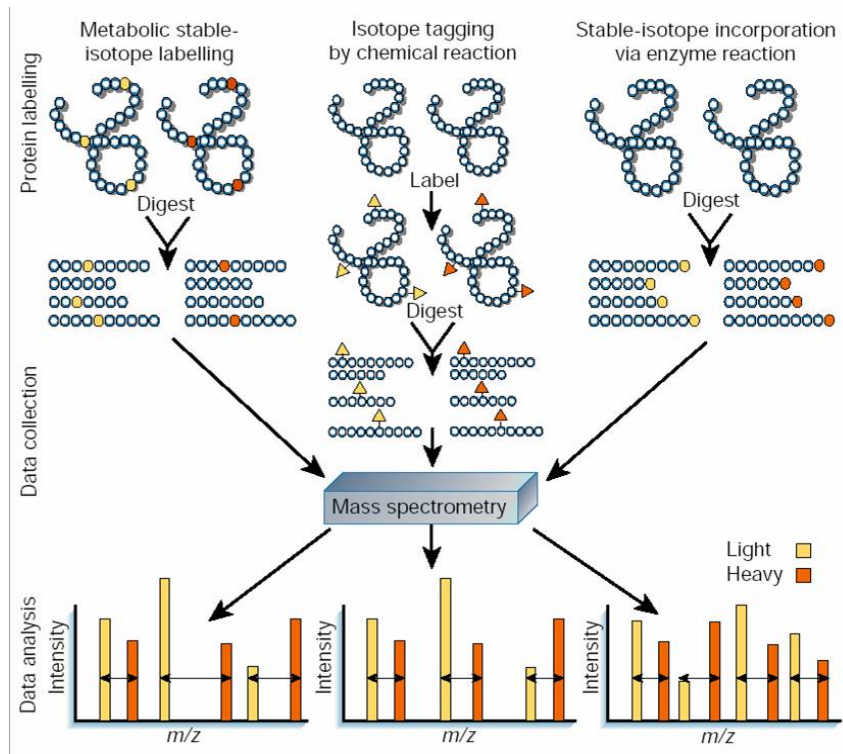


# Part III

Quantitation and file formats

# General Idea of Quantification With Labels



Different ways to introduce labels.

- Label two samples with different isotopes.
- Mix the same amount of samples together and do MS.
- Compare the intensities of peaks in the same scan of MS (or MS/MS).

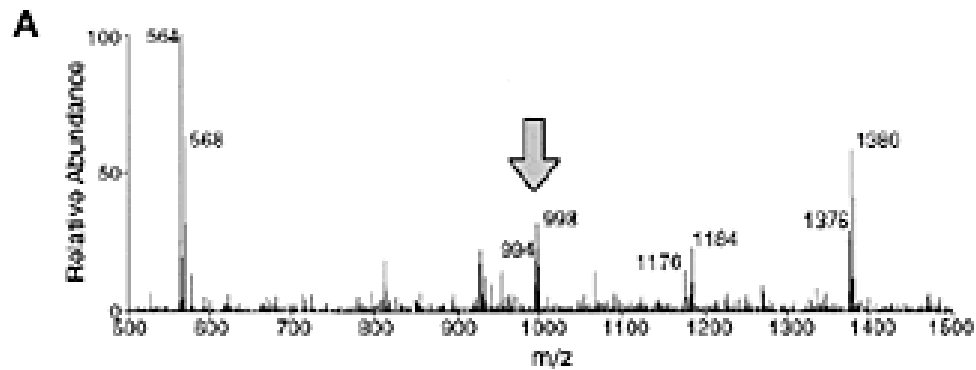
# Isotope coded affinity tag (ICAT)

- 2 Fun ICAT Animations!!

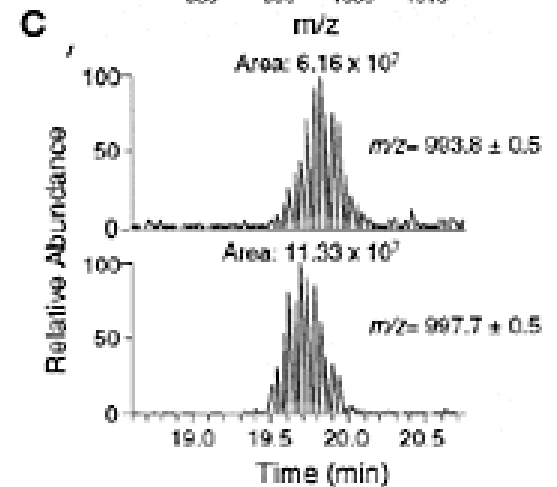
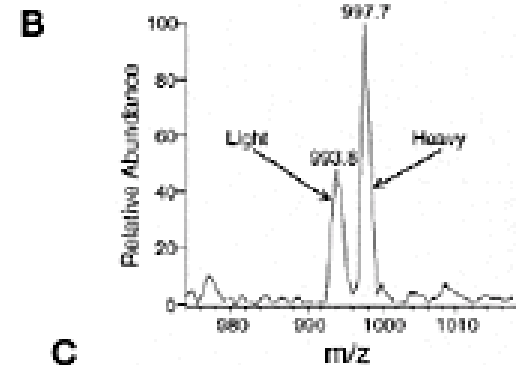
(1) <http://www.bio.davidson.edu/courses/genomics/ICAT/ICAT.html>

(2) <http://www.chemsoc.org/exemplarchem/entries/2002/proteomics/icat.htm>

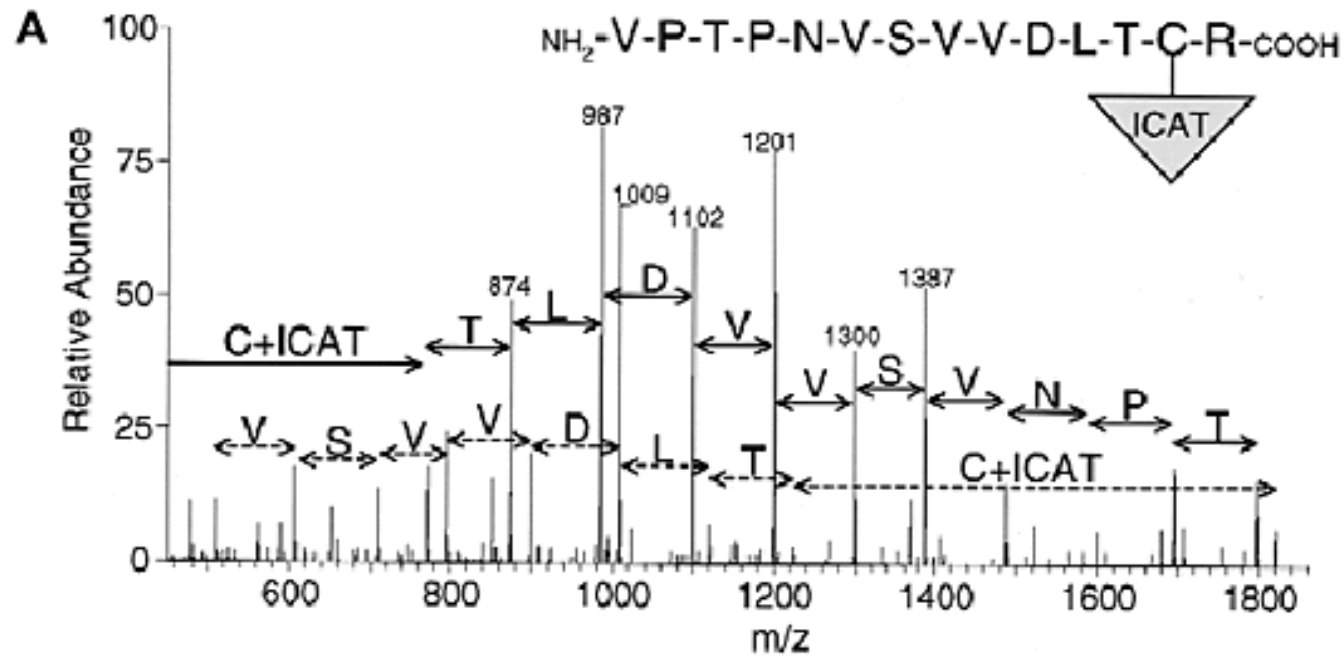
# Real Data



Ratio = 6.16/11.33



# MS/MS to identify sequence



# Compute Protein Ratios

**Table 1. Sequence identification and quantitation of the components of a protein mixture in a single analysis.**

Gene name <sup>a</sup>	Peptide sequence identified <sup>b</sup>	Observed ratio (d0/d8) <sup>c</sup>	Mean ± SD	Expected ratio (d0/d8) <sup>d</sup>	% error
<i>LCA_BOVIN</i>	ALC#SEK	0.94	0.96 ± 0.06	1.00	4.2
	C#EVFR	1.03			
	FLDDLTDDIMC#VK	0.92			
<i>OVAL_CHICK</i>	ADHPFLFC#IK	1.88	1.92 ± 0.06	2.00	4.0
	YPILPEYLQC#VK	1.96			
<i>BGAL_ECOLI</i>	LTAAC#FDR	1.00	0.98 ± 0.07	1.00	2.0
	IGLNC#QLAQAER	0.91			
	IIFDGVNSAFHLWC#NGR	1.04			
<i>LACB_BOVIN</i>	WENGECAQK	3.64	3.55 ± 0.13	4.00	11.3
	LSFNPTQLEEQC#HI	3.45			
<i>G3P_RABIT</i>	VPTPNVSVDLTC#R	0.54	0.56 ± 0.02	0.50	12.0
	IVSNASC#TTNC#LAPLAK	0.57			
<i>PHS2_RABIT</i>	IC#GGWQMEEADDWLR	0.32	0.32 ± 0.03	0.33	3.1
	TC#AYTNHTVLPEALER	0.35			
	WLVLC#NPGLAEIIAER	0.30			

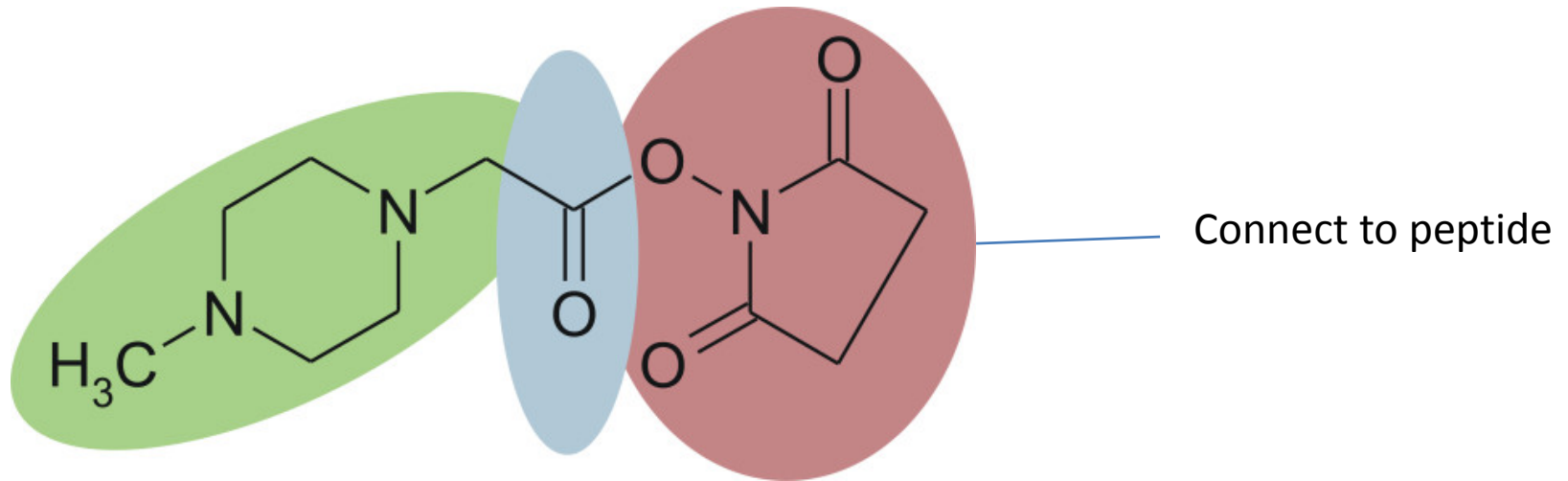
<sup>a</sup> Gene names are according to Swiss Prot nomenclature ([www.expasy.ch](http://www.expasy.ch)).

<sup>b</sup> ICAT-labeled cysteinyl residue is denoted by # sign.

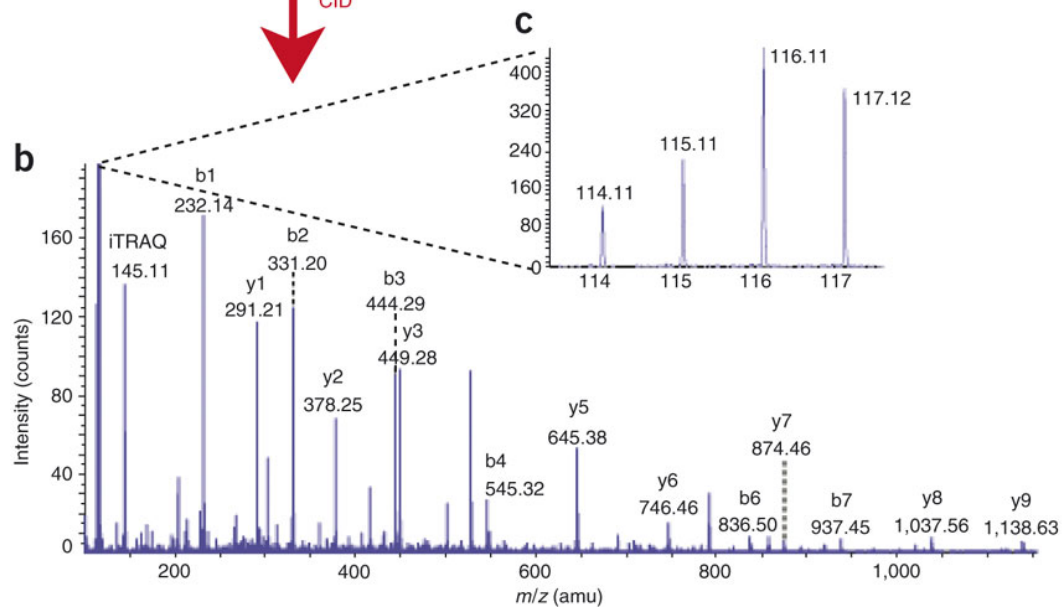
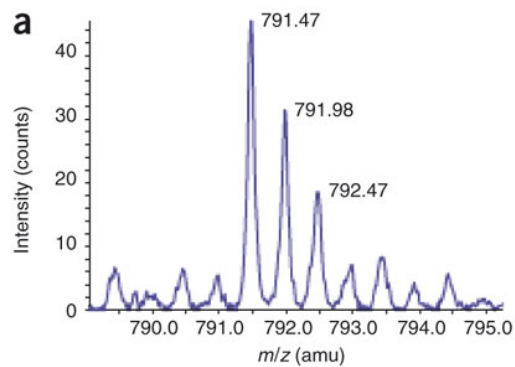
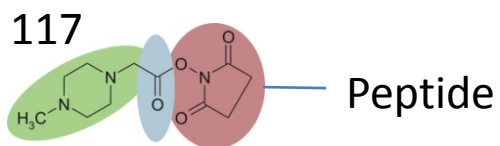
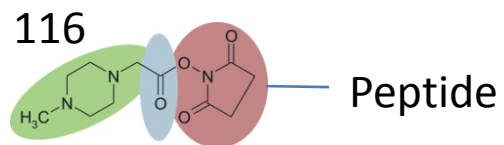
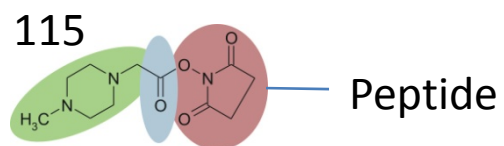
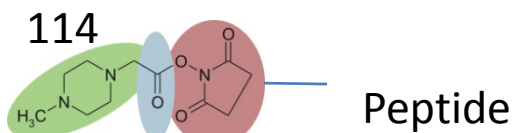
<sup>c</sup> Ratios were calculated for each peptide as shown in Fig. 3.

<sup>d</sup> Expected ratios were calculated from the known amounts of proteins present in each mixture.

# ITRAQ



Chemical structure of the iTRAQ™ reagent. The label is composed of a peptide reactive group (red, NHS ester) and an isobaric tag of 145 Da, which consists of a balancer group (blue, carbonyl group) and a reporter group (green, N-methylpiperazine). The four available tags of identical overall mass vary in their stable isotope compositions such that the reporter group has a mass of 114–117 Da and the balancer of 28–31 Da. The fragmentation site between the balancer and the reporter group is responsible for the generation of the reporter ions in the region of 114–117 m/z.



# File formats

- Instrument specific formats
  - A big headache in proteomics
- Common Text Files
  - PKL
  - DTA
  - mzxml

# PKL

Precursor1: m/z	intensity	z
415.7643	1155.4309	2
60.0632	1.9605	
61.0568	0.2318	
70.0702	757.7263	
.....		
863.4744	14.8326	
867.4798	0.3363	
871.4765	0.8046	
881.4725	3.8688	

Precursor2: m/z	intensity	z
688.0026	1083.5714	3
50.8367	0.0034	
55.9742	0.0027	
57.0060	0.0088	
60.5055	0.0091	
65.0418	0.0159	
71.0407	0.0114	
.....		

Support multiple spectra per file

# DTA

Precursor's MH+ Z

830.528	2
60.0632	1.9605
61.0568	0.2318
70.0702	757.7263
.....	
863.4744	14.8326
867.4798	0.3363
871.4765	0.8046
881.4725	3.8688

Support single spectrum per file

# mzXML

- mzxml is an XML format.
- It contains a lot of information regarding the experiment, instrument, etc.
- The most important information is the scans. Each scan is a spectrum (MS, or MS/MS)

```
<scan num="2"
  msLevel="1"
  peaksCount="39"
  retentionTime="PT930.237007S"
  lowMz="429.0797"
  highMz="830.2746"
  basePeakMz="445.1016"
  basePeakIntensity="3178"
  totIonCurrent="7423">
  <peaks precision="32">Q9aKNUSTQABD1wsvQ+YAAEPXixBDZwAAQ9gMykJ4AABD2IsFQgAAAE
PdkjJByAAAQ94R3UEwAABD3kKWJQAAAAEPeXJlAgAAAQ95uKEBAAABD3o0BRUagAEPfDq9EhcAAQ9+ON0
QtAABD4A7fQx0AAEPgj6lCHAAAQ+cSeUJ8AABD55JhQcAAAEPoEq1BcAAAQ+iUEkCAAABD/A5bQEAAA
QBSdtAAAAARAHJOUHQAABEAgI9QagAAEQCSOJAQAAARAKKo0CgAABEBctpQEAAAQUCcVAAAAARBSJ10
AAAABEF8wFQOAAAQmy5JAgAAARCPPhkBAAAABEKo4NQKAAAEQrDfhAAAAARDzP3UHAAABEPQ+kQUAAA
E Q9TxpBIAAARD2R10AAAABET1EEQbAAAERPkZNBcAAA</peaks>
</scan>
```

# mzXML

- Each spectrum's peak list has a float array
  - [m/z int m/z int m/z int .....
- This is regarded as 4n bytes and encoded with base-64 encoding.
- Each of the 64 printable characters are used to represent 6 bits. Therefore, every 3 bytes can be represented by 4 characters.

ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789+/'