

# **Aspects of liquid chromatography prior to mass spectrometry**

(Prefractionation using Liquid Chromatography)

Mark Raftery  
Bioanalytical Mass Spectrometry Facility  
University of New South Wales  
Sydney, NSW

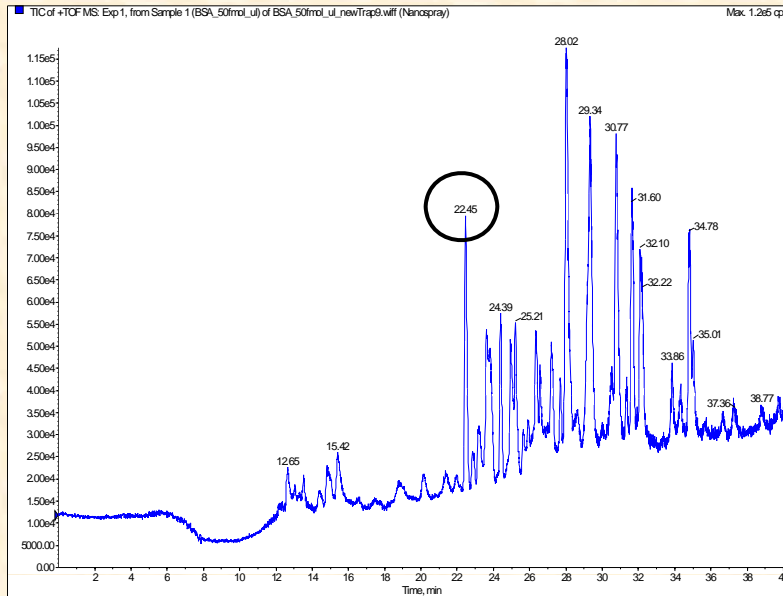
# Many samples do not require prefractionation

- A typical 20-50,000 MW protein (in solution or within a Gel) will probably yield 10-30 tryptic peptides
- All LC/MS systems and data-dependent acquisitions can easily handle this number of peptides and information
- These ~30 different tandem spectra would allow peptides/proteins to be identified after database searches

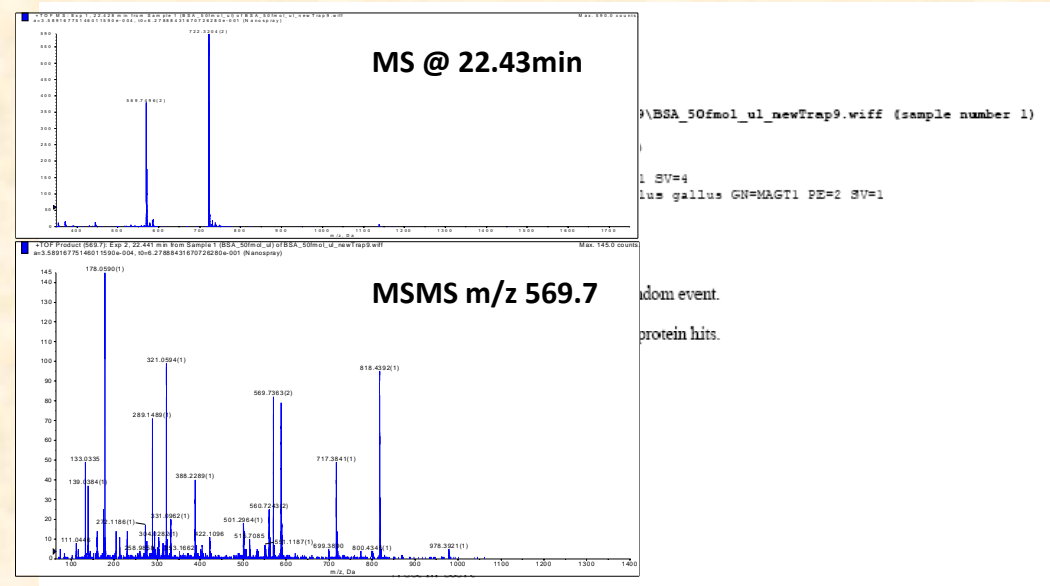
Typical conditions: Digest protein with trypsin, run LC/MS DDA and search using Mascot

# Simple LC/MS of a Single Protein (~100fmol BSA Digest)

Base Peak Ion Chromatogram



Mascot identifies BSA



Variable modifications: Carbamidomethyl (C), Oxidation (M)  
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P  
 Sequence Coverage: 67%

Matched peptides shown in **Bold Red**

```

1 MKWTFISLL LFSAYSRG VFRDTHKSE IARRFKDLGE EHFKGLVLIA
51 FSQYLQCFE DEHVKLVELL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCKQEP ERNECFLSHK DDSPLPKPK FDPNTLCDEF
151 KADEKKFWGK YLYBIARRHP YFYAPELLYY ANKYNGVQE CCQARDKGAC
201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
251 FVEVTKLVD LTKVHKECCCH GDLLECADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVERDA IPENLFPLTA DFAEDKDVCK NYQEAKDAPL
351 GSFLYEYSRR HPEYAVSVLL RLAKEYMATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCDQFEK LGEYGFCNAL IVRYIRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPSBRMP CTEDYLSLIL NRLCVLHEKT FVSEKVTKCC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TPHADICTLP DTEKQIKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 STQTALA
    
```

Good sequence coverage

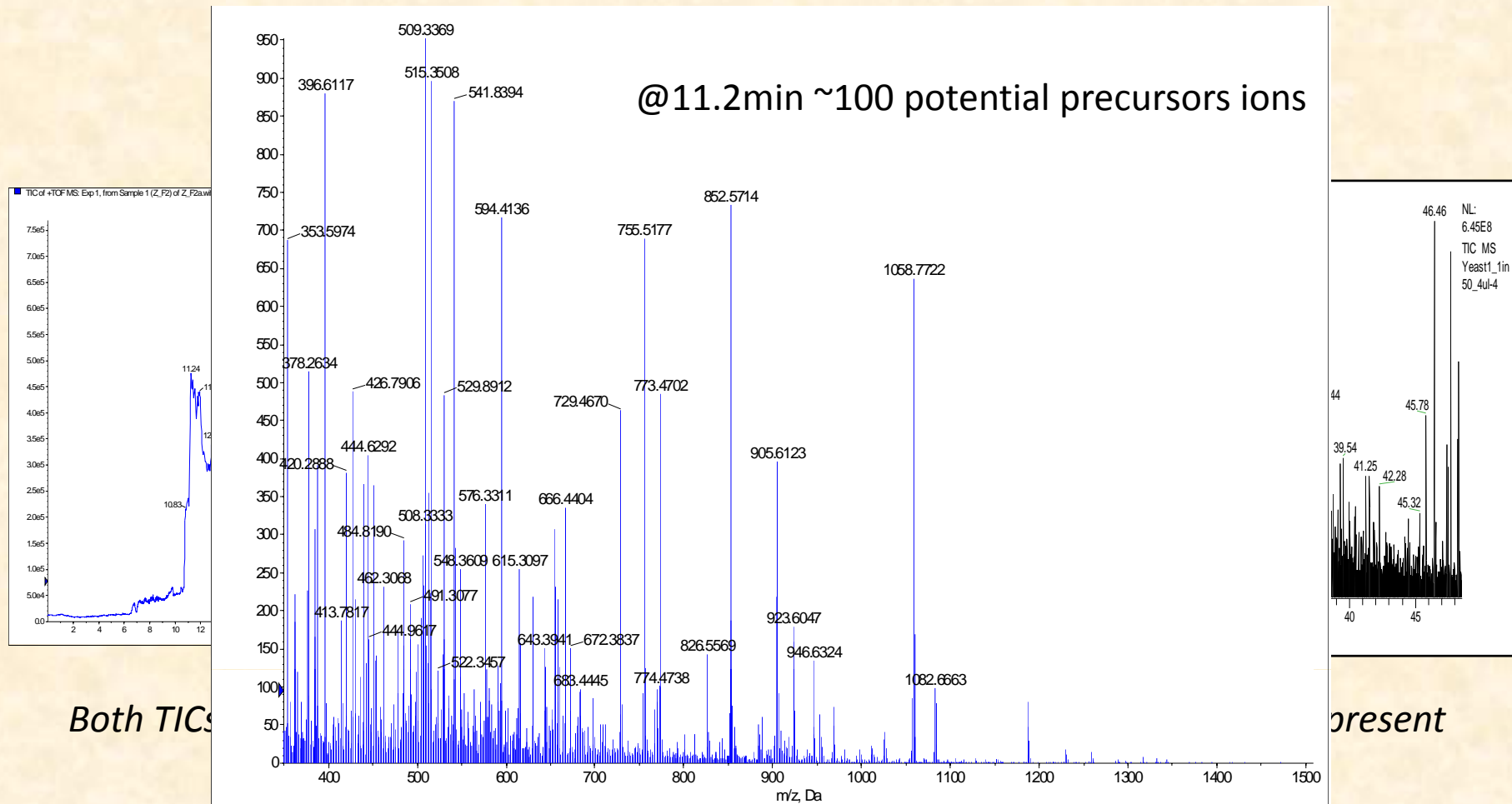
# A Complex Sample is Different

- Complex samples may be digests of total cell lysates or plasma
- These samples may contain many 100's or 1000's of proteins - and may yield many 1000's or 10,000's of peptides

Yeast has ~4,500 proteins ranging from very low (a few copies per cell) to very high abundance (many 10,000s of copies cell)

Need to analyse these types of samples

# Total ion chromatograms of a complex tryptic digests



# Analysing increasing amounts tryptic peptides from a complex sample (1D)

- A yeast digest was injected onto a LTQ Velos with different amounts of tryptic peptides using nanoC<sub>18</sub> LC and different separation times (30 and 60 min) followed by Mascot searches

- **Volume**

- **~Quantity**

- 0.01 µl
- 0.02 µl
- 0.04 µl
- 0.08 µl

- 20 ng
- 40 ng
- 80 ng
- 160 ng

100 fold amount range

- 0.1 µl
- 0.5 µl
- 1.0 µl
- 2.5 µl
- 5.0 µl

- 100 ng
- 200 ng
- 400 ng
- 1200 ng
- 2000 ng

***Does the number of identifications increase by 100?***

# LC MS Experimental



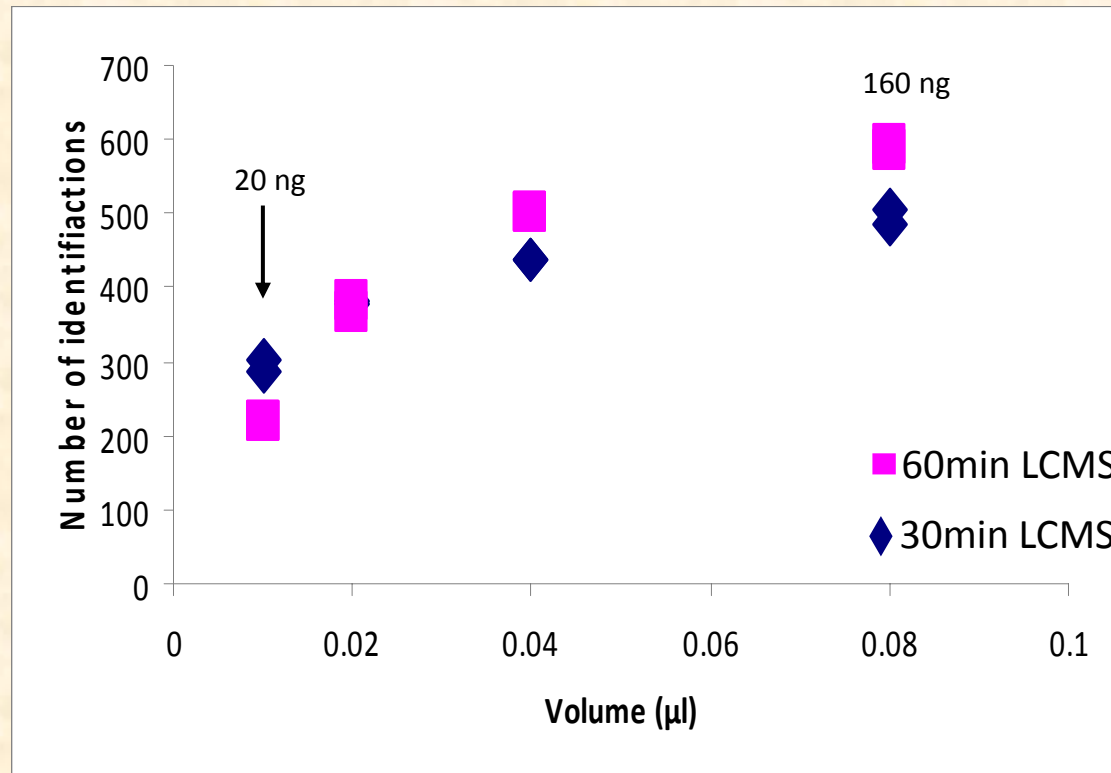
NanoLC: **LC Packings Ultimate**  
~300nl/min, C18 (75 $\mu$  x ~10cm) 5 micron. 2% to 36% acetonitrile (0.1% formic acid) in 30 or 60 min



Mass spectrometry: **LTQ Velos**  
Top 10 method (survey scan m/z 350-1750 followed by 10 possible MSMS, dynamic exclude for 30 sec

Data was search using Mascot

# Protein identifications from 1D LC MS of yeast tryptic peptides (30 and 60min)



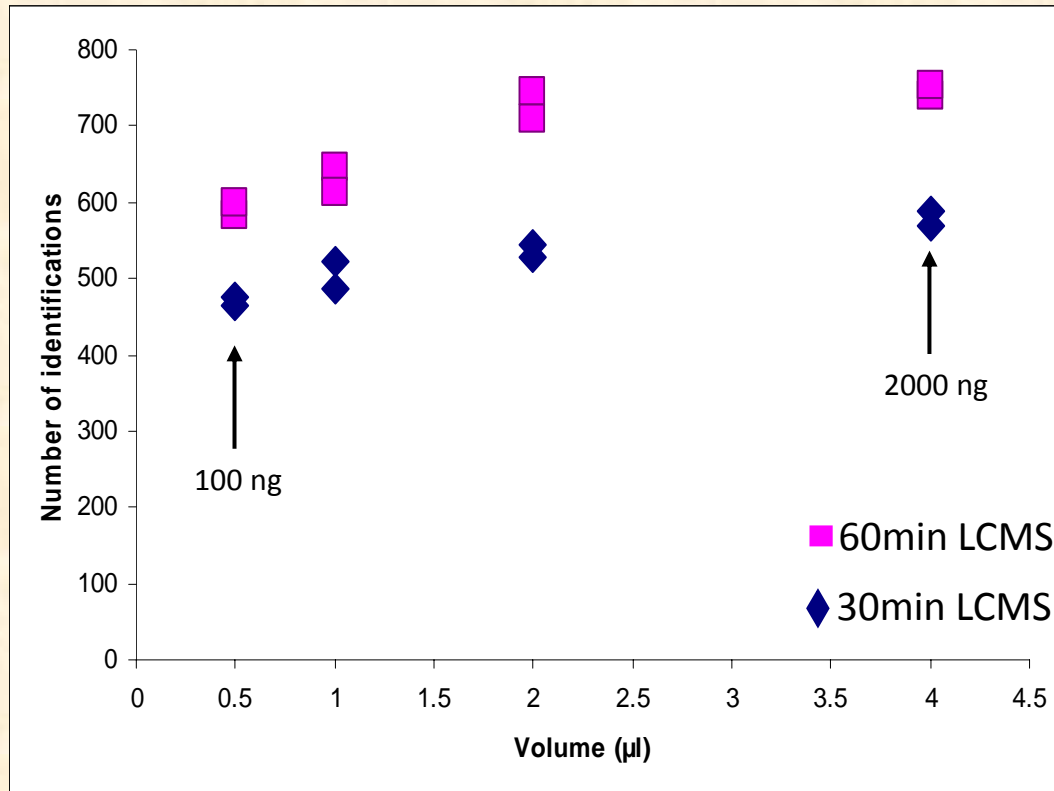
Doubling injection volume (amount) does not double number of proteins identified

Approx 5 x the sample must be injected

**More peptides are identified from the same protein (ie greater coverage) and these are normally the more abundant proteins**

Doubling the gradients time increases the number of identifications (by ~25%)

# Protein identifications from 1D LC MS of yeast tryptic peptides



Amount  
20 ng (0.05 μl)  
2000 ng (4 μl)

Identifications  
~250  
~700

See also. Systematical optimization of reverse-phase chromatography for shotgun proteomics. Xu P, Duong DM, Peng J. J Proteome Res. 2009 Aug;8(8):3944-50.

A similar trend is observed with even more quantity injected, however the number of proteins identified does not increase linearly (it plateaus)

100 x the sample injected resulted in only ~3 x the number of identifications

Doubling the gradient time increases the number of identifications (by ~25%)

# **Different methods are needed to analyse complex tryptic digests to maximise the number of protein identified**

## **Sample fractionation can help in 2 ways**

- Increase the number of proteins identified
  - Greater depth of proteome (lower abundance)
- Also expect increased sequence coverage of proteins
  - More chance to locate modifications

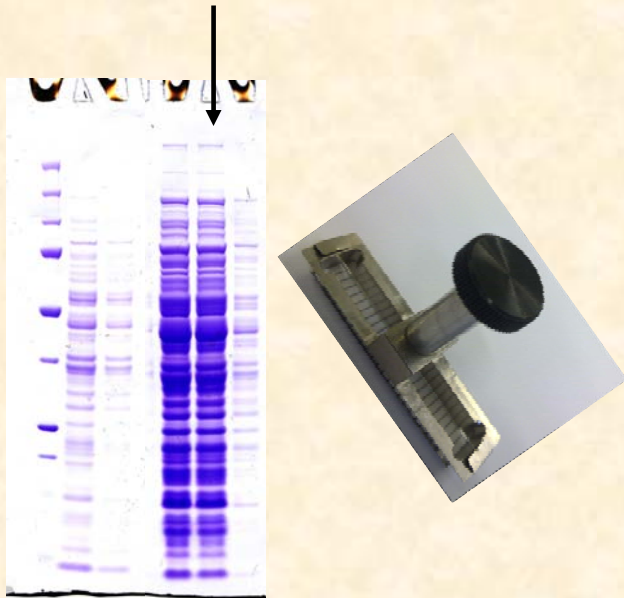
# There are Two Common Approaches

- **Separate the proteins before digestion**
  - Gel separation (GeLC/MS)
  - Column chromatography
  - Affinity chromatography
  - etc
- **Separate the peptides after digestion**
  - Mudpit (2D LC-MS, commonly SCX and RP)
  - Titanium enrichment
  - IEF
  - etc

***Both are widely used and successful***

# Protein Separation GeLC/MS

- GeLC/MS – separates proteins based on size (most common method)



10-25 slices are cut from a gel lane  
These are digested with trypsin  
Each of the samples is analysed  
using nanoLC/MS  
Individual results are combined for  
a final Mascot search

Helps clean up the sample as well as reducing the complexity of the mixture.

Normally obtain many more proteins and greater sequence coverage

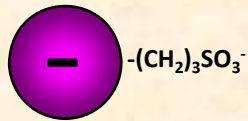
Analysis time increases with the number of slices

# Protein Separation using Column Chromatography

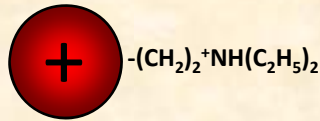
- Many different types of column packing materials are available
  - strong cation exchange (SCX),
  - anion exchange (WAX, SAX)
  - reverse phase (C4, C18) separations are possible
- Enrichment may be focused to help eliminate unwanted protein
  - Eg Lectin columns for enriching glycoproteins
  - Antibody column

# Types of Stationary phases

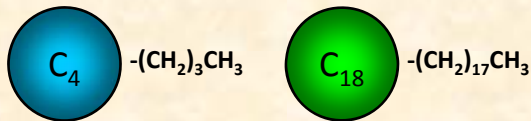
Proteins may be captured and separated by 3 common types of chromatography



Strong cation exchange (SCX) resins are negatively-charged particles (sulfonic acid) that bind positively-charged proteins



Weak anion exchange (WAX) resins (Diethylaminoethyl, DEAE) are positively-charged particles that bind negatively-charged proteins



$\text{C}_4$  or  $\text{C}_{18}$  resins contain covalently-linked alkyl chains that bind proteins by hydrophobic-type interactions

Buffers will normally need to be optimised for a particular stationary phase to allow binding. Ion exchange only works well if samples contain low mM salt

# Column Types

The type of column used will depend on the stationary phase, solvents, flow rates, back pressure and application.  
And what is commercially available



**Low pressure (manufactured from glass or plastic)**

**High pressure (manufactured from stainless steel)**



# Typical Conditions

All will normally require HPLC equipment capable of forming gradients & UV detection + fraction collector



## Reverse phase

Solvent: A. Water (0.1% TFA) B. Acetonitrile (0.1% TFA)

Gradient: 20% to 80% B (over 30-60 min)

Flow: 0.2-1 ml/min

Column: C4 or C18 5 $\mu$ , 300 Å, 2 x 150 mm or 4.6 x 250 mm  
(eg Waters, Grace, Agilent, Thermo...)

## Strong cation exchange (SXC)

Solvent: A. Water (5mM NaPO<sub>4</sub>) B. Water (5mM NaPO<sub>4</sub>, 500mM KCl) + acetonitrile (25%) pH 2.8-3.

Gradient: 0% to 100% B (over 20-40 min)

Flow: 0.2-1 ml/min

Column: eg PartiSphere 5 $\mu$ , 4.6 x 250 mm (Whatman)

## Strong anion exchange (SAX)

Solvent: A. Water (20mM Tris) B. Water (20mM Tris, 1M NaCl) + acetonitrile (25%) pH 8.

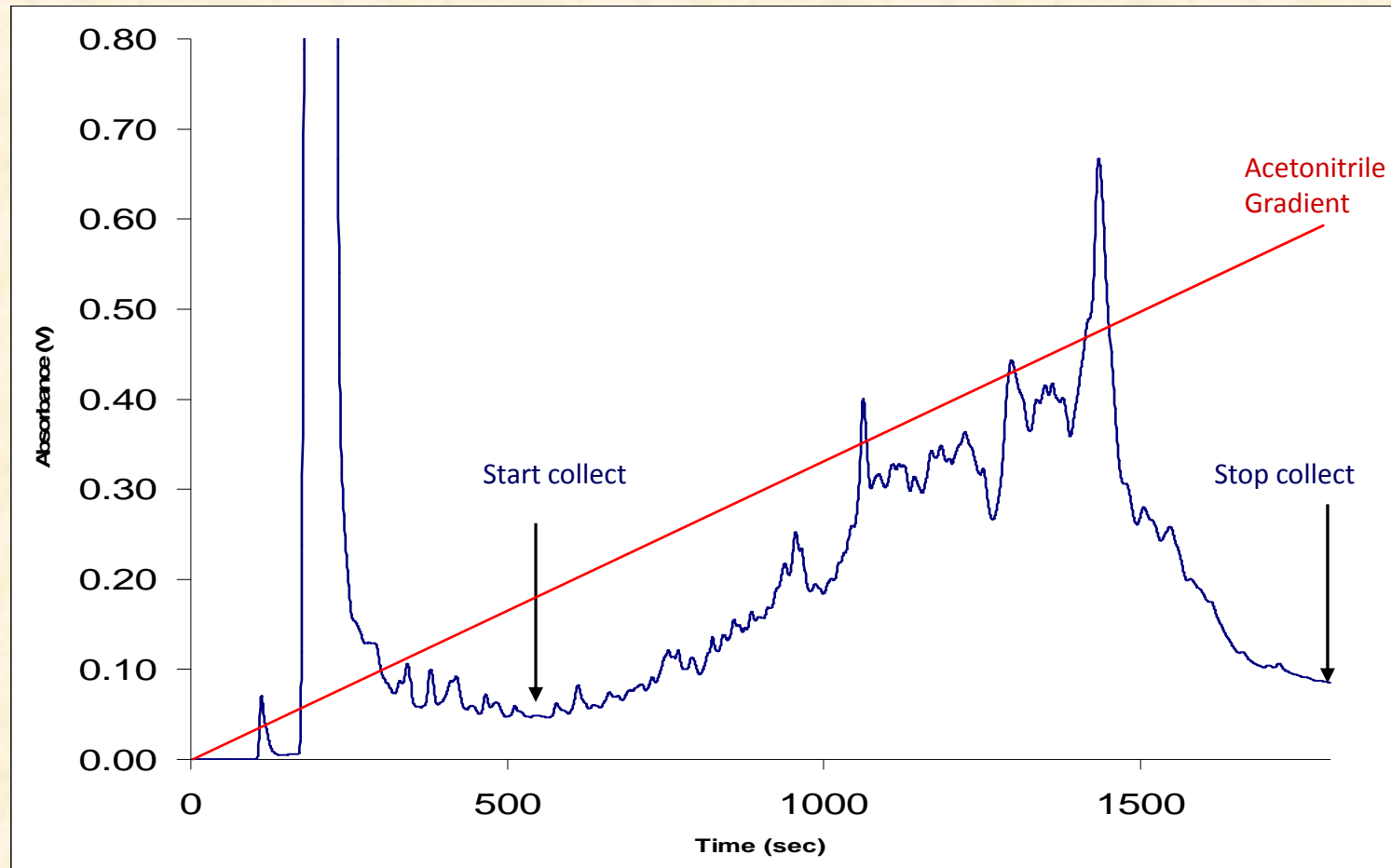
Gradient: 0% to 100% B (over 20-40 min)

Flow: 0.5-1 ml/min

Column: eg MonoQ 10 $\mu$ , 10 x 100 mm (GE Lifescience)

(Note: Low ionic strength needed to effectively bind proteins with ion exchange chromatography)

# HPLC separation of yeast proteins



Fractions collected at 1 minute intervals starting at 480 sec until end of run  
~21 fractions dried down, trypsin digested and analysed by nanoLC/MS



# LC/MS of Collected Fractions

- 21 fractions were collected over 30min C4 LC run
- These fractions were digested with trypsin
- LC/MS DDA obtained of each fraction @ 52 min (~18 hr MS time)
- Peak lists/Mascot searches 5-20 min each depending on program and search database (3-10hr)
- ~218,000 queries
- When combined ~1,680 unique identifications (more analysis possible but false discovery rates ~2%)
- Number of proteins identified more than doubled when proteins were separated before digestion
- 45% average increase in coverage (compared to 1D proteins)

1D LC/MS gave ~700 protein

Protein separation before 1D LC/MS gave ~1700 protein

# Summary

- 1D LC/MS is great for simple or moderately complex samples
- Greatly increasing amounts loaded or longer LC/MS run times does not increase (lineally) the numbers of peptide identifications
- Proteins (or peptides) separated before LC/MS reduces the complexity allows many more proteins to be identified
- Large amounts of time required to collect and analyse data
- May be the only way to increase proteome coverage

# Acknowledgements

## Funding

- UNSW
- ARC Lief
- NCRIS (NSW Government contribution)
- Cancer Institute NSW