETREBGLTVMENFVAFVDECCARDQKACFVFGKLVV 600
601 STQZALA 607

spectrum	log(e)	log(I)	m-h	delta	z	sequence
240.1	-1.5	3.12	1554.043	1.990	3	vhr ⁶ GLVLIAPSOV ²¹ frd
45.1	-5.2	2.73	1249.622	-0.011	2	ish ²⁵ VDLGEENK ⁴⁴ ghl
36.1	-1.2	3.05	974.459	0.009	2	hrk ³⁷ DLGEENK ⁴⁴ ghl
130.1	-4.5	4.04	1163.632	1.002	2	ehk ⁶⁴ LVNLTFAK ⁷⁷ ova
5.1	-11.1	2.85	1463.567	0.053	3	efak ⁷⁴ VADESHAK ⁸⁸ shb
6.1	-5.7	2.36	1463.567	0.060	2	efak ⁷⁴ VADESHAK ⁸⁸ shb
140.1	-6.8	3.74	1419.683	0.036	2	qcek ⁸⁹ SLYTLFGDEL ¹⁰⁰ ast
30.1	-4.5	3.23	1478.501	0.053	2	ast ¹⁰⁴ ETVGMAL ¹¹⁷ pepe
71.1	-16.0	2.90	2524.141	1.074	3	coek ¹¹⁸ PERKALSHKDDSPDKL ¹³⁰ apd
61.1	-6.7	2.67	2541.157	0.059	3	coek ¹¹⁸ PERKALSHKDDSPDKL ¹³⁰ apd
63.1	-6.3	2.89	1901.859	1.059	3	epar ¹²³ QLSHKDDSPDKL ¹³⁰ apd
89.1	-6.8	2.72	2019.959	0.072	2	dipk ¹³⁹ LKPDNITDEFKADEK ¹⁵⁵ fnq
97.1	-6.2	2.80	1576.757	0.054	2	dipk ¹³⁹ LKPDNITDEFKADEK ¹⁵⁵ fnq
62.1	-5.7	2.65	2613.153	0.092	3	yaaq ¹⁷⁷ LLVYANKYNGVFOAEK ¹⁹⁷ pad
60.1	-5.5	3.14	1747.683	0.057	2	yaaq ¹⁷⁷ LLVYANKYNGVFOAEK ¹⁹⁷ pad
49.1	-1.1	3.74	922.489	-0.001	2	kfgk ²⁴⁴ AEEVETK ²⁷⁸ hyd
73.1	-17.5	2.88	2247.909	0.089	3	kyrk ²⁴⁷ SHDILLKADRADLAK ²⁸³ hyd



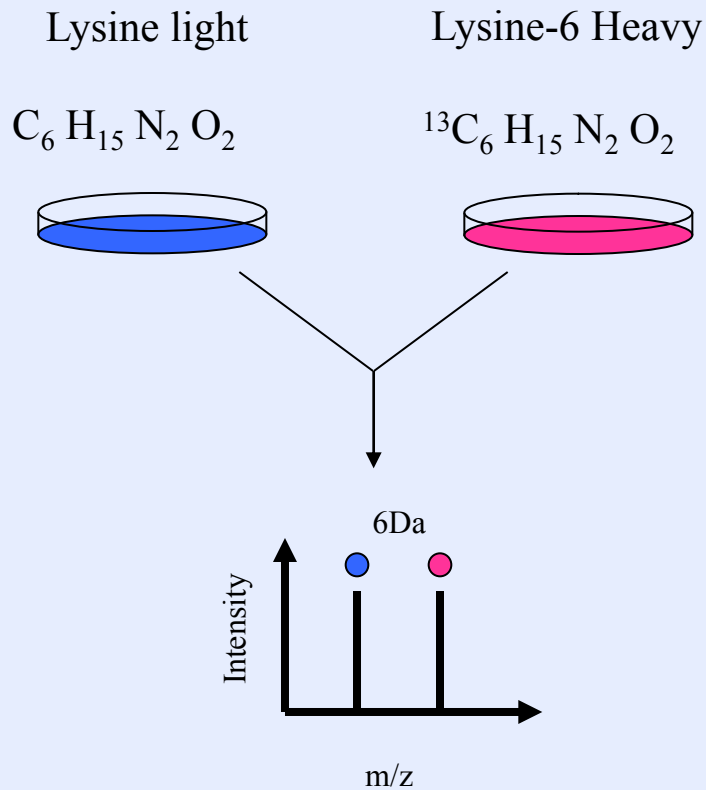
^{15}N metabolic labeling of the unicellular algae *Ostreococcus tauri*

Common Stable isotopes and Their natural abundance

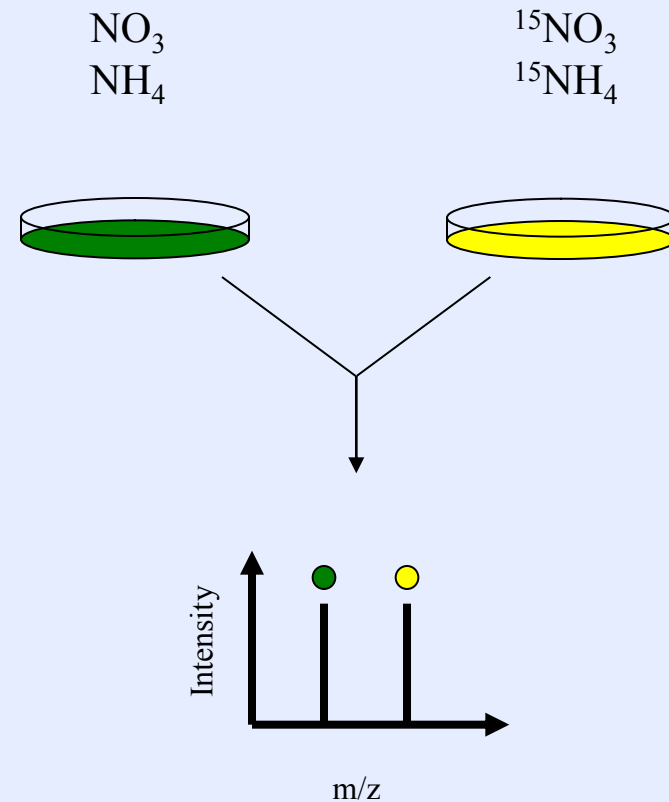
Element	Number of isotopes	Mass	Abundance
H	2	1.0078	0.9999
		2.0141	0.0002
C	2	12.0000	0.9889
		13.0034	0.0111
N	2	14.0031	0.9963
		15.0001	0.0037
P	1	30.9738	1.0000
S	4	31.9721	0.9502
		32.9715	0.0075
		33.9679	0.0421
		35.9671	0.0002
O	3	15.9949	0.9976
		16.9991	0.0004
		17.9992	0.0020

In vivo labeling strategies based on Stable isotope incorporation

SILAC



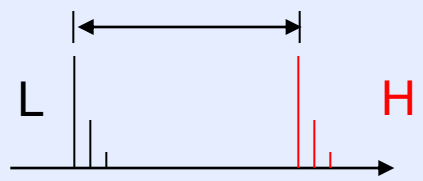
Metabolic labeling



SILAC

Light		Heavy	
G	F	G	F
A	Y	A	Y
S	W	S	W
T	D	T	D
C	E	C	E
V	N	V	N
L	Q	L	Q
I	H	I	H
M	P	M	P
K	R	K	R

$\Delta m =$ either +8 or +10Da



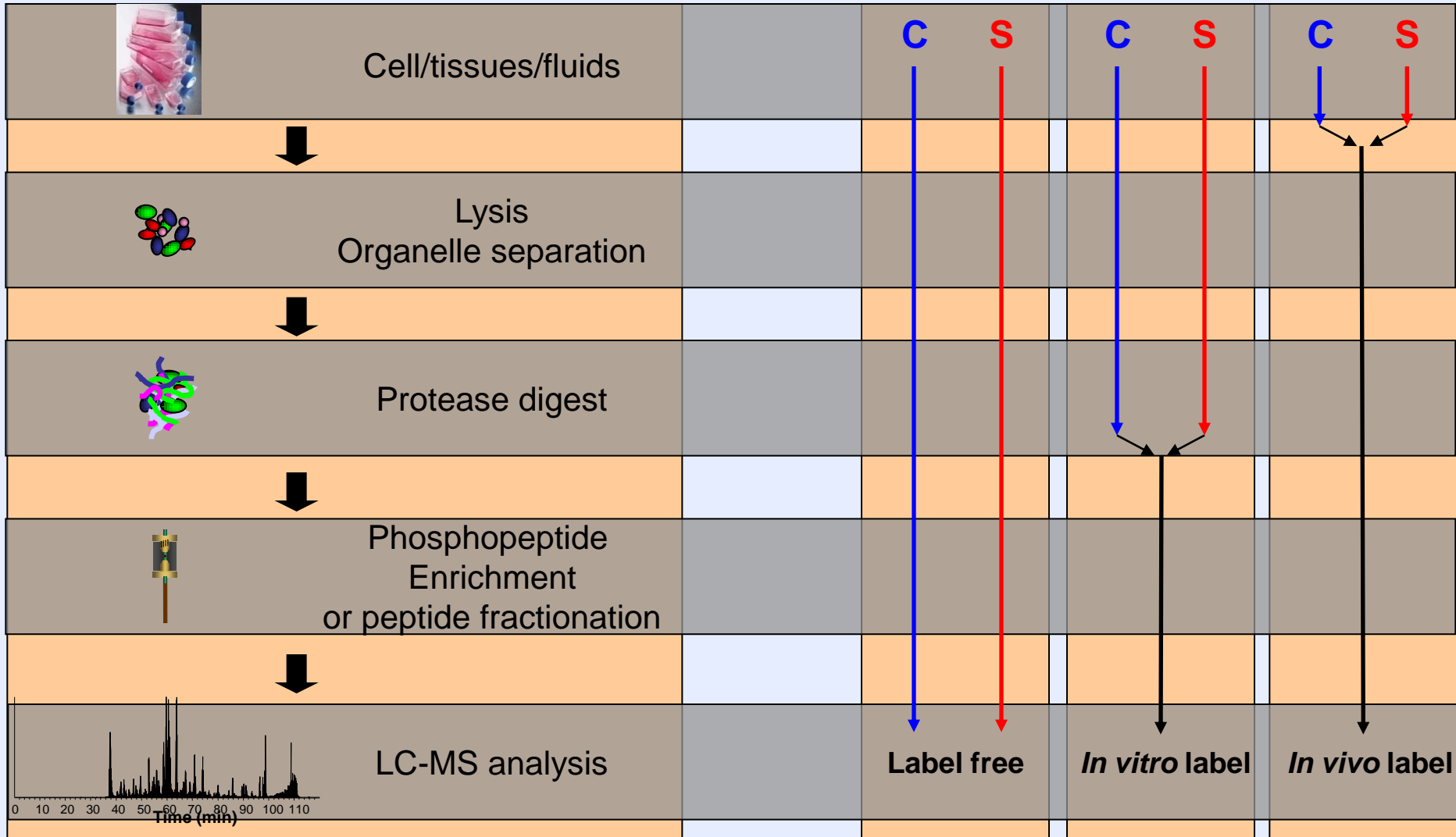
Metabolic labeling with ¹⁵N

Light		Heavy			
G	F	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;">1Da</div> </div>			
A	Y				
S	W				
T	D				
C	E				
V	N				
L	Q				
I	H				
M	K				
P	R				
		2Da			
		G	F		
		A	Y		
		S	M		
		T	D	W	
		C	E	N	
		V	L	Q	
		I	P	K	
				3Da 4Da	
				H	R

Δm function of peptide length and content in WNQHKR (can be used for seq confirmation)

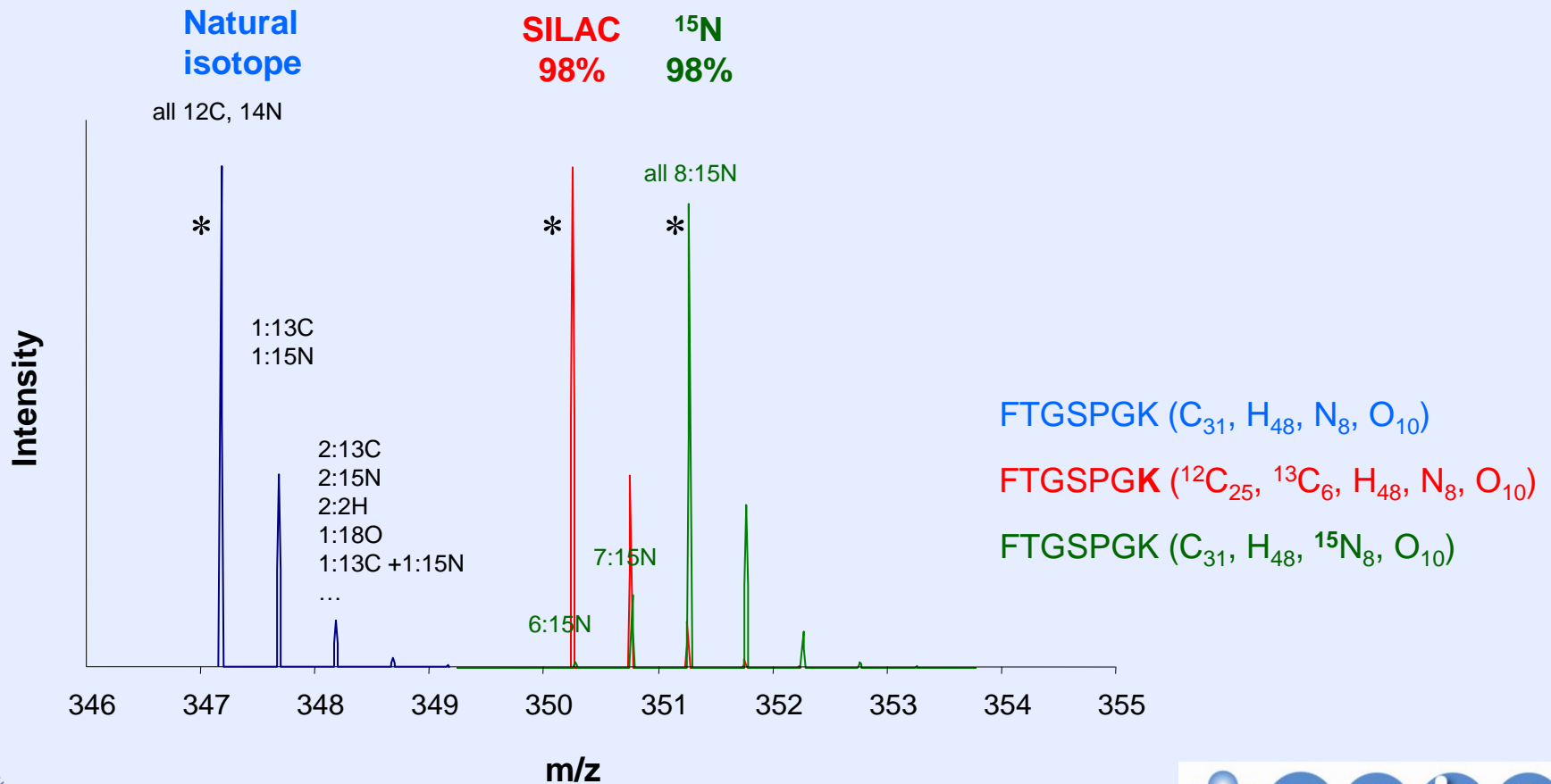


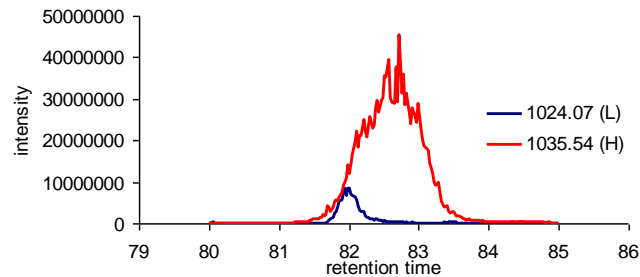
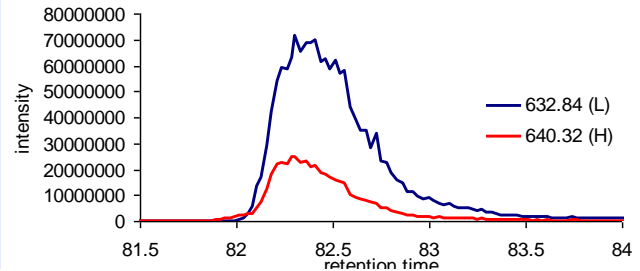
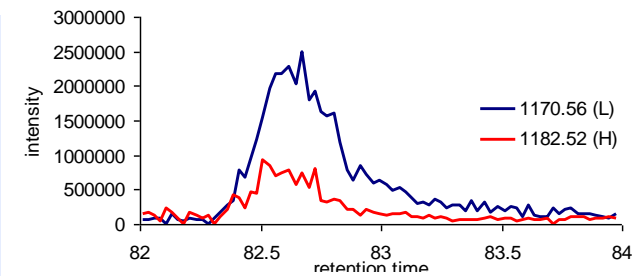
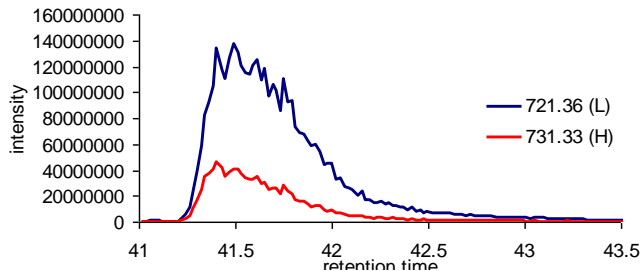
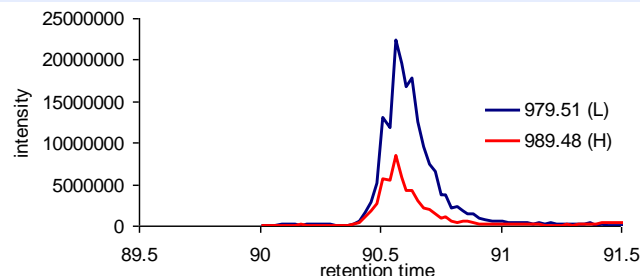
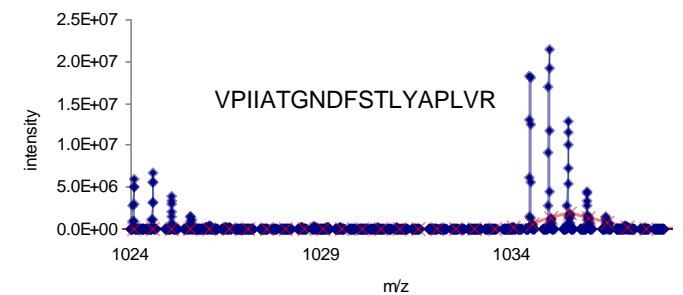
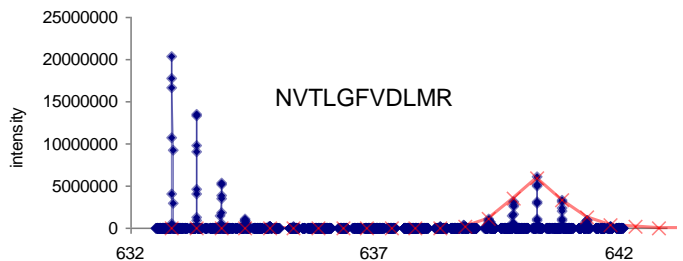
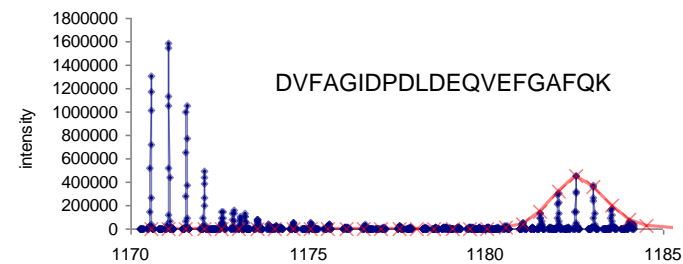
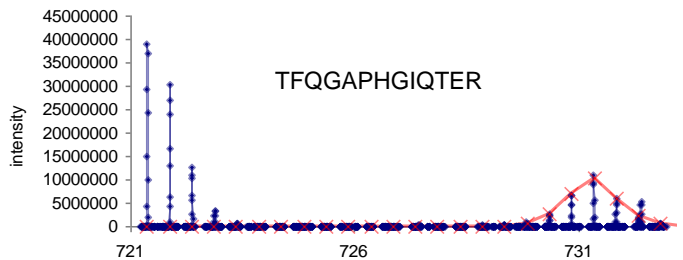
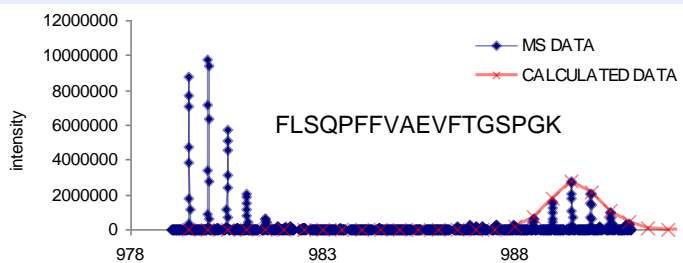
Different quantitation strategies



Comparison of Different Labeling Strategies at the Spectra Level

FTGSPGK ($C_{31}, H_{48}, N_8, O_{10}$)
2+ charge





Base peak ratio	MS peak ratio
2.92	3.09
3.63	2.97
2.67	2.82
3.29	2.81
0.1	3.08

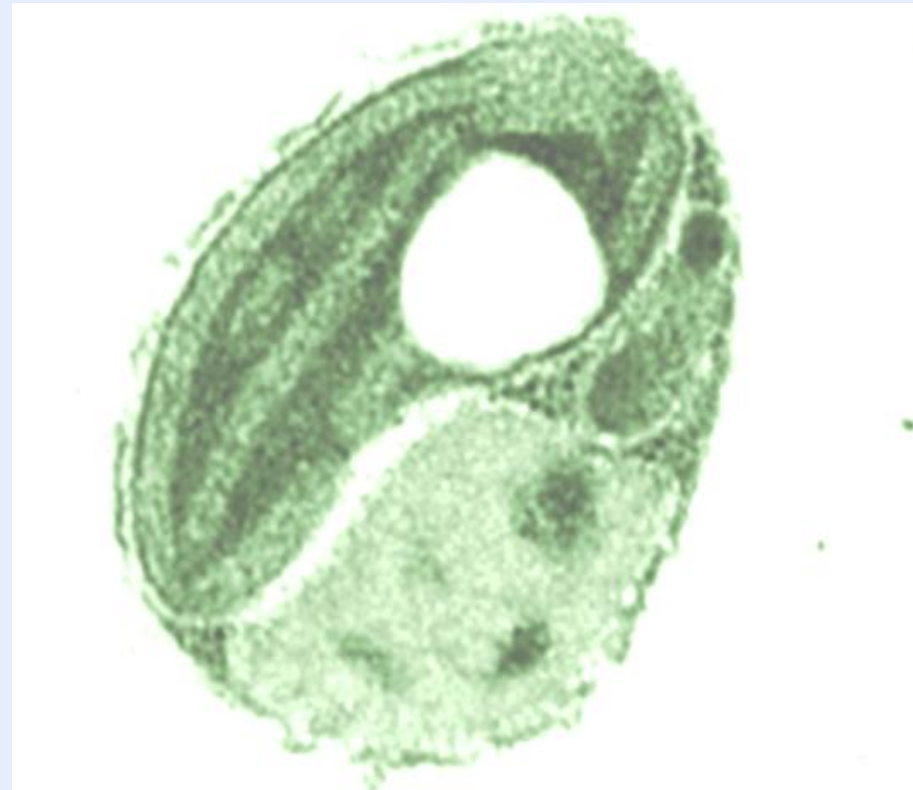


Why studying *Ostreococcus tauri* (Ot)

- Easy to grow, genome fully sequenced (but poorly annotated)
- One of the most primitive and the smallest free living eukaryote
- 12.6Mb sized genome/7892 genes
- Very limited set of core cell cycle genes
- “Good plant like model” (no cell wall)
- Population naturally synchronised (strong relation between circadian and cell cycle)

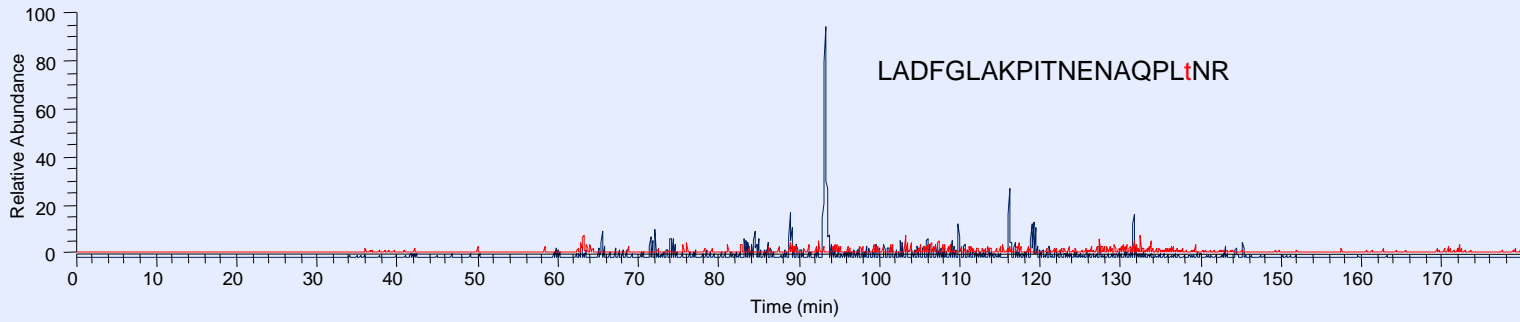
Our interests are:

- Culture media optimization for metabolic labelling
- Relationship between cell cycle and circadian clock
- Development of tools for the evaluation of ^{15}N incorporation
- Comparison of the proteome under different conditions/stresses



Cell cycle dependant kinase C (CDKC)

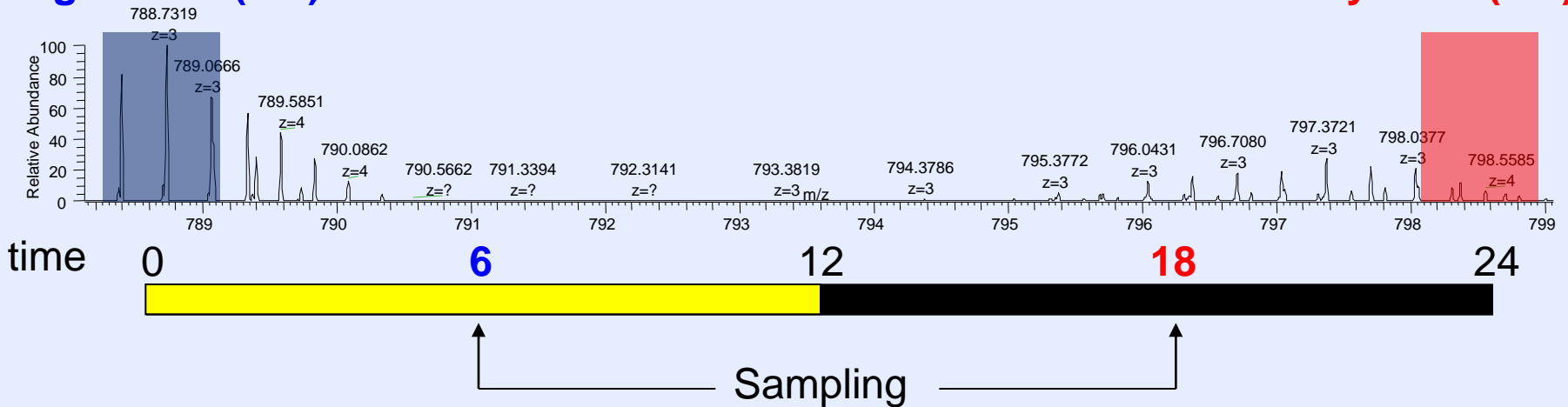
Metabolic labeling $^{14}\text{N}/^{15}\text{N}$ and Titanium dioxide enrichment



Light form (^{14}N)

ZT6L-ZT18H

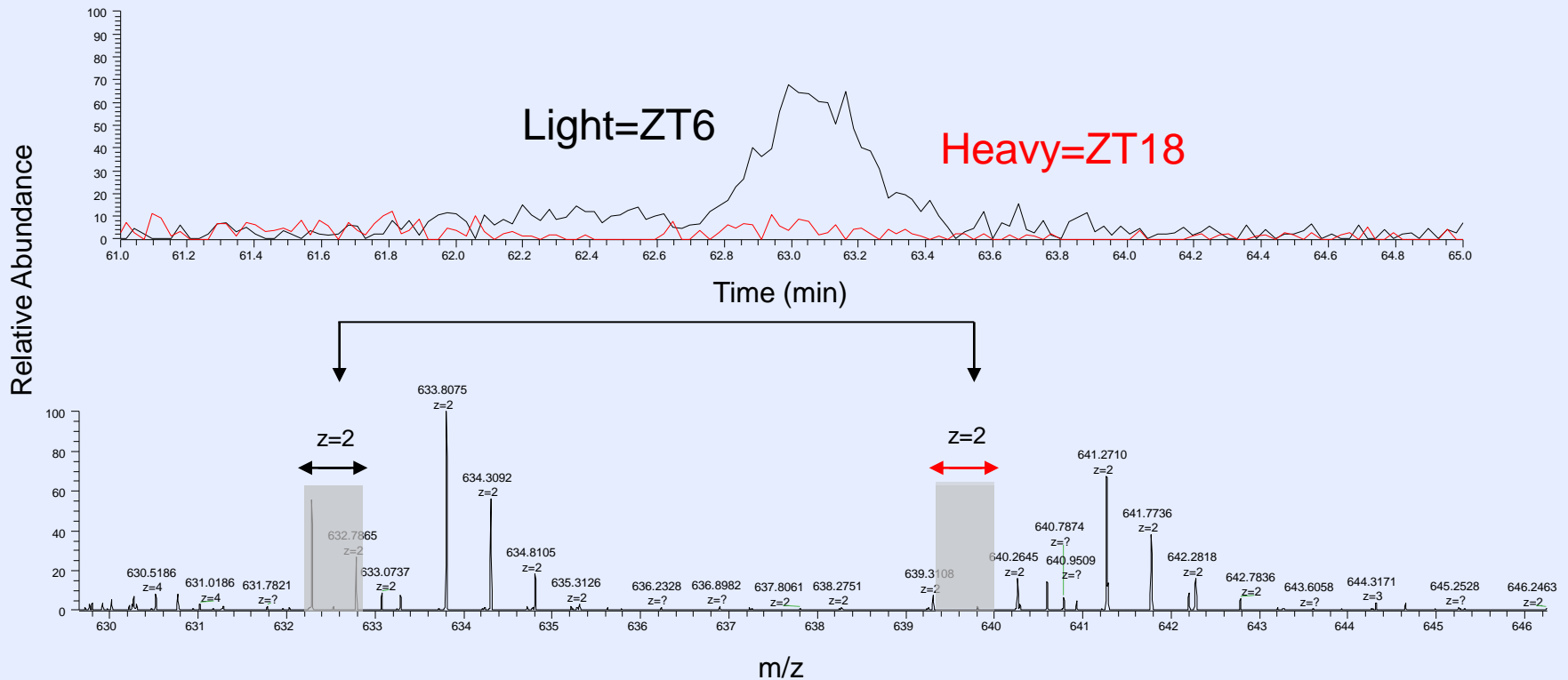
No Heavy form (^{15}N)



Regulatory domain phosphorylated in the middle of the day and dephosphorylated at night

PAS protein: Histidine kinase-like ATPases

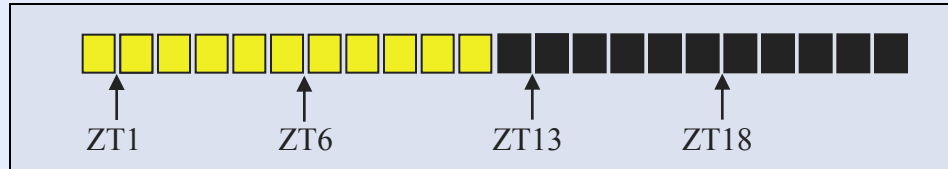
Phosphorylation state



SYsAGSLAATQK

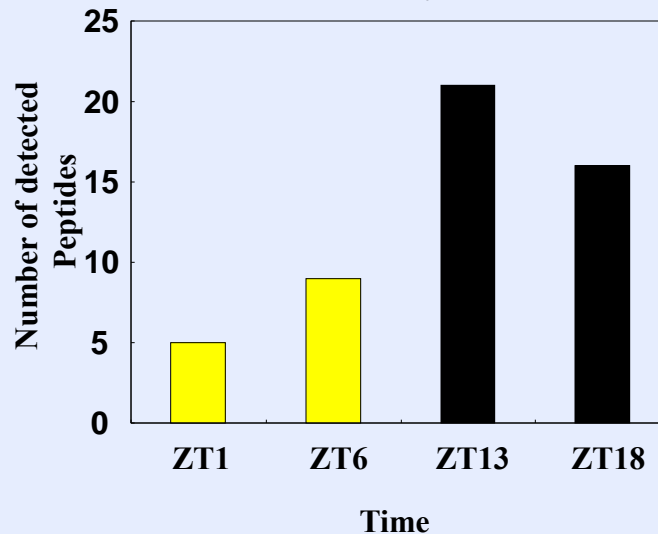


Some observations of various proteins at different time points



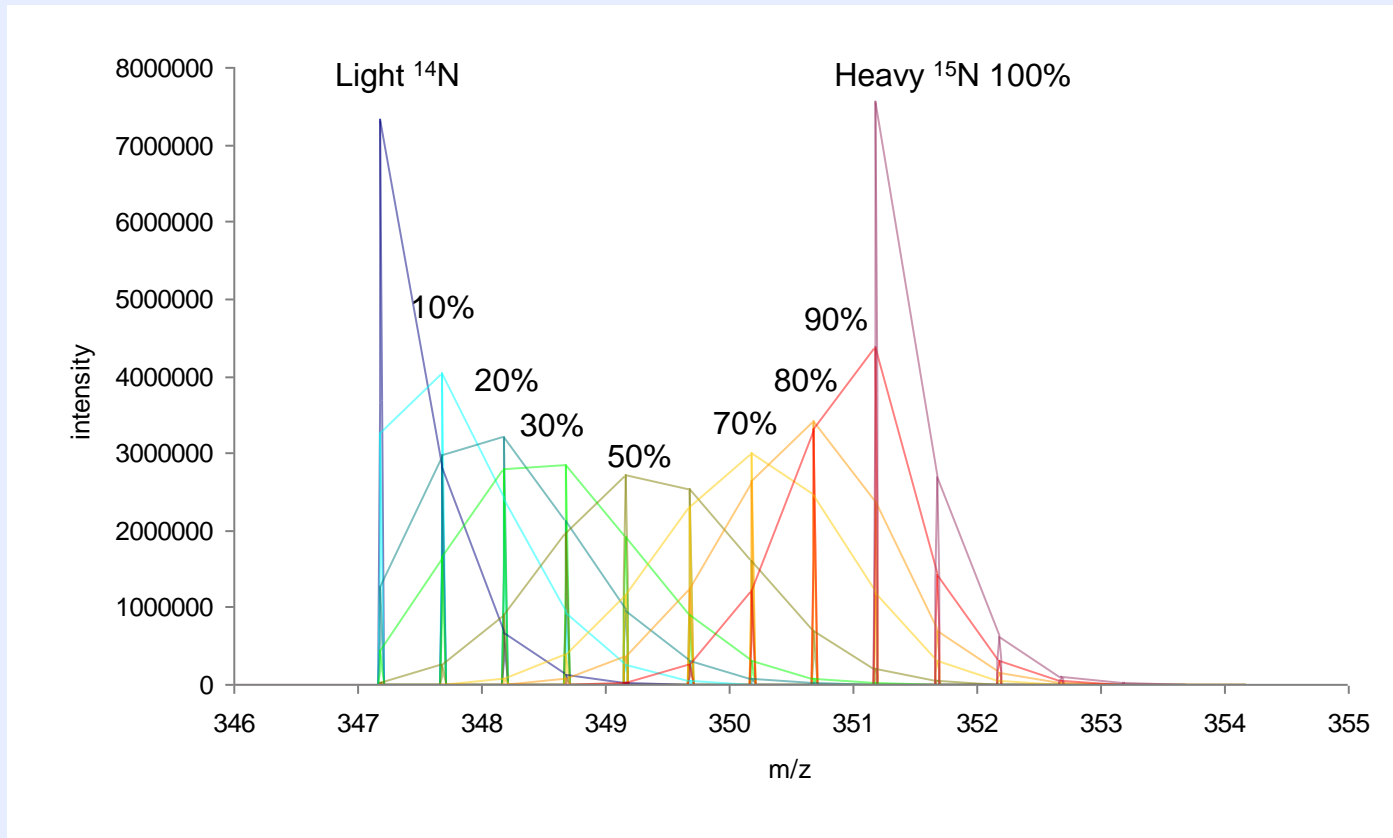
Identifier	Protein name	ZT1	ZT6	ZT13	ZT18	Fraction
gi 116057477	Dynamin	Down		Up		unsoluble
gi 116059445	his-containing phosphotransfer protein		Down		Up	cytoplasm
gi 116056366	ABC transporter C family protein	Up	Up	Down	Down	unsoluble
gi 116056680	putative vesicle-associated protein		Up	Up		unsoluble
gi 116060889	unnamed protein			Up	Up	unsoluble
gi 116055375	Pyruvate kinase	Down	Up	Up	Down	cytoplasm

Granule Bound Starch Synthase1 level (GBSSI)



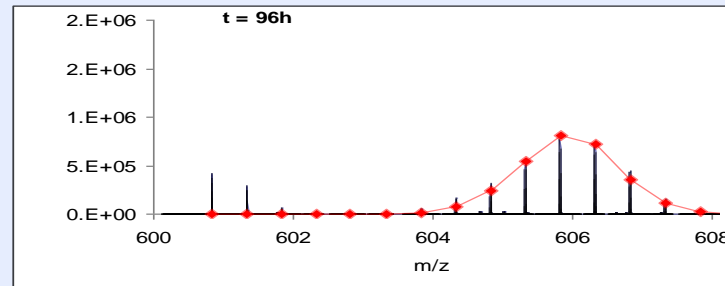
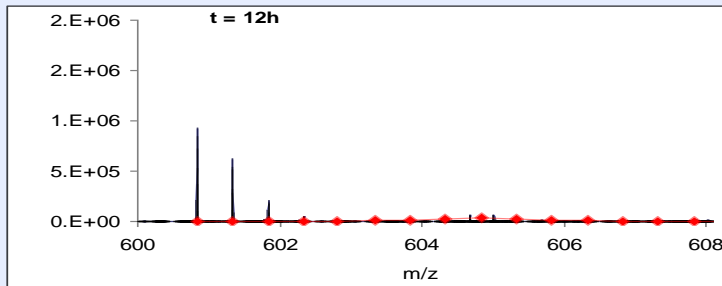
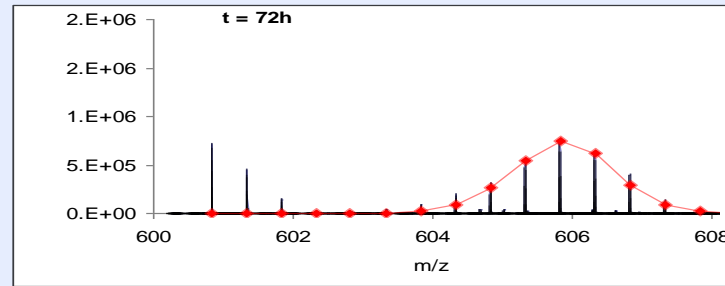
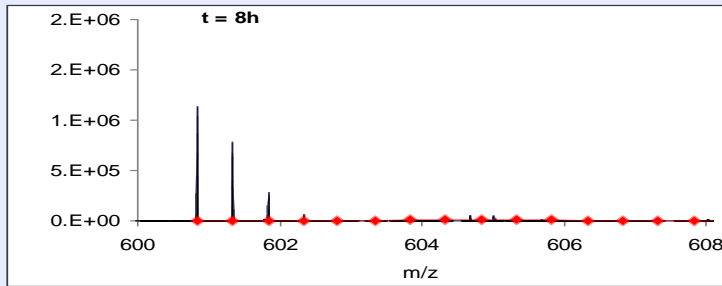
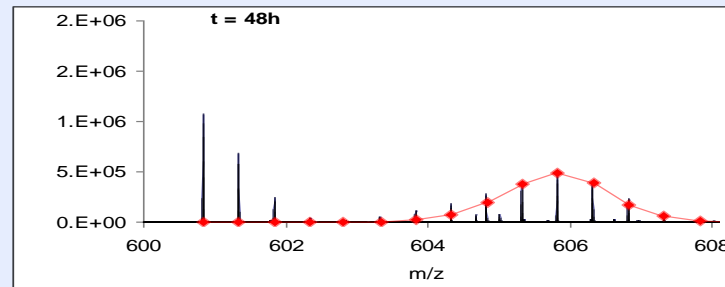
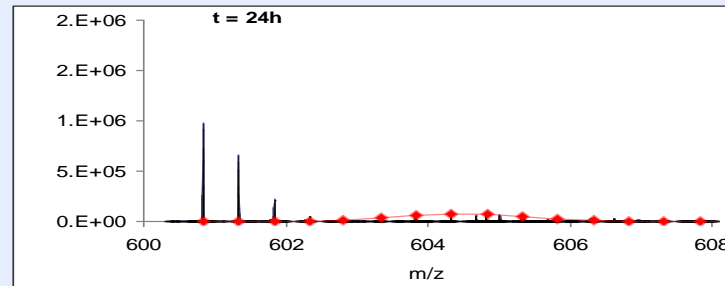
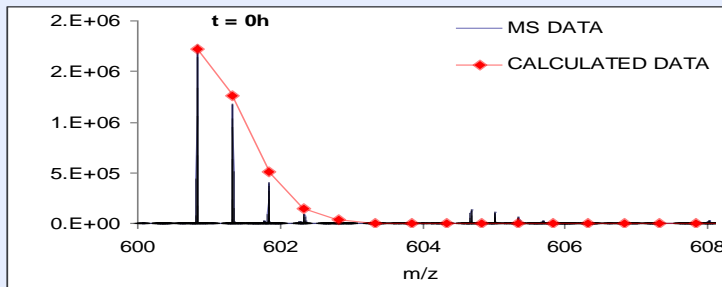
Simulation of a ^{15}N incorporation Time course

FTGSPGK ($\text{C}_{31}, \text{H}_{48}, \text{N}_8, \text{O}_{10}$)
2+ charge

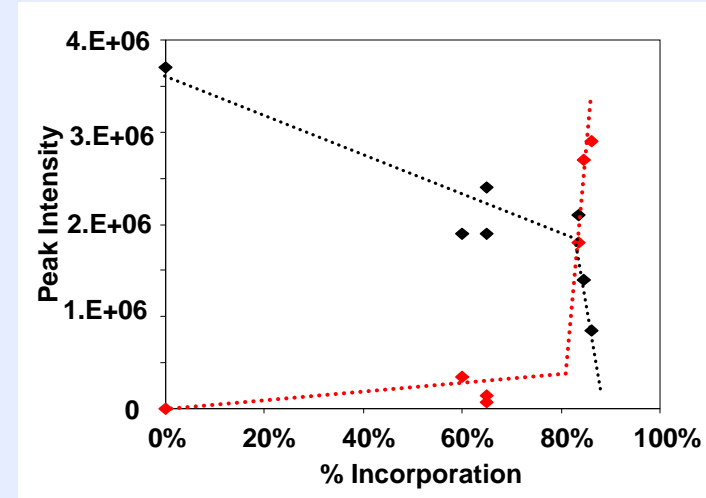
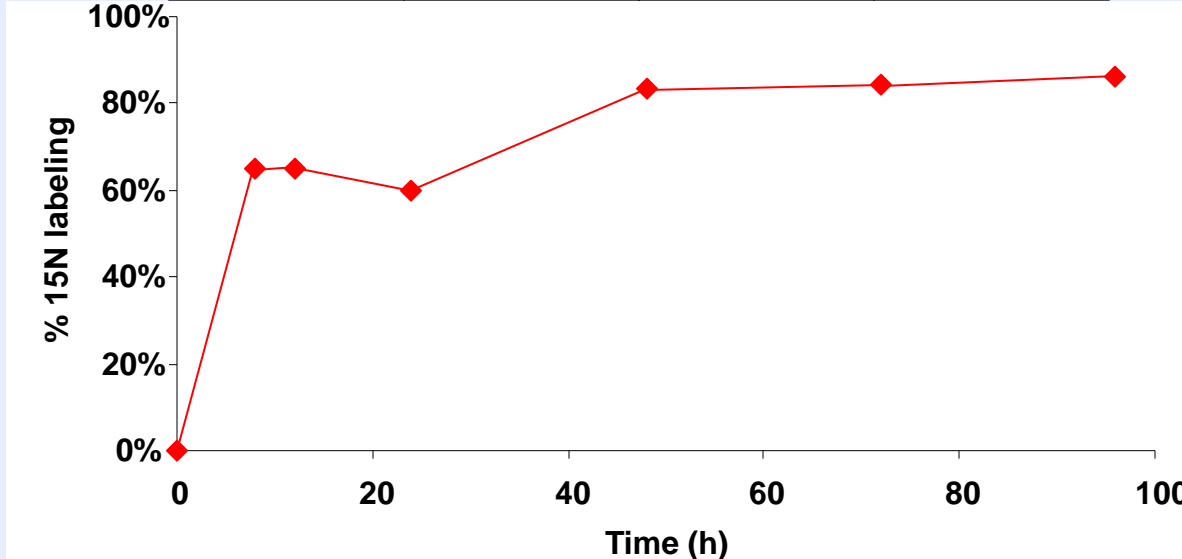
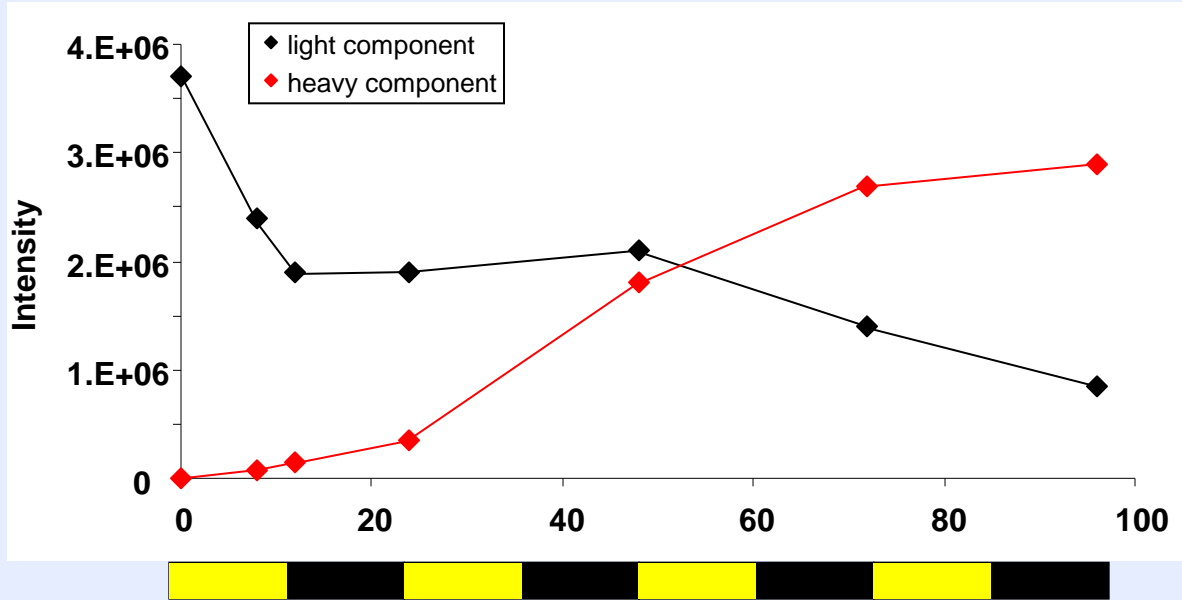


Experimental ^{15}N incorporation: Time course

Peptide sequence:
FLFVAEAIYK

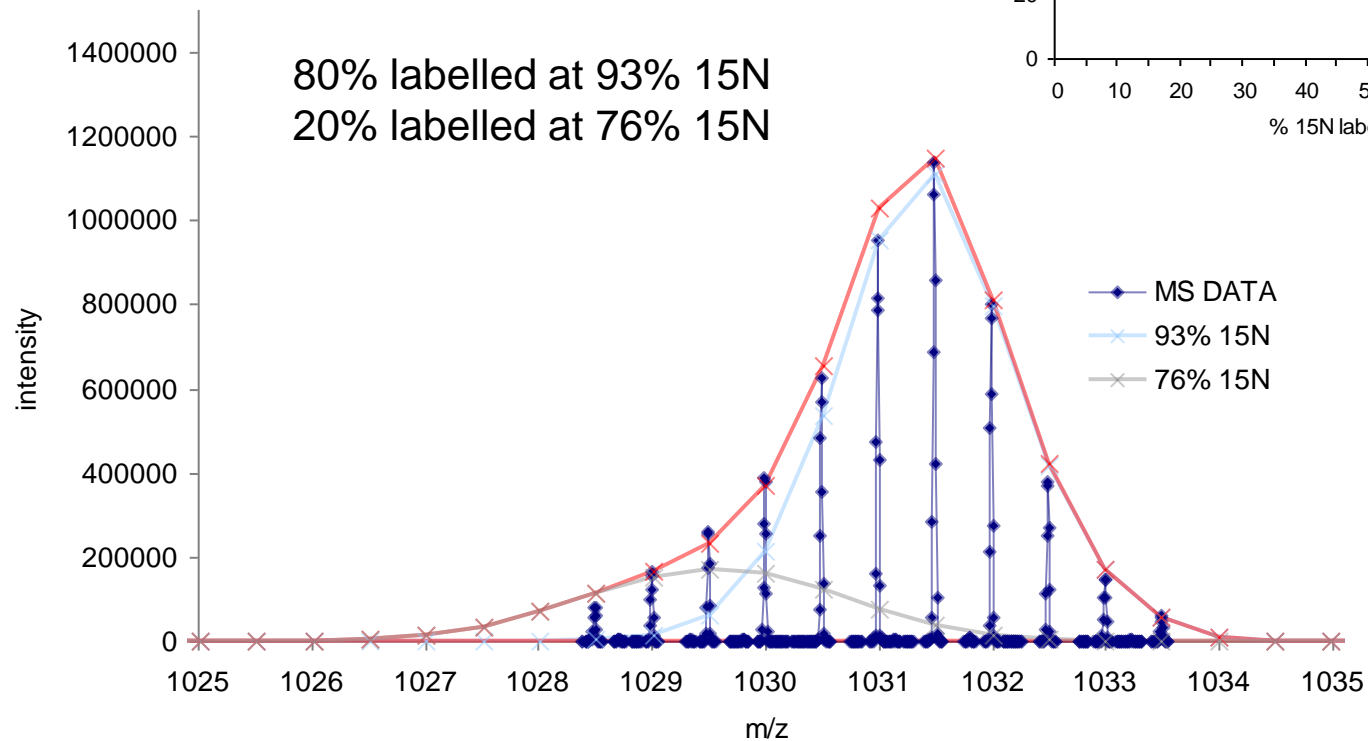
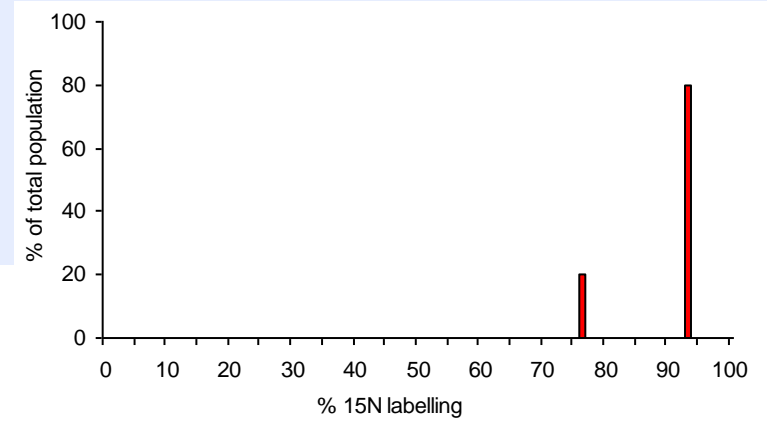


Experimental ^{15}N incorporation: Time Course



Example of an unexpected result: Low ^{15}N incorporation and isotopic pattern distortion

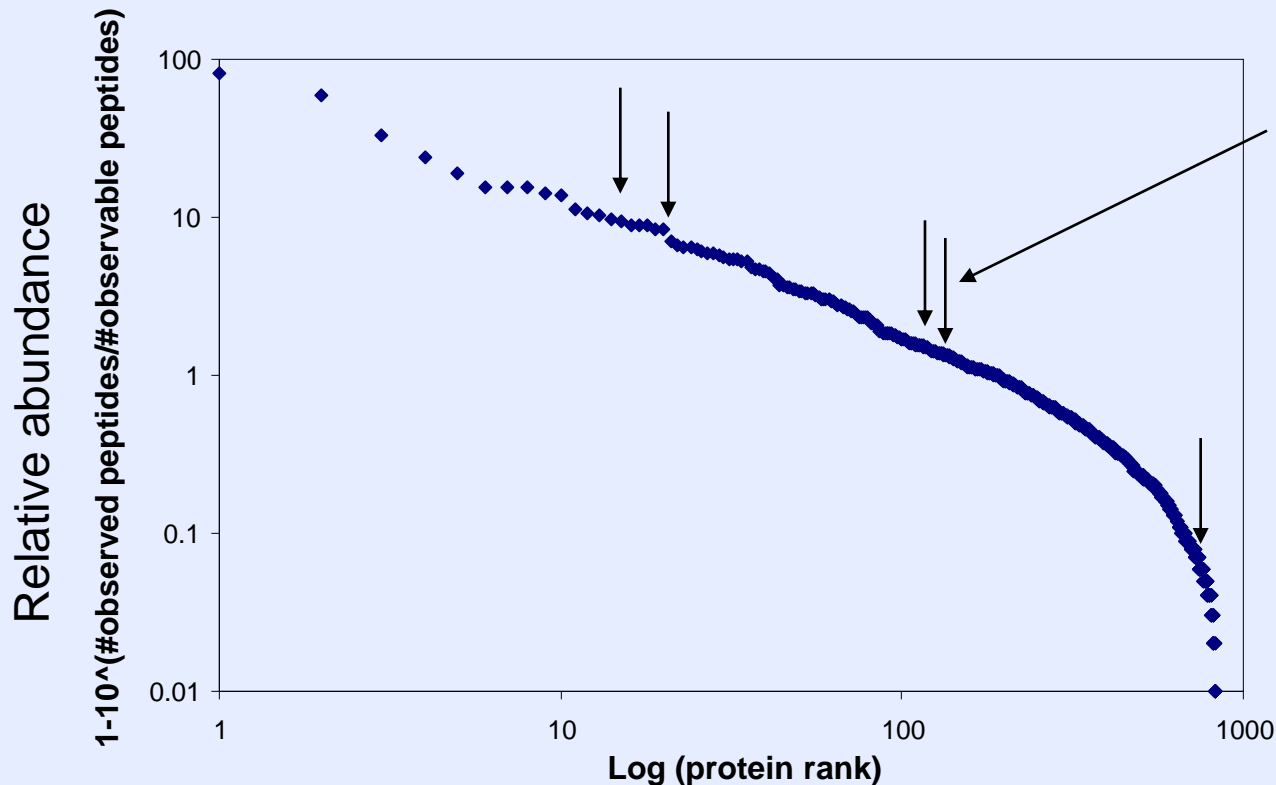
MFETFSFLPPLSDAEIAK



Targeted approach



Protein distribution in the Cytoplasm of HeLa cells



NAP1L1 Nucleosome assembly protein 1

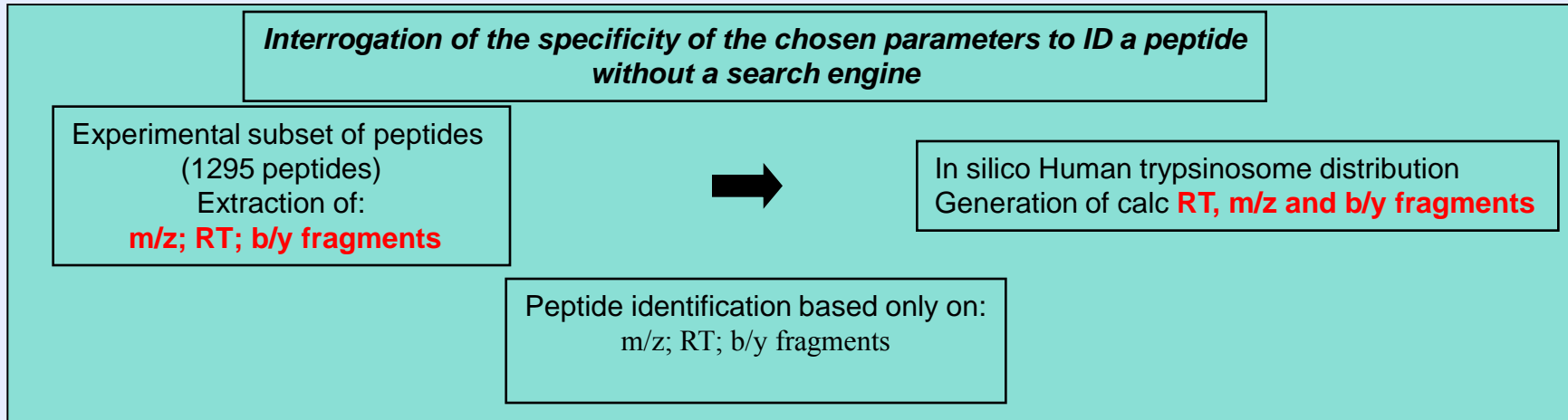
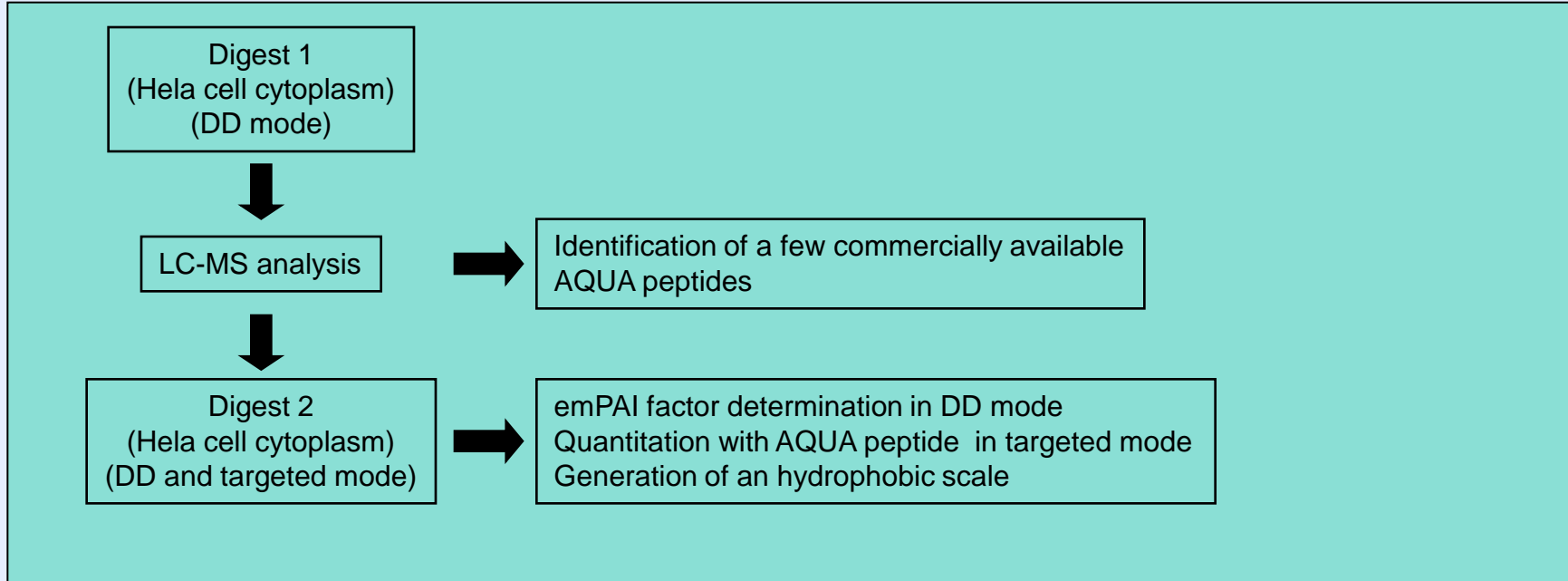
```
SEPDDSDPFSFDGPEIMGCTGCQIDWK  
QLTVQMMQNPQILAALQER  
FEIINAIYEPTEEEECEWKPDEEDEISEELKEK  
ADIDNKEQSELDQDLDDVEEVEEETGEETK  
FSDAGQPMSFVLEFHFEPNEYFTNEVLTK  
LDGLVETPTGYIESLPR  
FYEEVDLER
```

AQUA peptide:

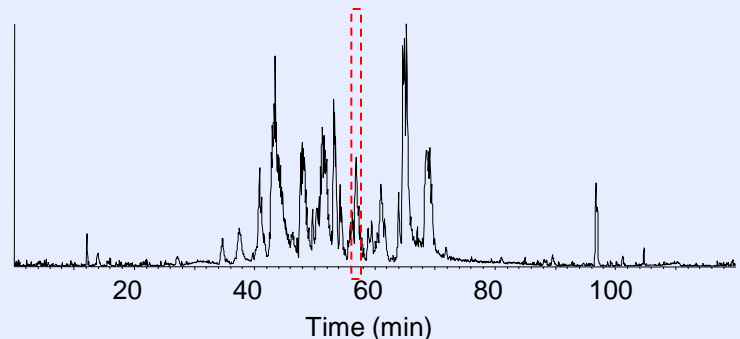
Peptides that can be bought in their heavy form and in a known quantity

Evaluation of the absolute quantity of a few anchor points and adjust for the rest of the curve.

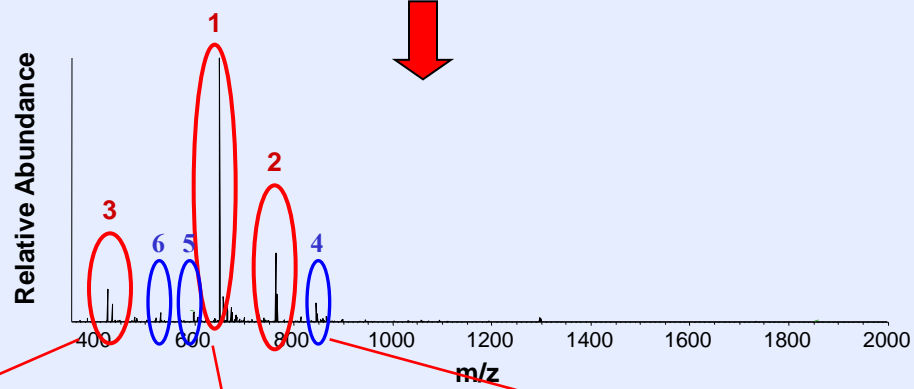
Experimental and *in-silico* design



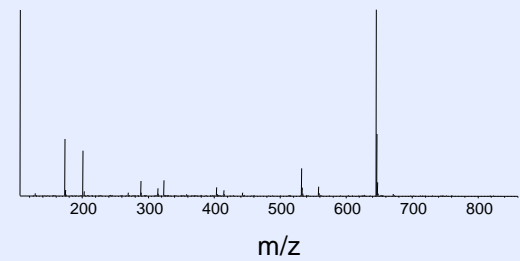
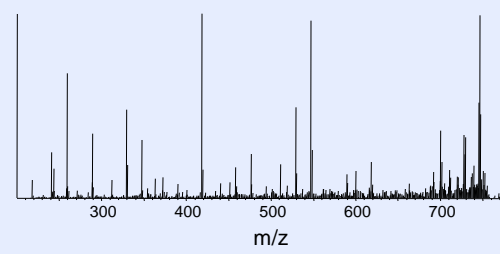
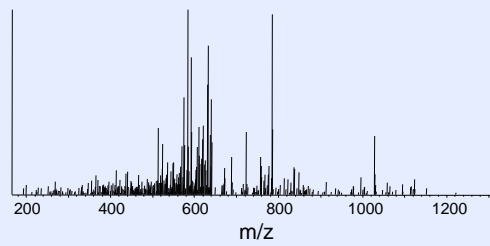
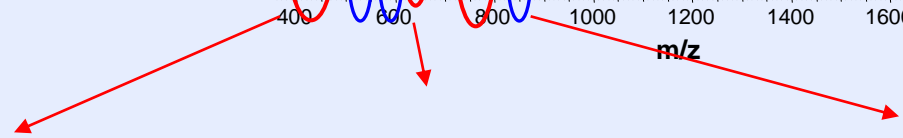
Data Dependant approach:



Time range



LC-MS
Mass range
400-2000amu

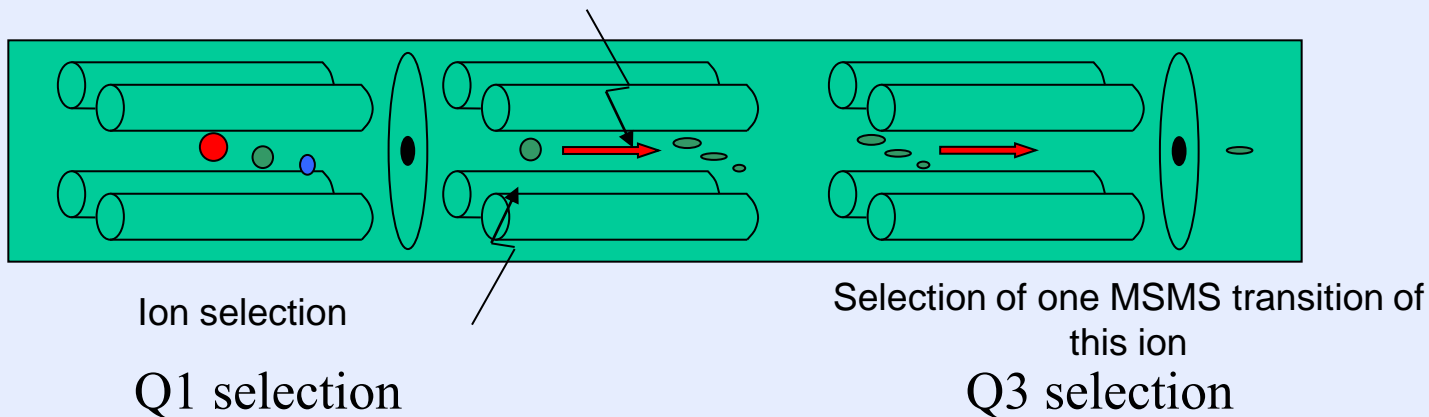


LC-MS²
Ca 100-1500amu

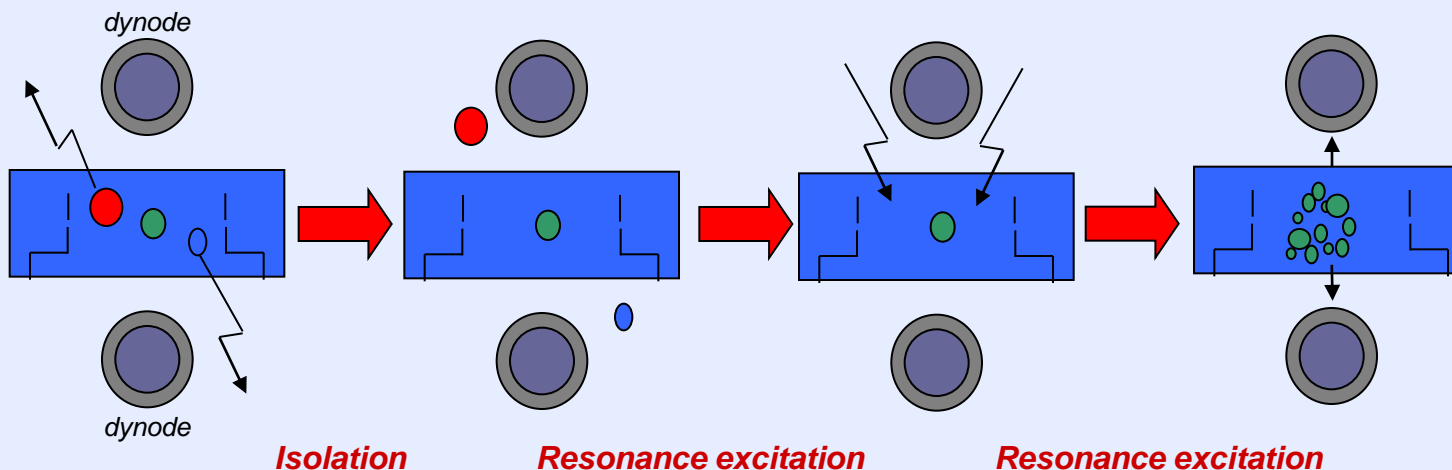


Targeted Approach Using SRM/MRM Assay

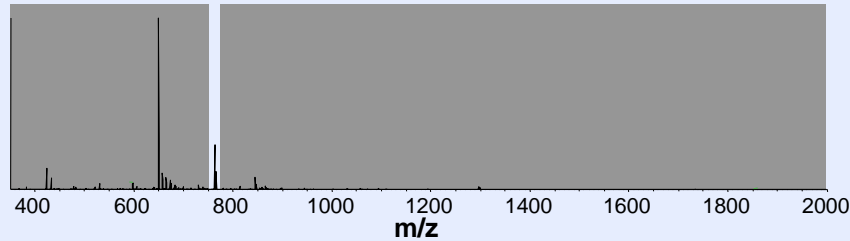
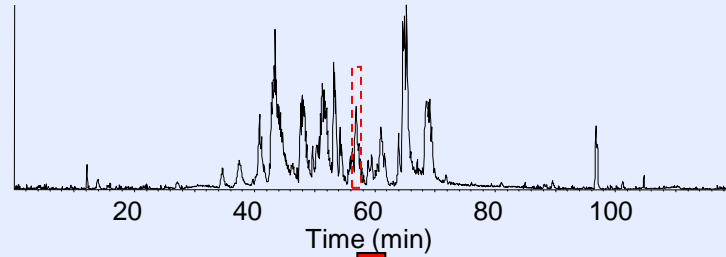
1) Classical SRM with a triple Quadrupole MS



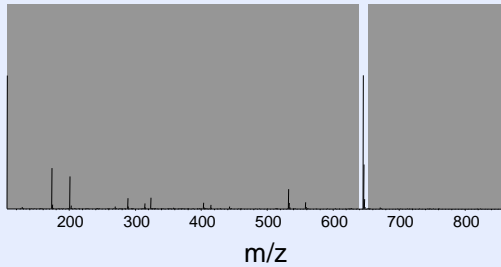
2) Pseudo SRM with an ion trap MS: (1 to 10 MSMS on specific mass)



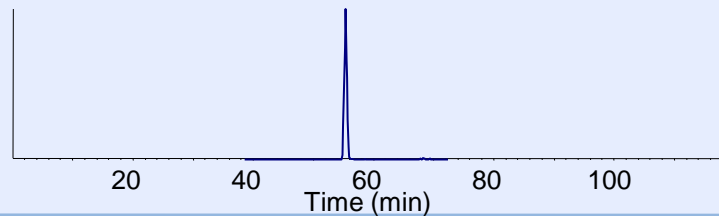
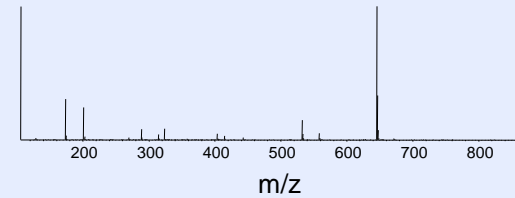
Target approach using Ion Monitoring:



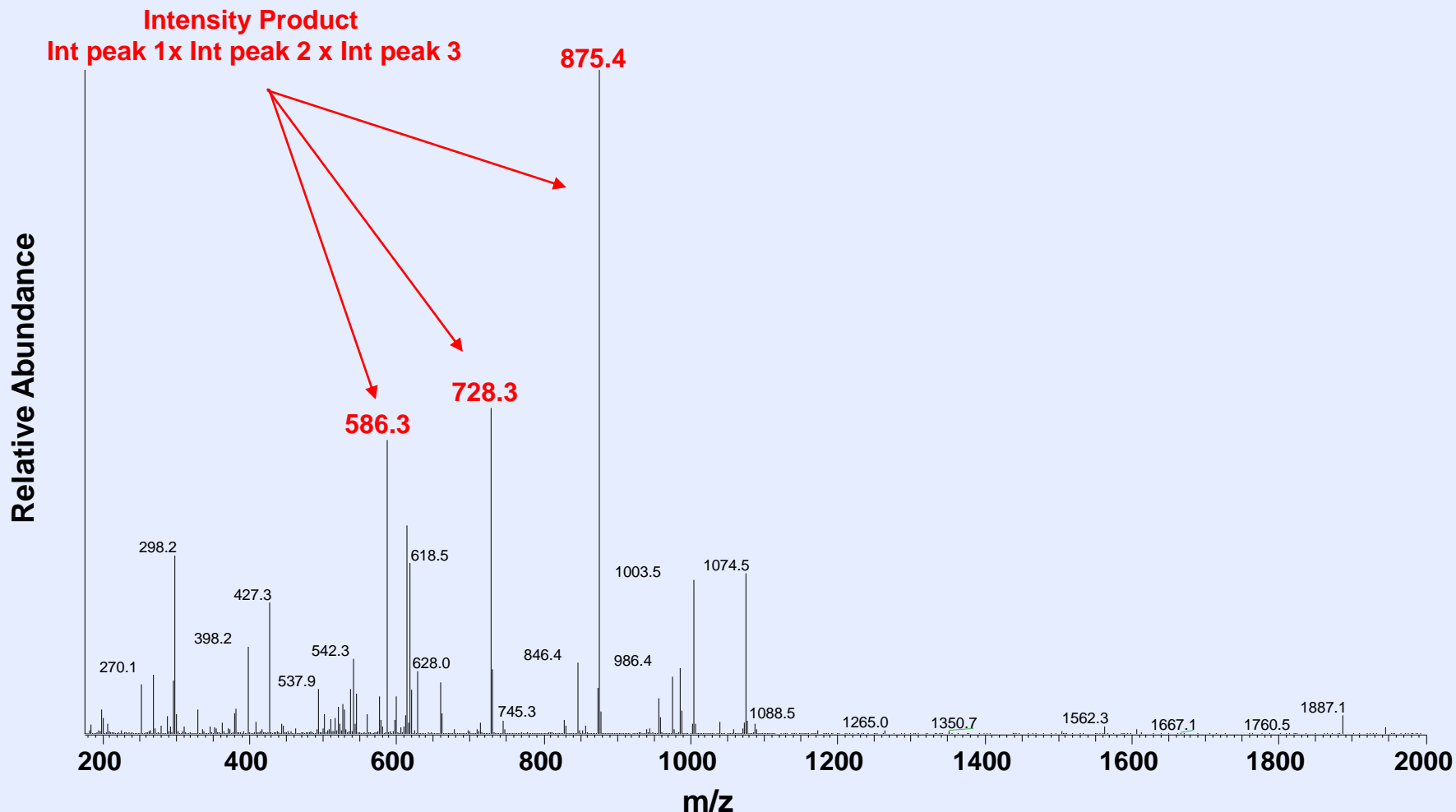
Standard approach



Ion trap based assay



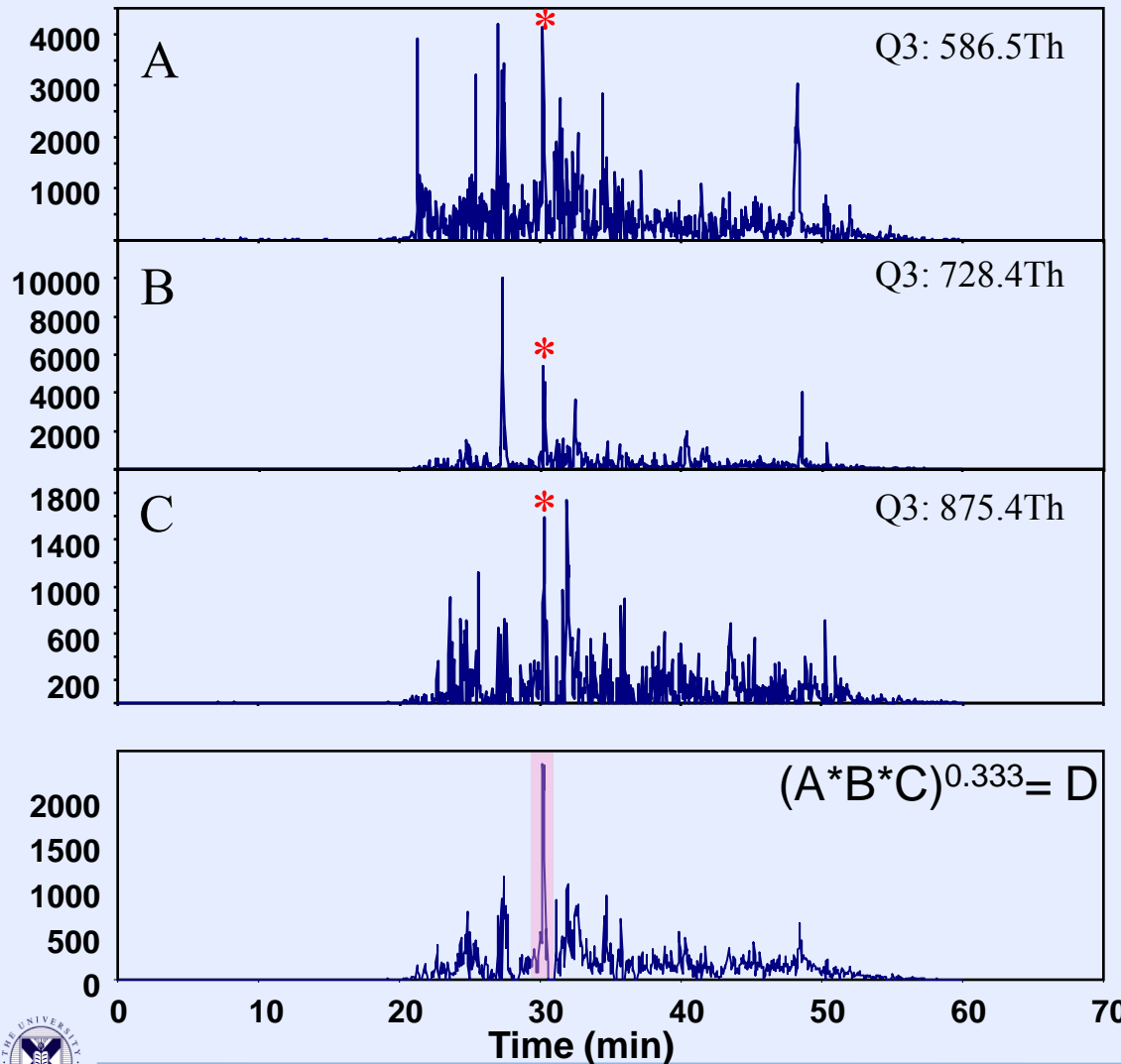
Pseudo-mrm : Signal Processing



*Empirical rules: choosing the most intense fragments but:
y3/b3 and up
at least 1 fragment having a m/z higher than the parent ion*

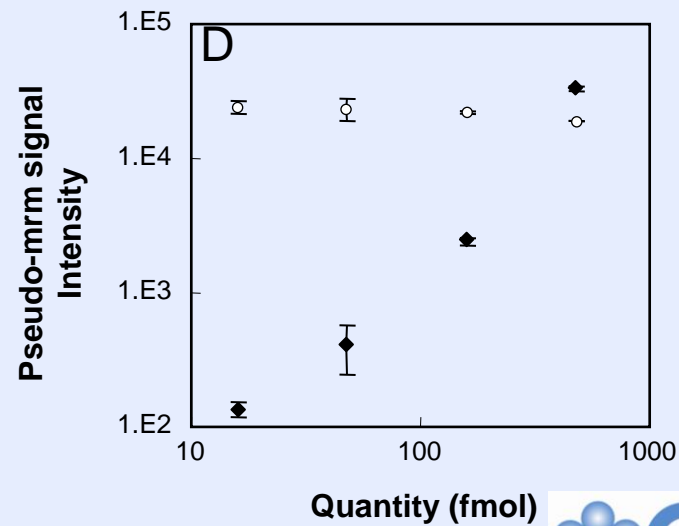
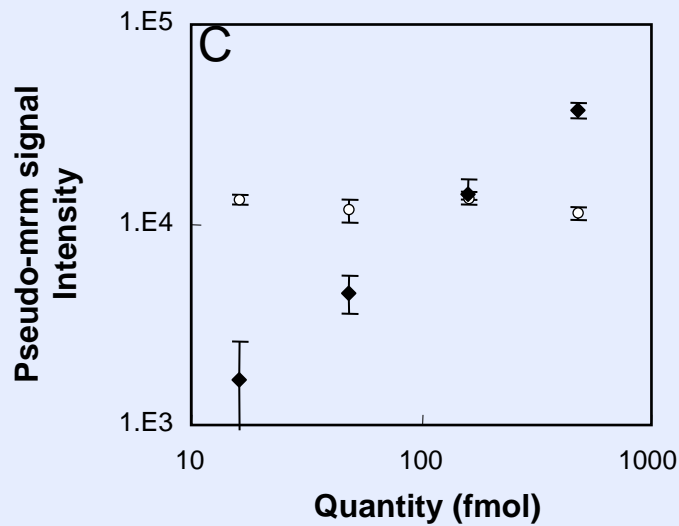
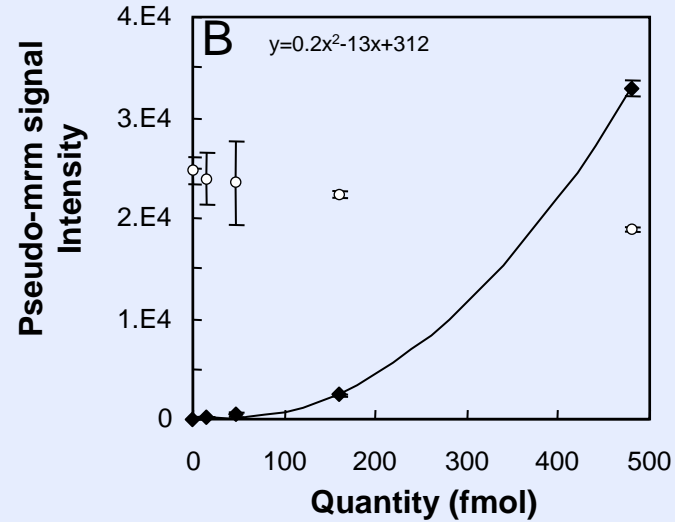
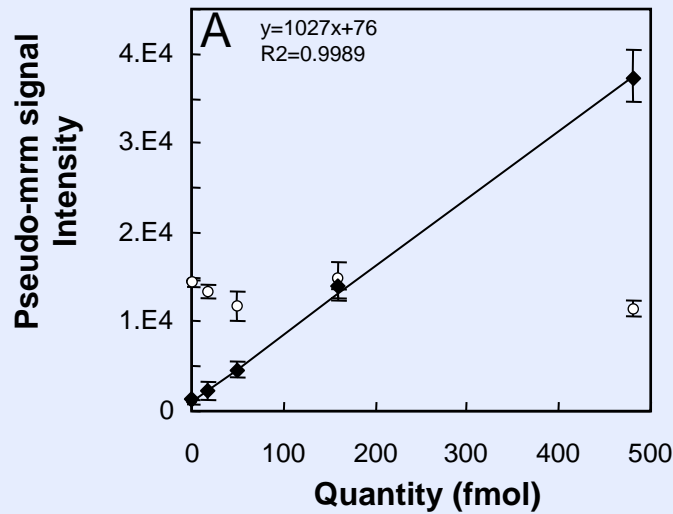
Signal manipulation to increase specificity of the assay

(AQUA peptide spiked at 16fmol in 4ug cell lysate)

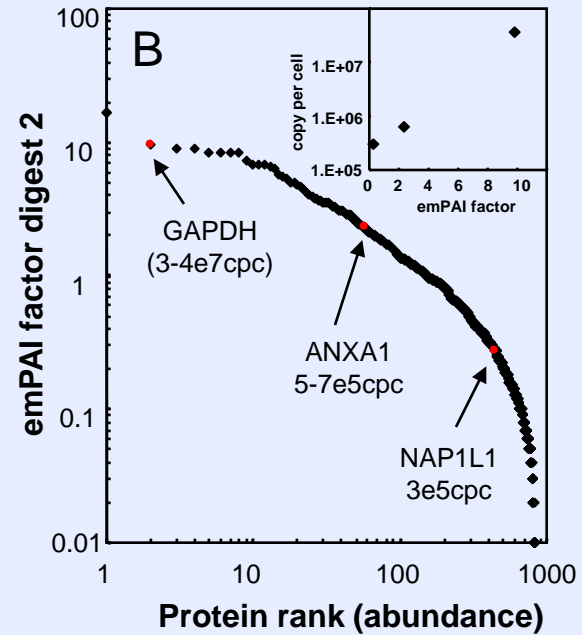
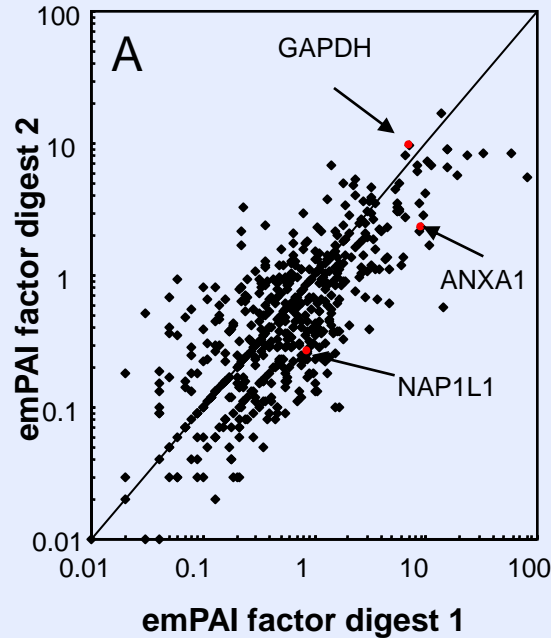


Q1 like 636.8 Th
AQUA peptide TPAQFDADEL**R**

Example of quantitative measurement performed using an ion-trap

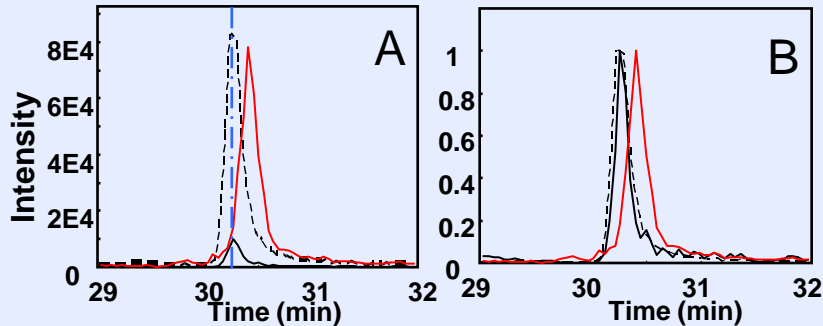


Semi-quantitative analysis using the emPAI factor

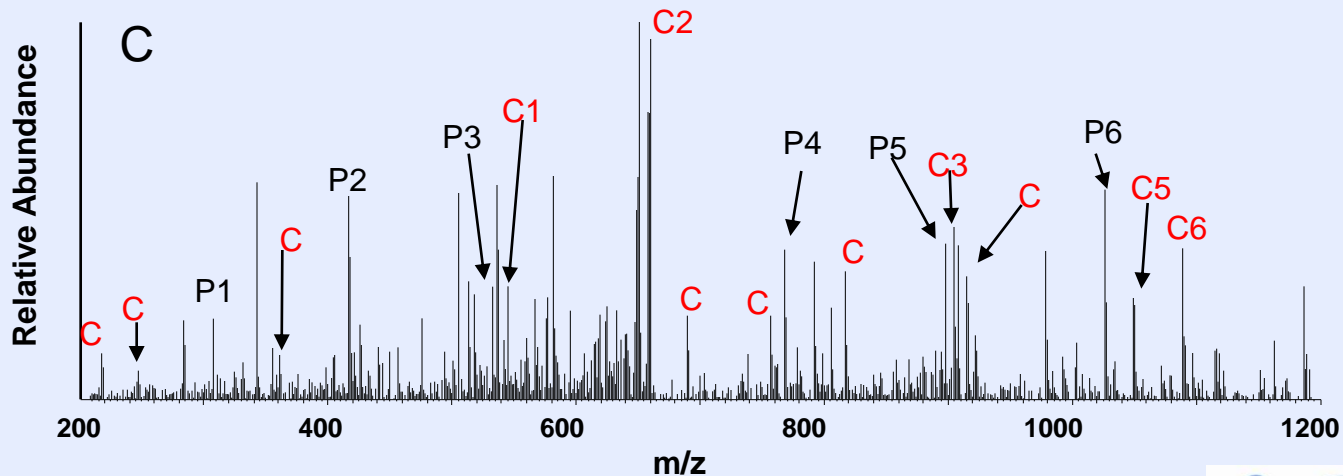


Example of a peptide not identified using a search engine

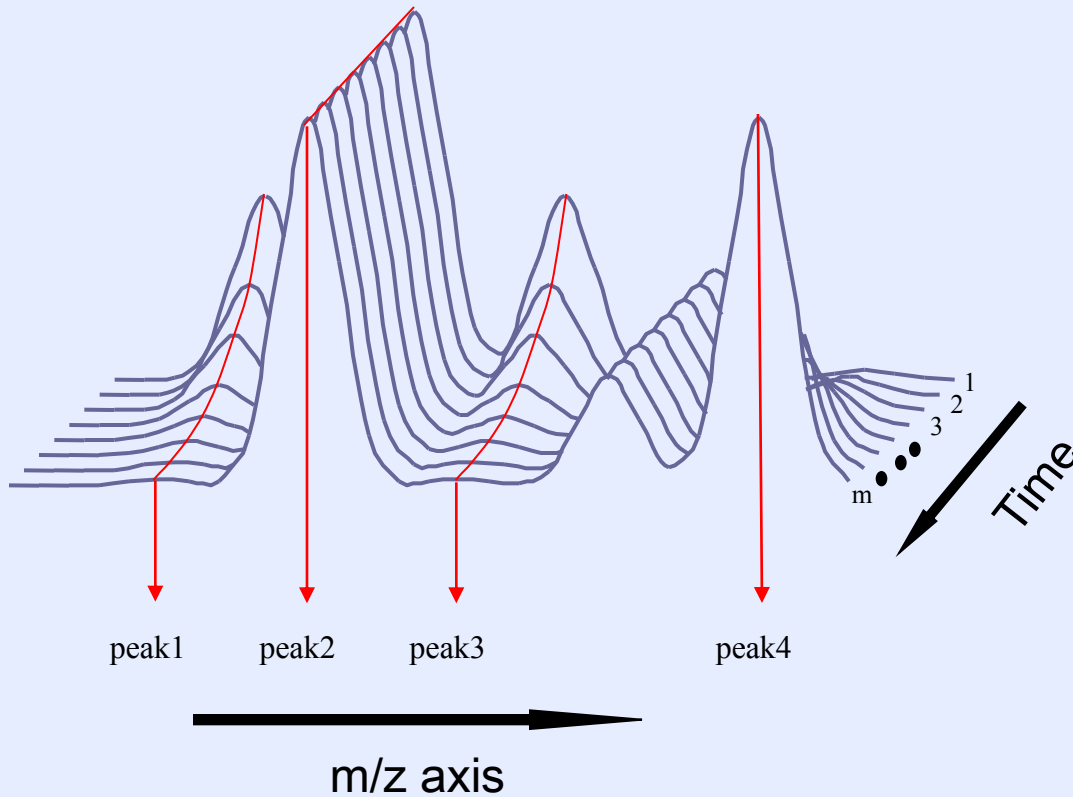
(distinction between peptide detection and identification)



Sequence	Z	m/z	RT _{exp} (min)	RT _{calc} (min)
VFGQSGAGNNWAK	2	668.3	30.42	33.5
FYEEVHDLER (natural)	2	668.8	30.28	33.7
FYEEVHDLER (AQUA)	2	673.8	30.26	33.7



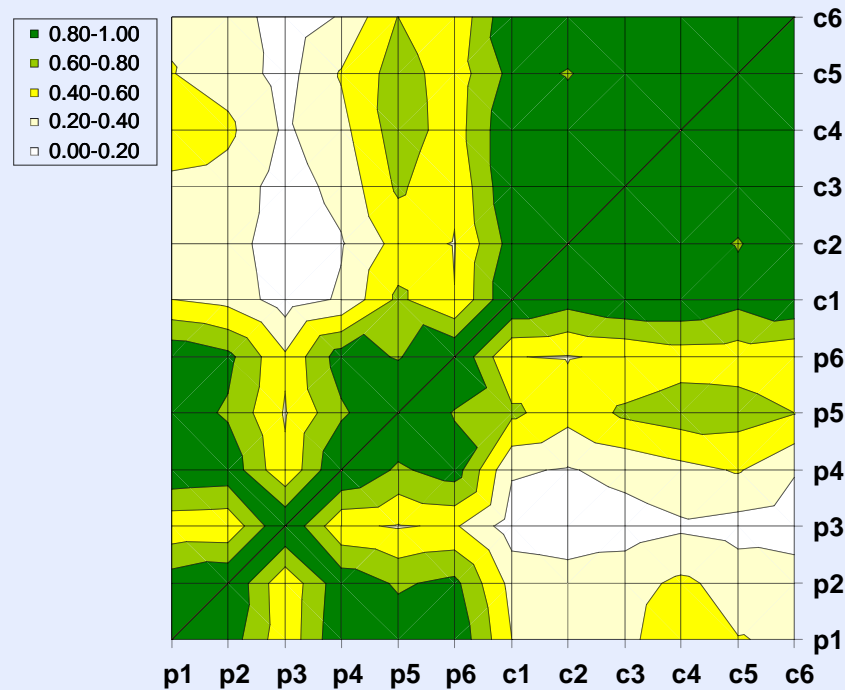
Cross correlation analysis: Brief description of the approach



Correlation

- peak1 vs peak 2: poor corr.
- peak1 vs peak 3: good corr.
- peak1 vs peak 4: anti corr.

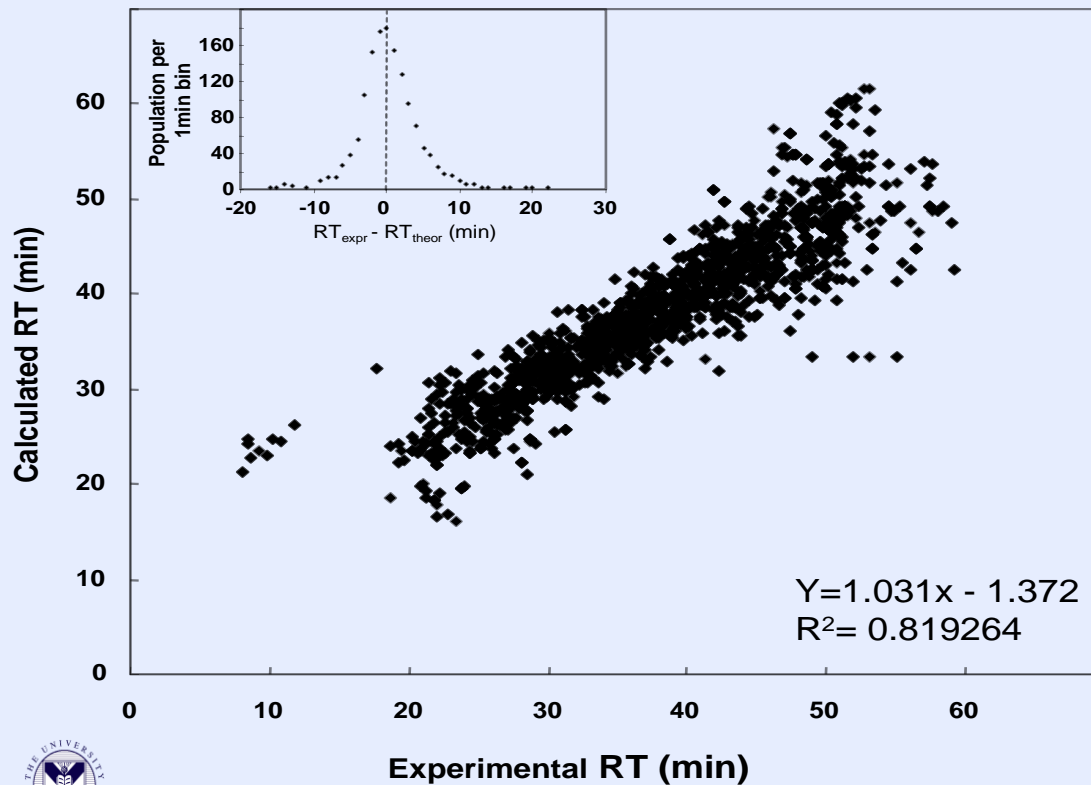
Cross-Correlation analysis between Contaminant, C, and targeted peptide, P



Hydrophobic scale for 1h LC-MS run

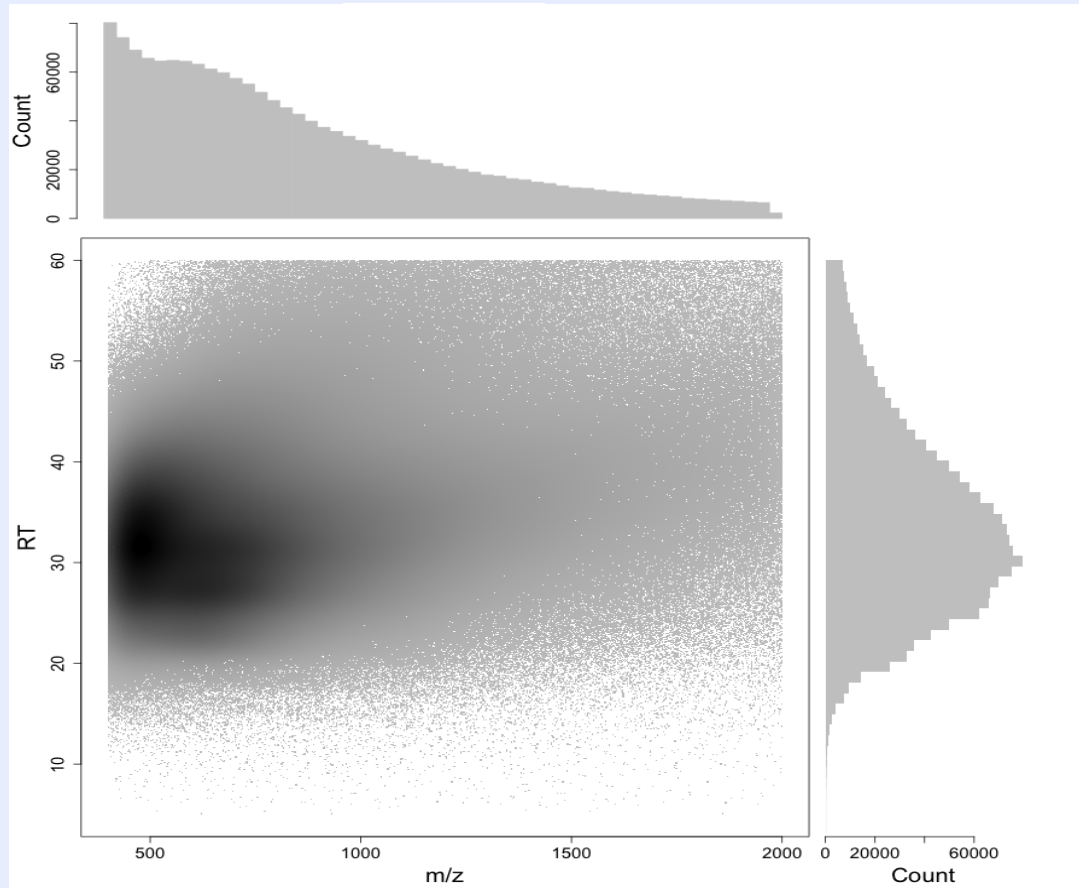
Peptide	System to solve	Time (min)
AVEHINK	A+V+E+H+I+N+K+DT	8.5
IEEELGSK	I+E+E+E+L+G+S+K+DT	21.6
KLNVTEQEK	K+L+N+V+T+E+Q+E+K+DT	17.0
⋮	⋮	⋮

Deduced
Hydrophobic scale



	(Min)
A	0.54
R	-3.11
N	-0.91
D	-0.07
Q	-0.50
E	0.05
G	-0.35
H	-3.39
I	4.06
L	4.36
K	-3.16
M	2.81
F	5.28
P	-0.34
S	-0.20
T	0.35
W	5.59
Y	2.21
V	2.19
C	-0.60
DT	26.10

Human “Trypsinosome” spread in the RT-m/z space



Combination of properties for Peptide detection

Specificity of the method

*Interrogation performed on 1295 experimental peptides
Over 850 000 generated in silico peptides*

Experimental information

seq ₁	m ₁	H _{exp1}	F _{1a,b,c}
•	•	•	•
•	•	•	•
•	•	•	•
seq ₁₂₉₅	m ₁₂₉₅	H ₁₂₉₅	F _{1295a,b,c}

Interrogation process

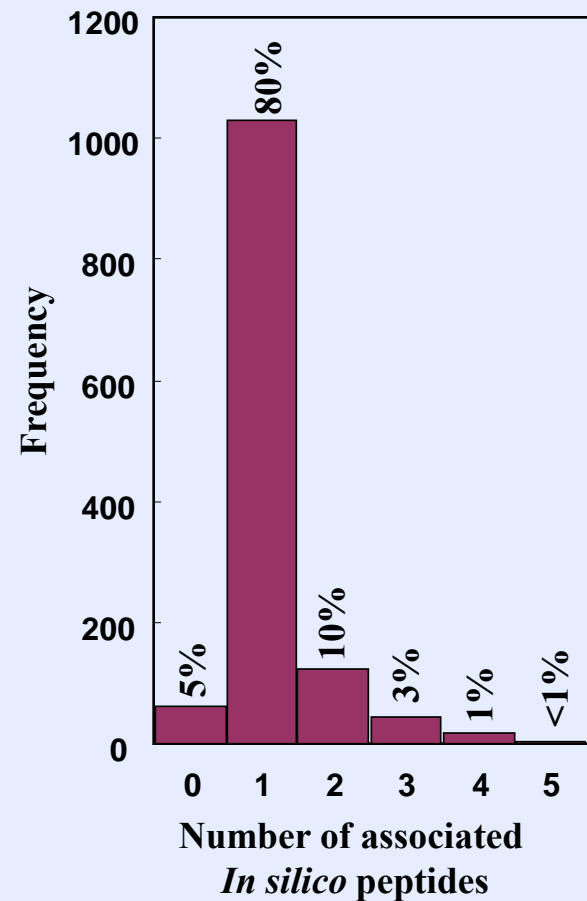


In-silico information

seq ₁	m ₁	H _{calc1}	F _{1a,b,c}
•	•	•	•
•	•	•	•
•	•	•	•
seq _{850k}	m _{850k}	H _{850k}	F _{850k,a,b,c}

*1) Can we identify the right peptide by only using:
m/z, H and 3 MSMS fragments?*

2) Does it give a unique answer?



Summary:

- We have developed methods for ^{15}N labeling for *Ostreococcus tauri* which allow us to quantitatively compare Ot under different conditions.
- We are able to detect some changes associated to the circadian cycle and potentially key elements involved in the control of this process in Ot.
- We have shown an example of ^{15}N incorporation at different time points; We are in the process of establishing a platform that will allow for the determination of the amount of ^{15}N incorporation on a proteome scale.
- Finally: it is possible to detect a peptide without actually identifying it using a search engine.

Acknowledgements

- Sarah Martin: (postdoc: experimentalist and modeler)
- Shanti Munagapati :(Bioinformatics, Programmer)
- Ramon Grima (lecturer; modeler)
- Thorsten Forster: (postdoc biostatistitian and programmer)
- Eliane Chirnside Salvo: (lab technician)

Funding (CSBE, BBSRC/EPSRC)





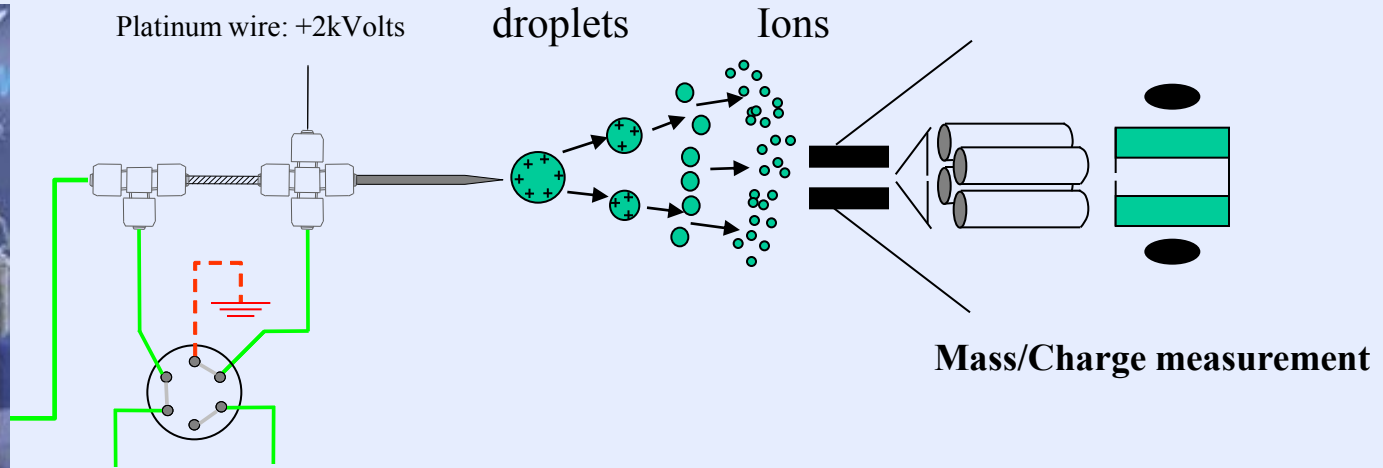


The HPLC-MS component

HPLC



Platinum wire: +2kVolts



MS

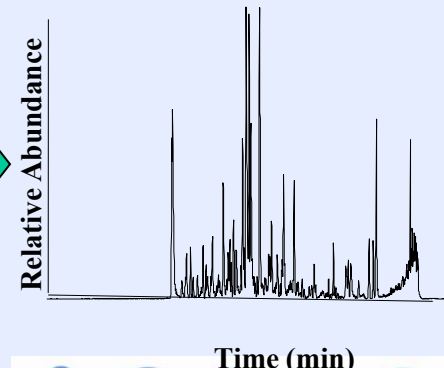
HPLC (separation)

Peptides separated based on their physico-chemical properties



MS (detector)

Peptides are measured by their mass properties (m/z)



Heterotrophe: need some essential amino acid
I,L,K,M,F,T,W,V

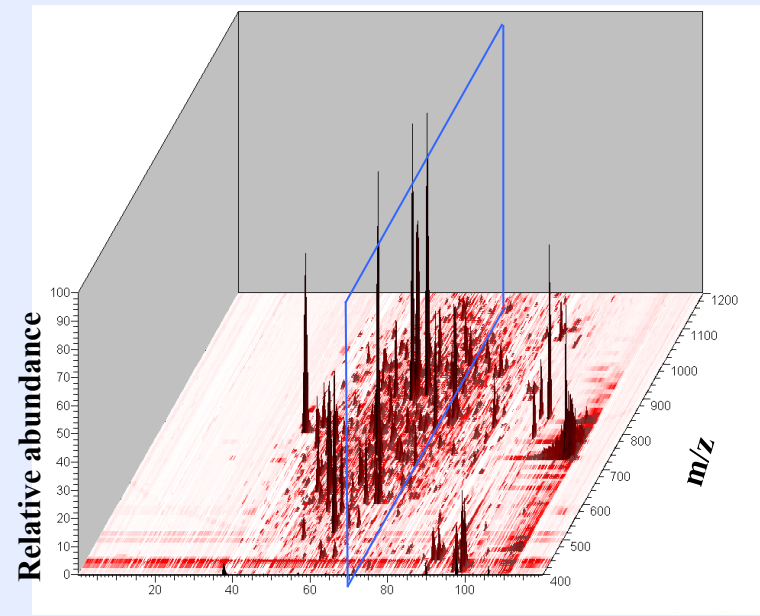
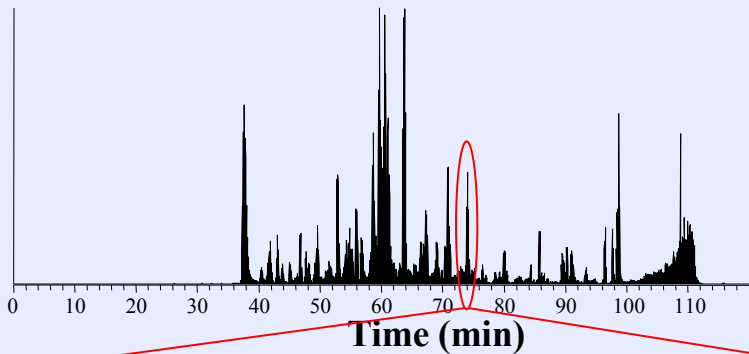
Autotrophe (plant, algae, bacteria): Organism that produce complex molecules from simple inorganic molecules.

Amino acid incorporation vs metabolic labeling mostly dictate by the nature of the studied organism (there are few examples of study done using AA incorporation in autotrophic organism).

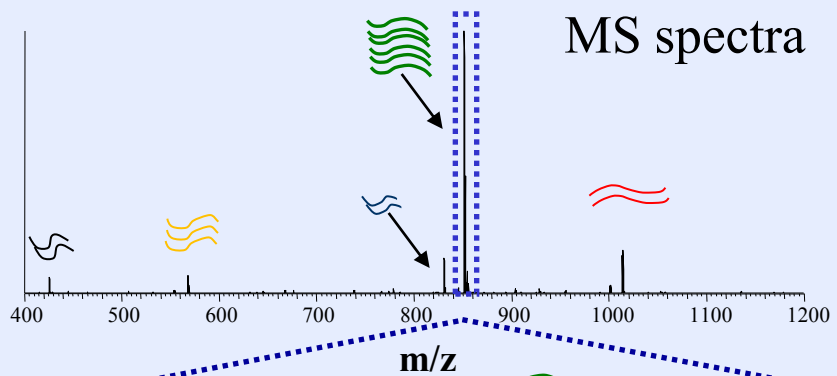
Labeling strategy using ^{13}C glucose, ^{13}C acetate, $^{15}\text{NH}_4\text{Cl}$, $^{15}\text{NO}_3$

MS Signal Deconvolution

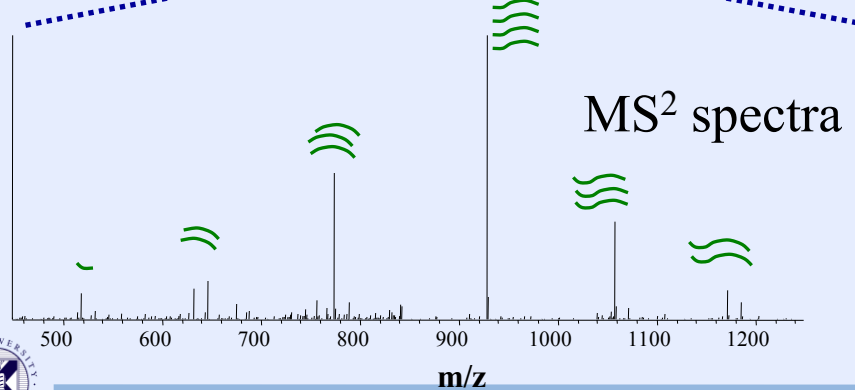
Basepeak chromatogram



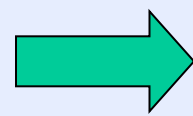
Relative Abundance



MS spectra



MS² spectra



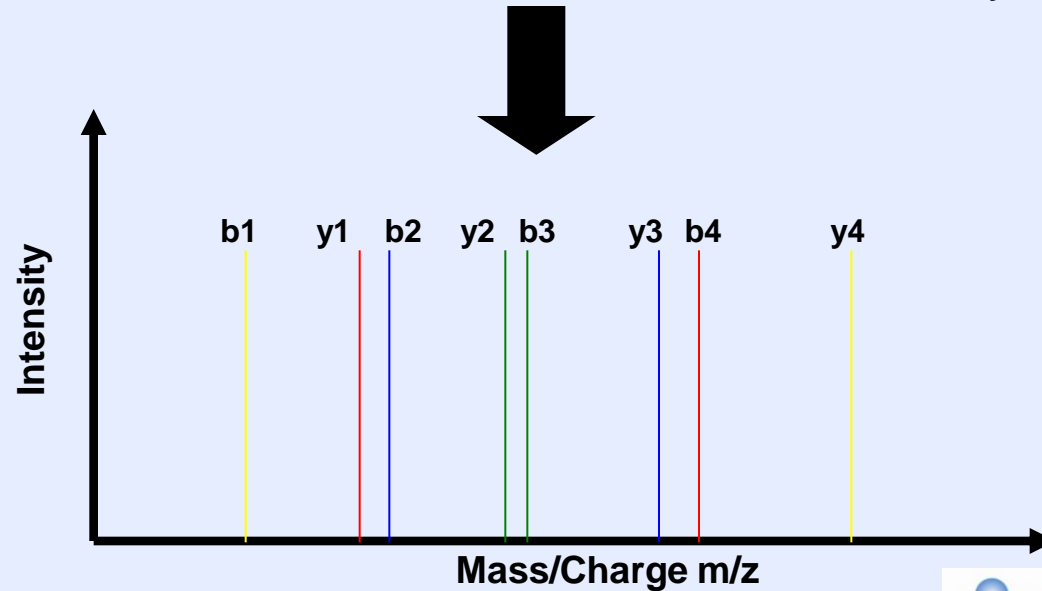
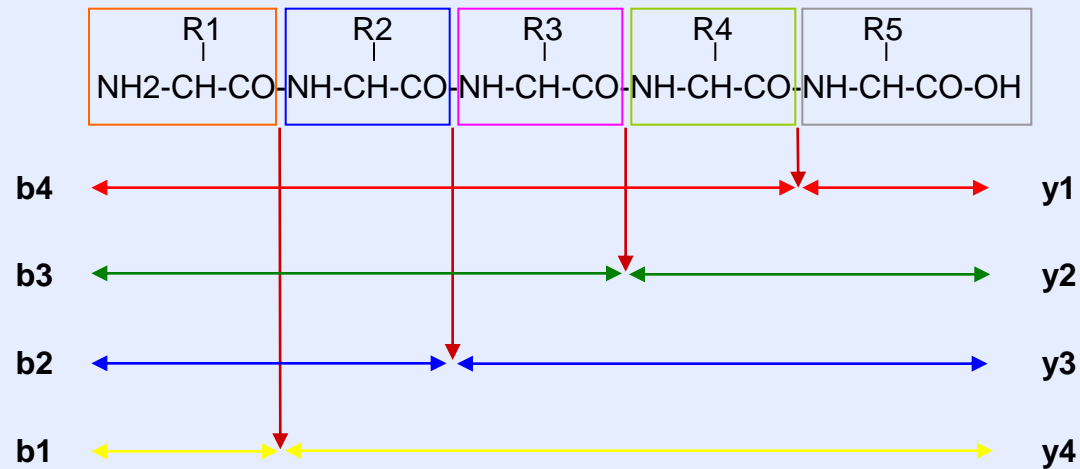
Individual Peptides are clustered into Protein information



Each peptide has an MSMS specific signature



Peptide fragmentation in MSMS mode



Artificial sea water (Keller media)

Compound	Molar Concentration
NaNO₃	8.83 x 10⁻⁴ M
NH₄Cl	3.63 x 10⁻⁵ M
β-glycerophosphate	1 x 10 ⁻⁵ M
Na ₂ SiO ₃ · 9H ₂ O*	1.07 x 10 ⁻⁴ M*
H ₂ SeO ₃	1 x 10 ⁻⁸ M
Tris-base (pH 7.2)	1 x 10⁻³ M
Na₂EDTA · 2H₂O	1 x 10⁻⁴ M
FeCl ₃ · 6H ₂ O	1 x 10 ⁻⁵ M
Na ₂ MoO ₄ · 2H ₂ O	1 x 10 ⁻⁸ M
ZnSO ₄ · 7H ₂ O	1 x 10 ⁻⁹ M
CoCl ₂ · 6H ₂ O	1 x 10 ⁻⁹ M
MnCl ₂ · 4H ₂ O	1 x 10 ⁻⁸ M
CuSO ₄ · 5H ₂ O	1 x 10 ⁻⁸ M