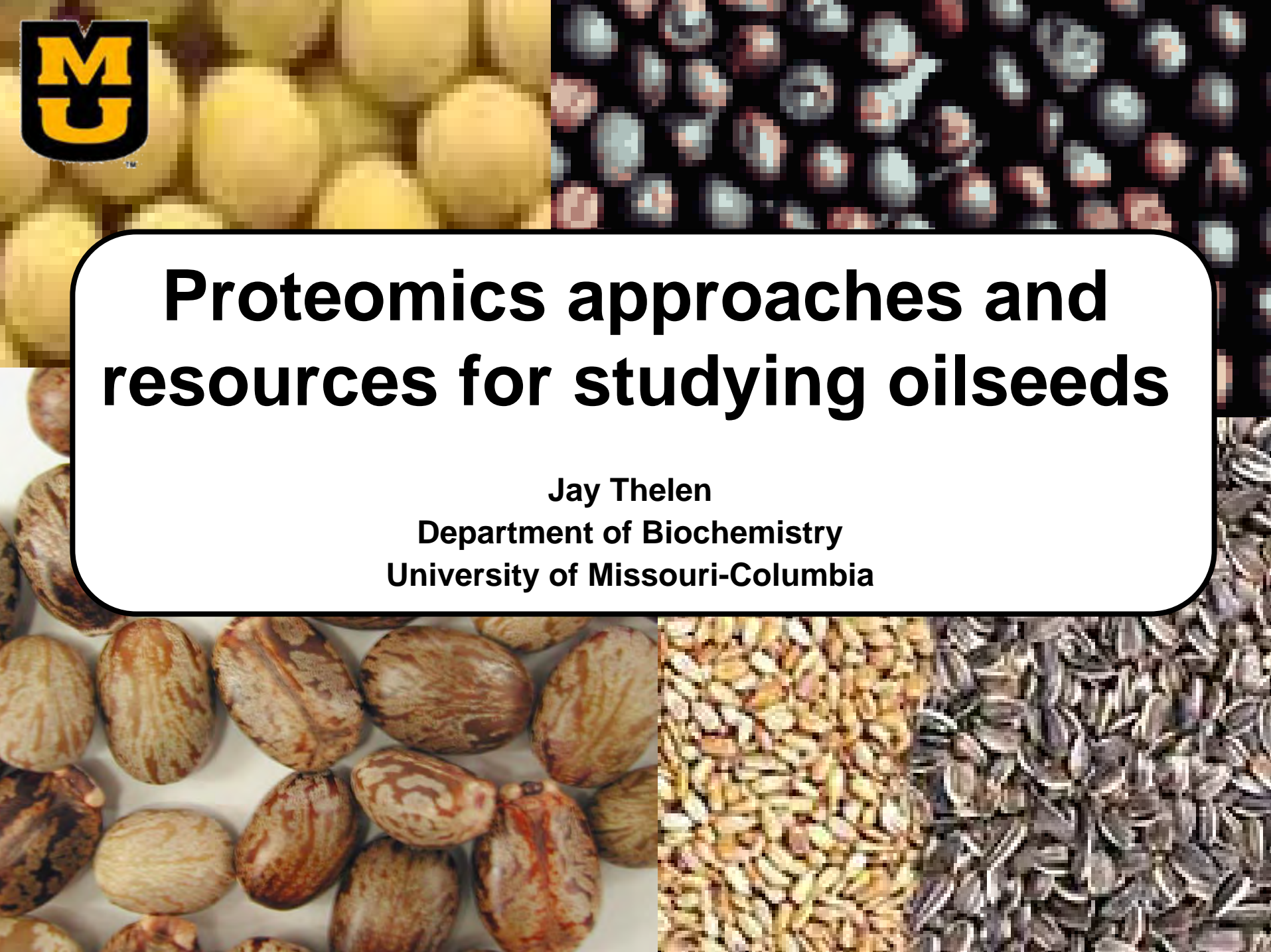




# **Proteomics approaches and resources for studying oilseeds**

**Jay Thelen**

**Department of Biochemistry  
University of Missouri-Columbia**



# Outline

---

- **Introduction**
- **Proteomics of seed filling in oilseeds**
  - 2-D electrophoresis quantitative proteomics
  - “Oilseed Proteomics” website
- **New quantitative proteomics approaches**
  - analysis of peanut Ara h 2 RNAi line, a collaboration with Peggy Ozias-Akins

# Acknowledgements

## Thelen Lab

**Martin Hajduch**

**Ganesh Agrawal**

**Severin Stevenson**

**Mingjie Chen**

**Diogo Demartini**

**Renuka Jain**

**Kyndal Marshall**

**Katherine Graham**

## Collaborators

**Peggy Ozias-Akins**  
**(UGA-Tifton)**

**Dong Xu (UMC)**



**Oilseed Proteomics website**

**<http://oilseedproteomics.missouri.edu>**

**Funding: NSF-Plant Genome Program**

# Oilseeds and vegetable oils

---

**Agriculture:** \$45 billion annually in world trade exceed \$70 billion by 2010

**Health:** Protein and oil rich. Up to 25% of human caloric intake is derived from vegetable oils

**Environment:** Oils are cheap (\$0.6 per kg), renewable alternative to petroleum fuel (biodiesel)

**Industrial:** Reservoir of fatty acid diversity – plants synthesize over 200 different fatty acids. Cheap feedstocks for chemical industry

# **Protein allergens in oilseeds**

---

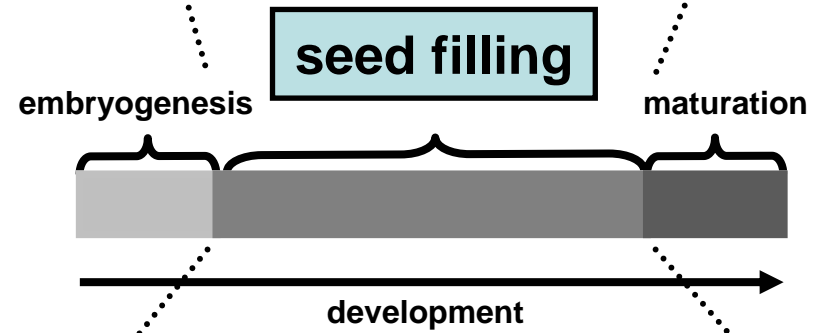
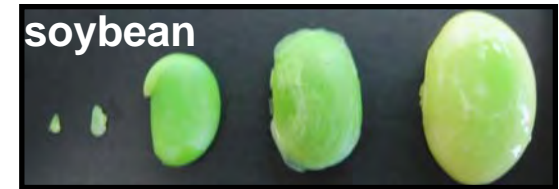
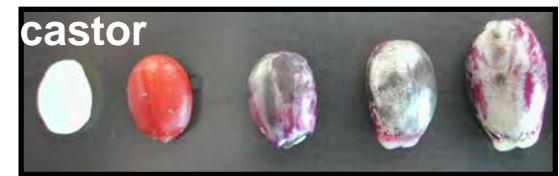
- **Human health and agricultural concern**
- **Present in many crop plants, particularly problematic for soybean and peanut (~1% of population is affected)**
- **Diverse function- storage proteins, structural, metabolic, defense**
- **Basic information about these proteins is not completely known...expression, diversity of form (protein species)**

# Objectives/deliverables for NSF-PGR: “Proteomics of seed filling in oilseeds”

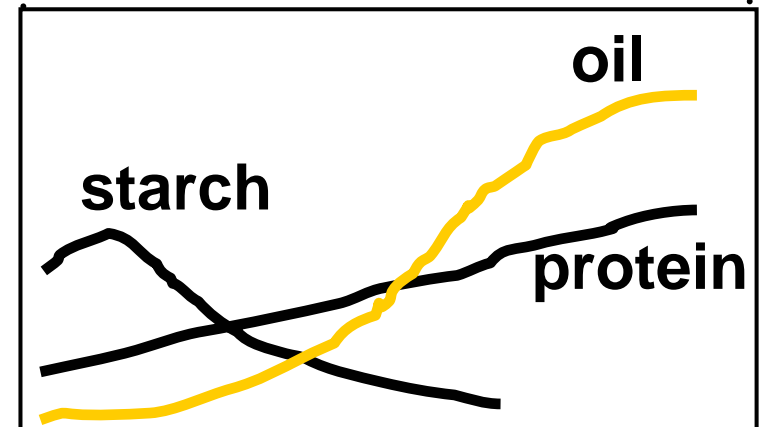
---

- Quantitatively profile proteins expressed during seed filling in soybean, canola, castor, and Arabidopsis using high res. two-dimensional gel electrophoresis (2-DE)
- Establish 2-DE gel reference maps of profiled spots and identify >500 proteins by mass spectrometry
- Develop website for data dissemination
- **Model and compare carbon assimilation in four diverse oilseeds**

Seed/embryo development  
is divided into three phases



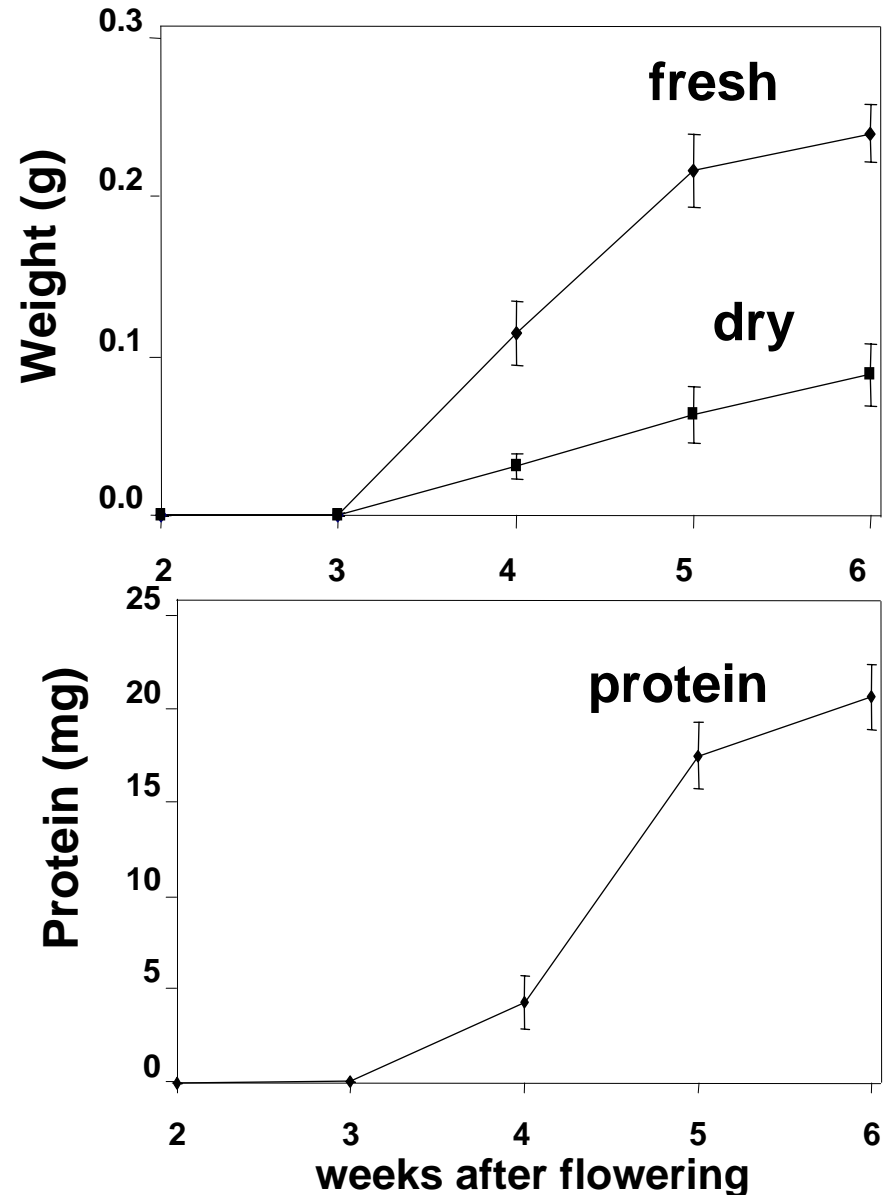
Seed filling is the phase  
when storage reserves  
are made



# Soybean seed filling



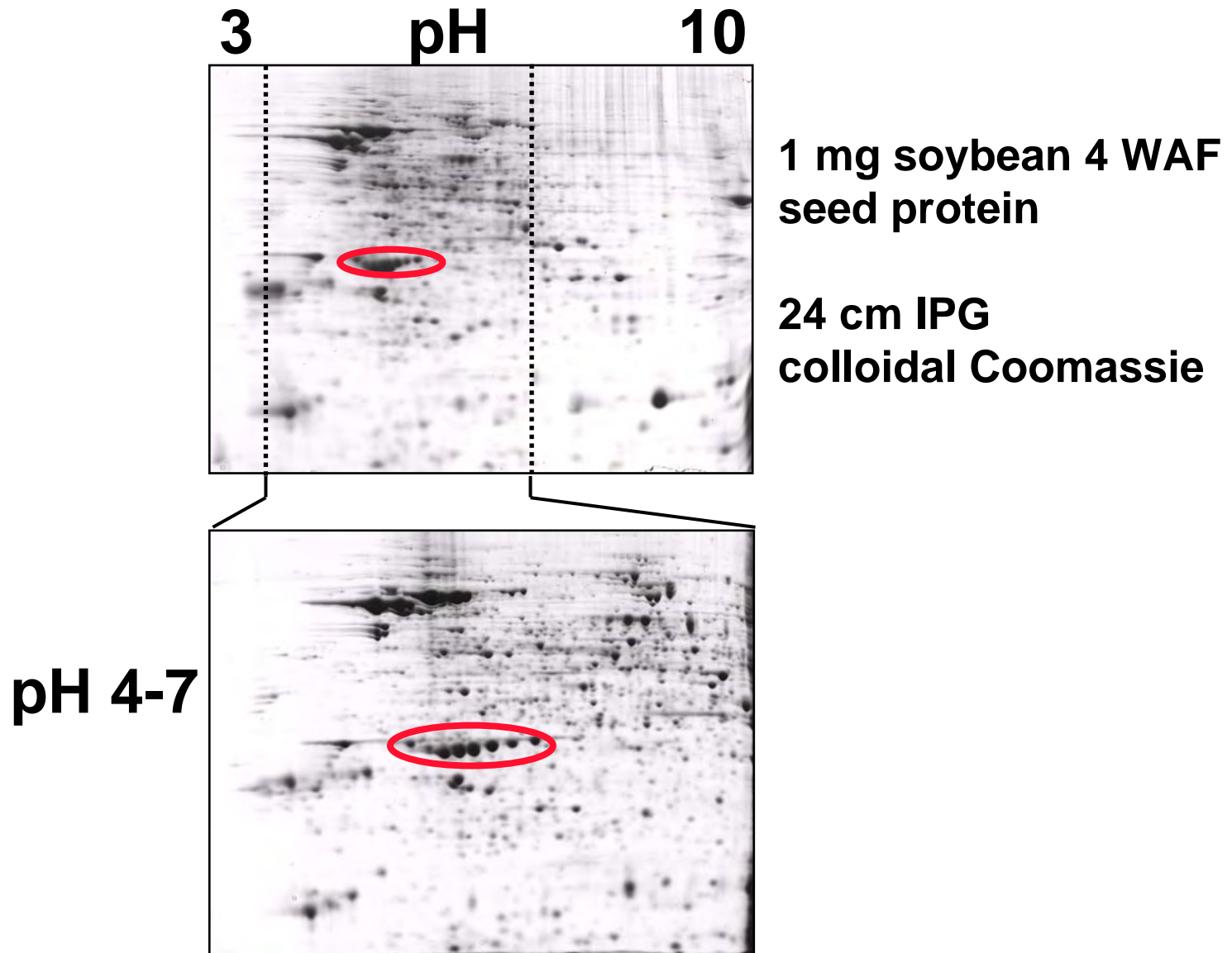
**soybean (Maverick)**  
**2, 3, 4, 5, 6 weeks after flowering**



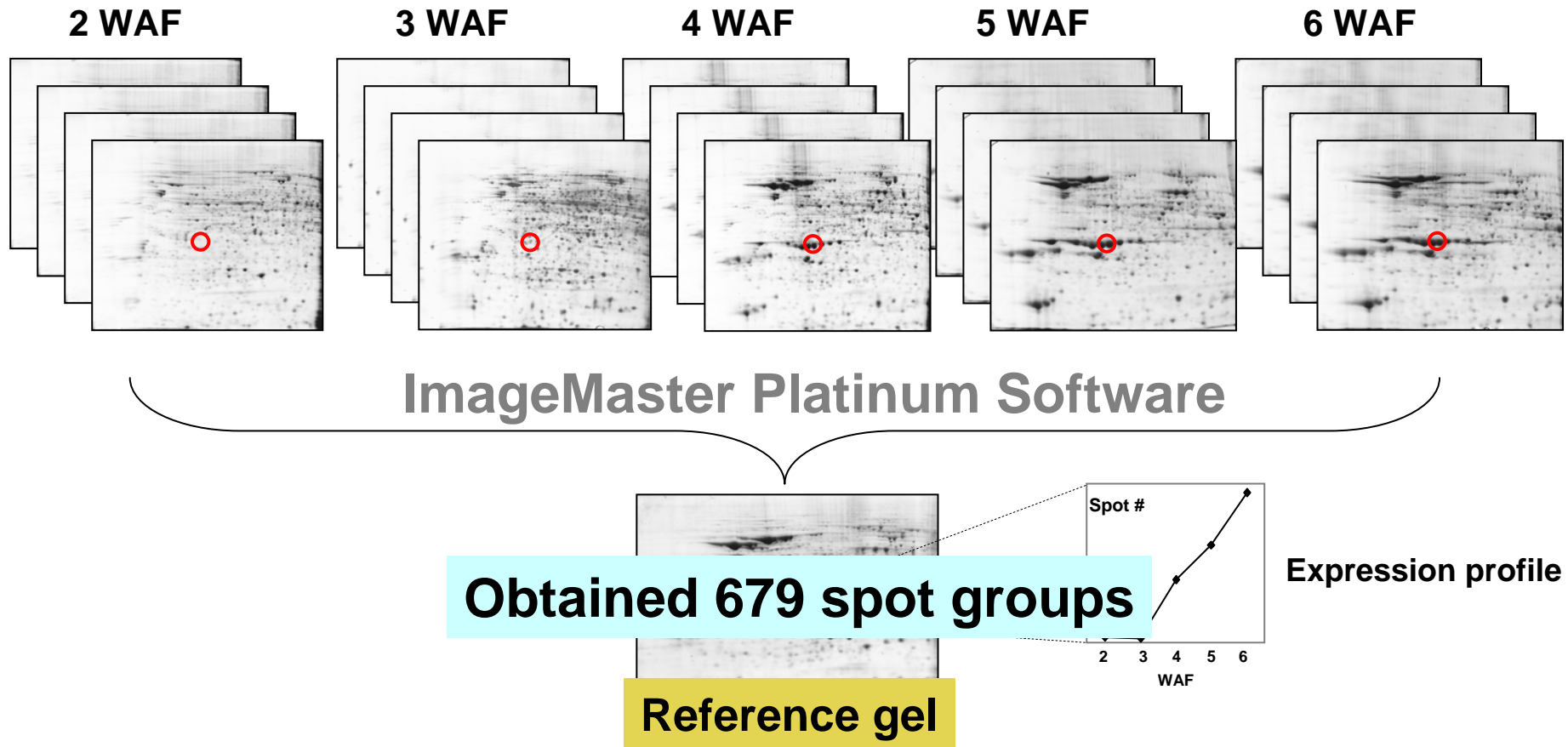


# Medium range IEF improves spot resolution and coverage

---



# Protein expression analysis



**Two sets of gels (3-10 and 4-7) = 40 gels**

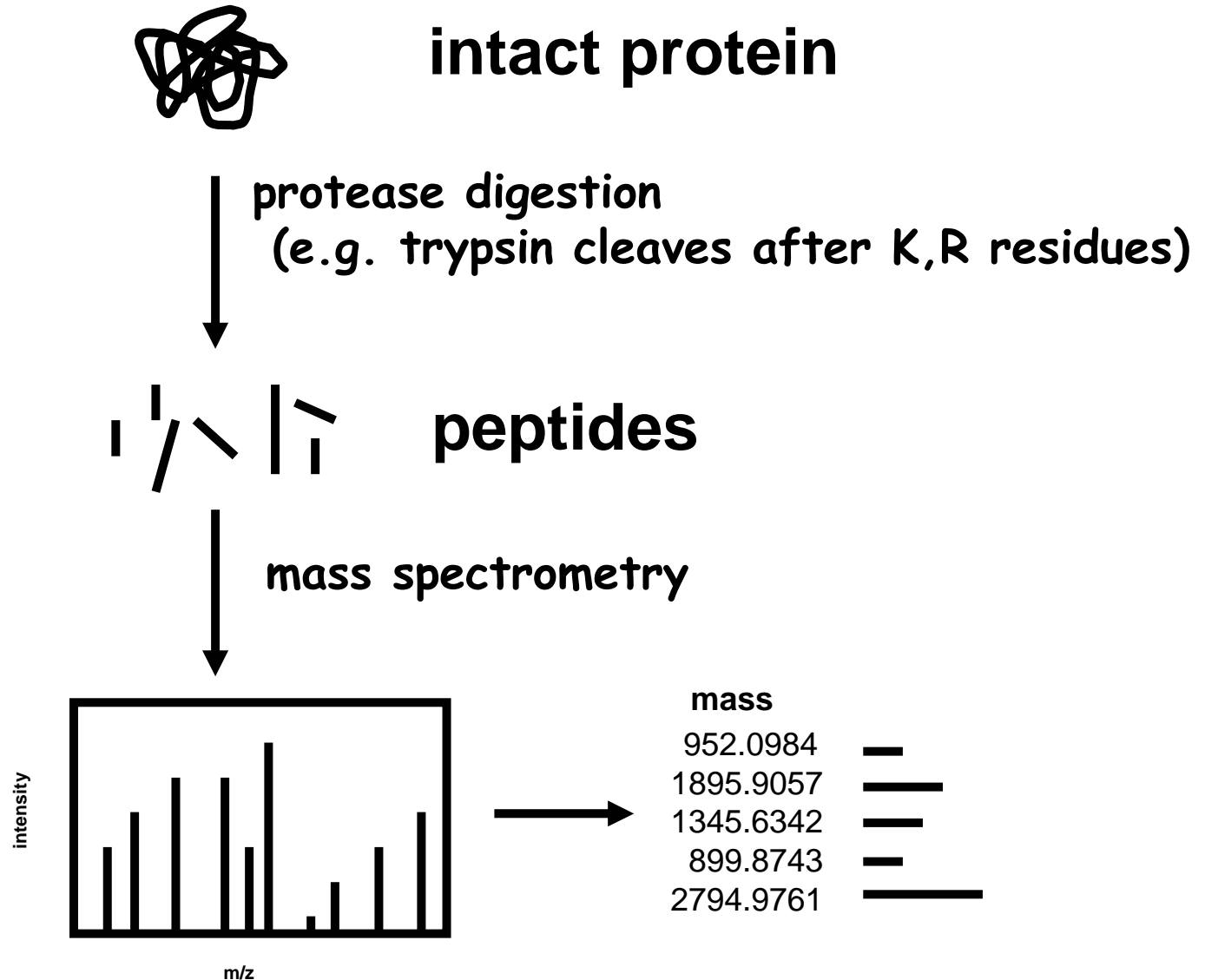
Hajduch et al., (2005) *Plant Physiol.*

# Protein identification by mass spectrometry

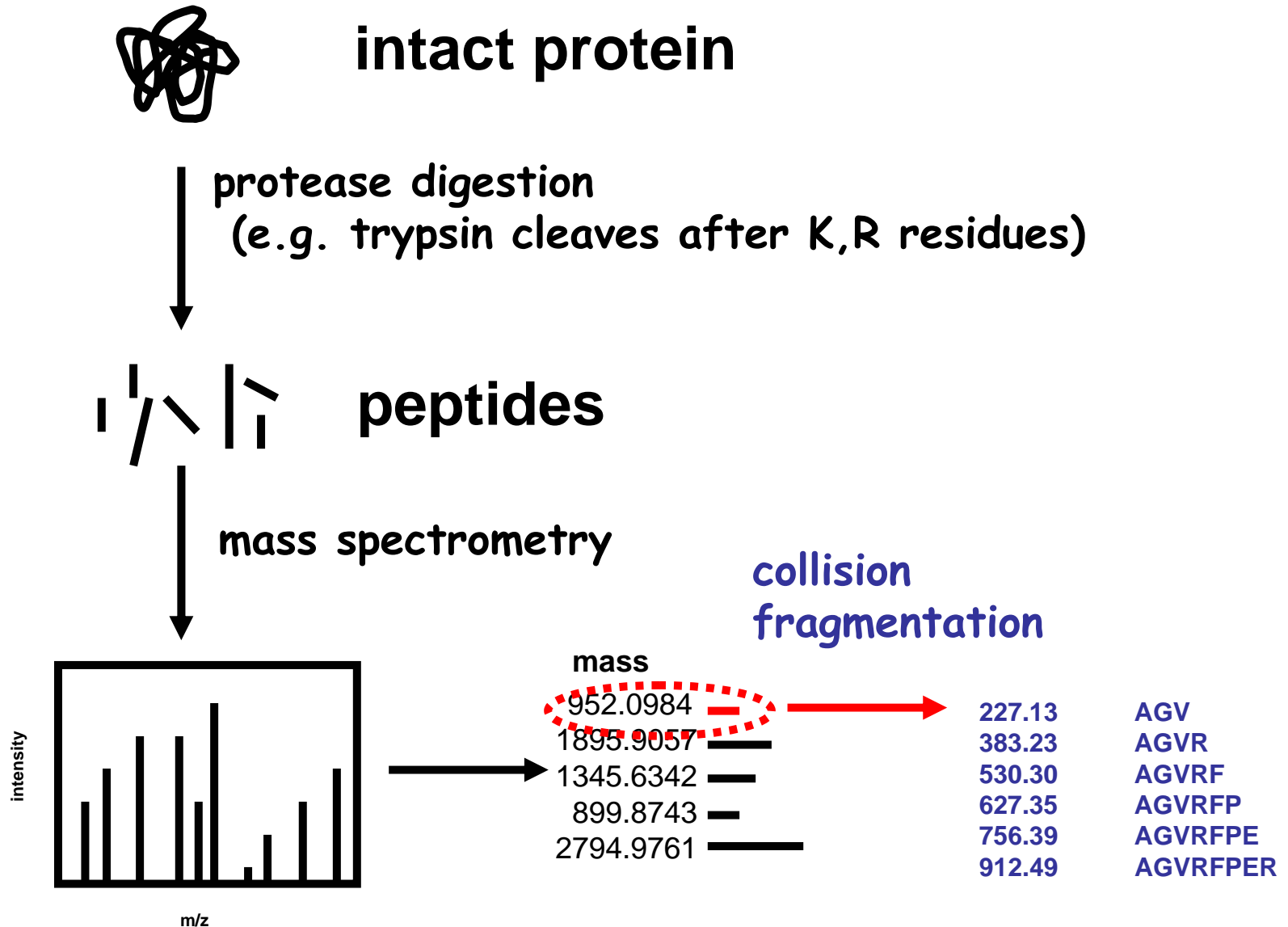
---

- 1) Peptide mass fingerprinting (PMF)
- 2) Tandem mass spectrometry (MS/MS)
- 3) *De novo* sequencing of peptides from tandem mass spectrometry

# Peptide mass fingerprinting



# Tandem mass spectrometry (MS/MS)



# Parallel MALDI-TOF and LC-MS/MS analysis of soybean 2-DE spots

## LC-MS/MS

**All distinct/resolved spots**

960 protein spots analyzed

531 protein spots identified

(unpublished)

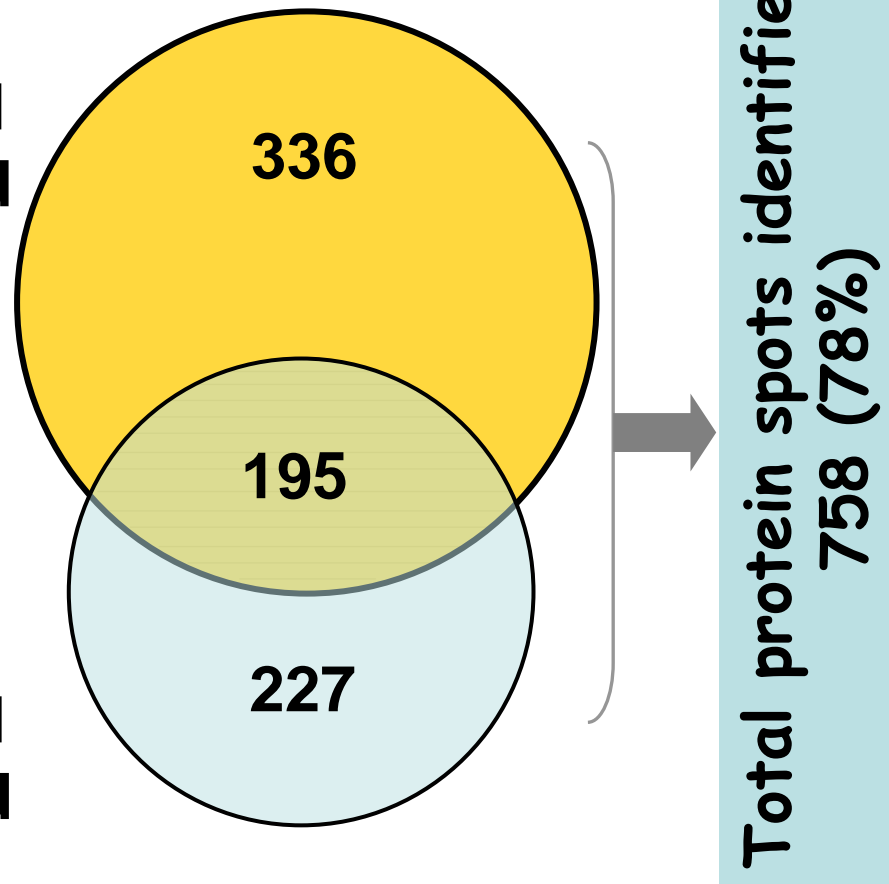
## MALDI-TOF-MS

**“spot groups” only**

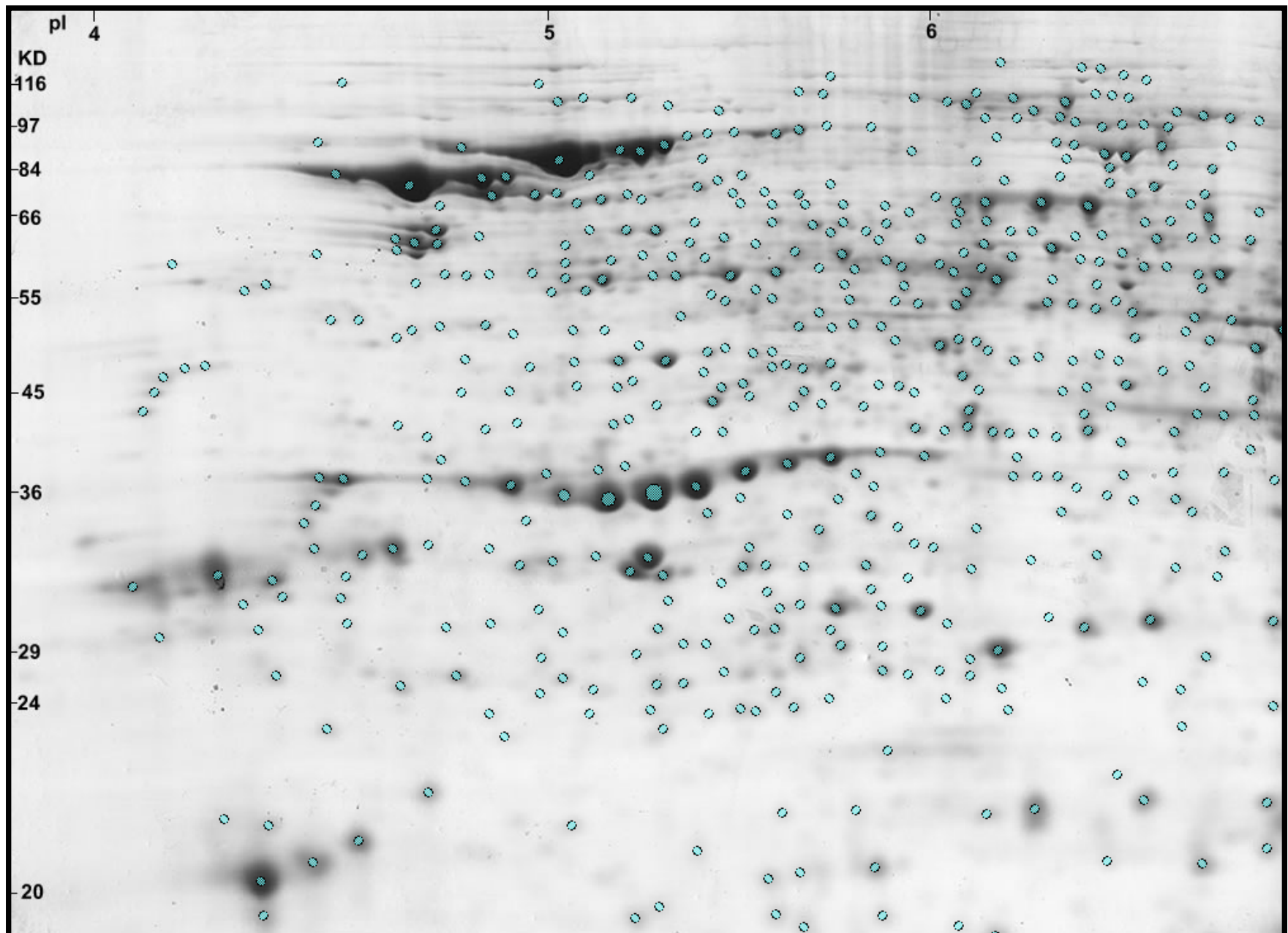
679 protein spots analyzed

422 protein spots identified

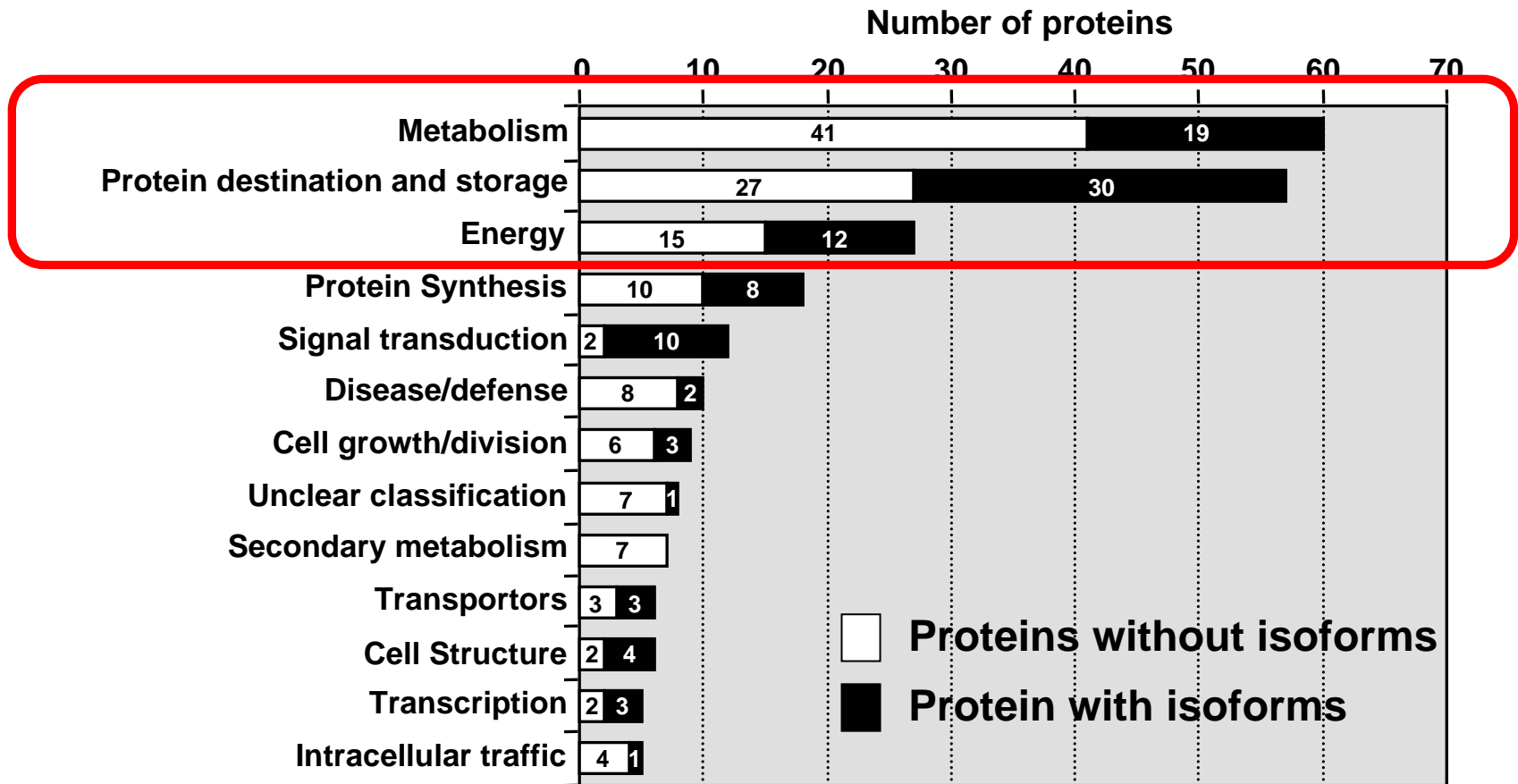
(Hajduch et al. 2005)



# Majority of 2-DE spots were analyzed



# Functional classification of soybean seed proteins - nonredundant





# Oilseed allergens

---

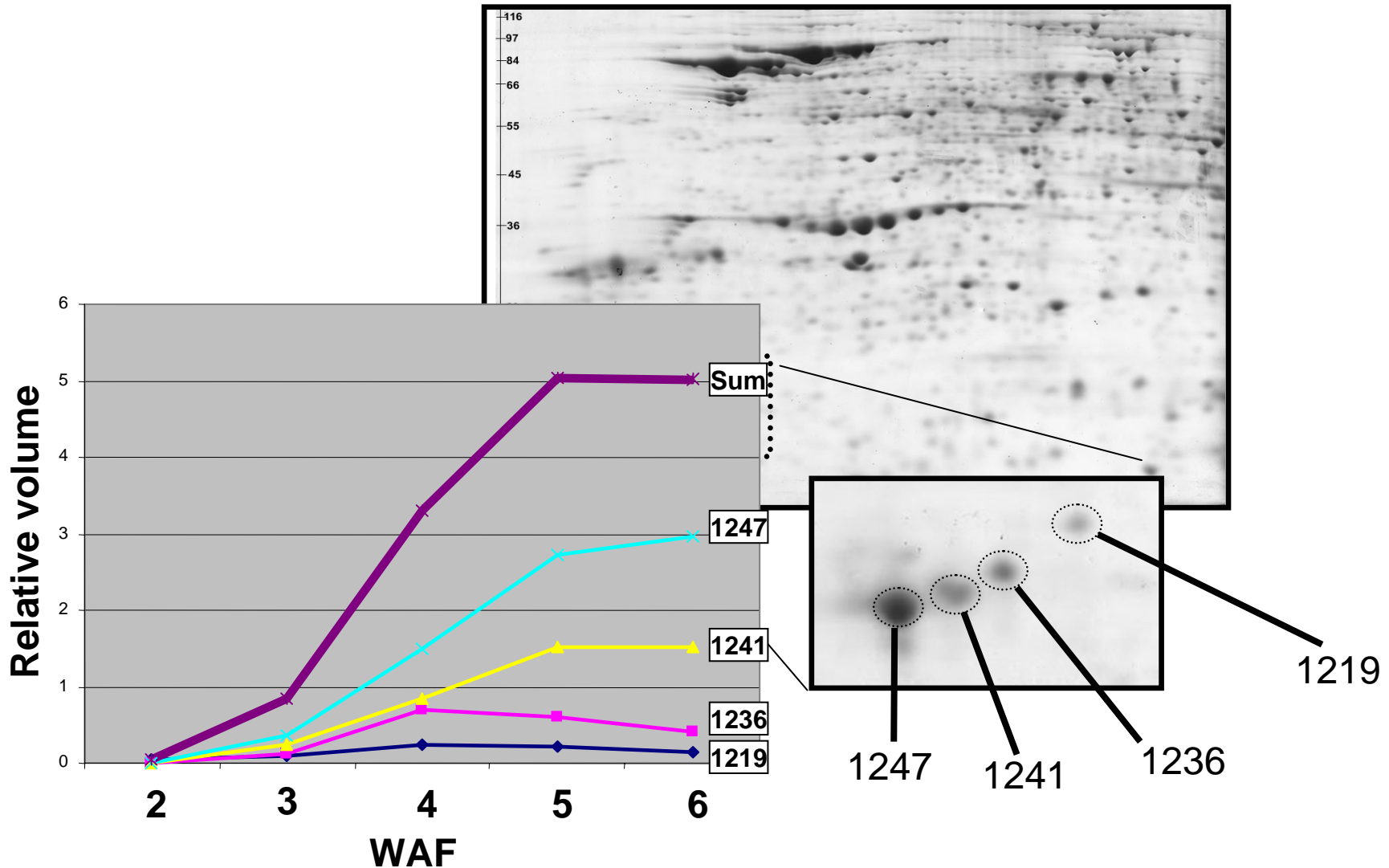
**Soybean - glycinin,  $\alpha$ -subunit  $\beta$ -conglycinin, kunitz trypsin inhibitor, thiol protease (Gly m Bd 30k), MP27 globulin (Gly m Bd 28k), profilin**

**Canola - napin (2S albumin), lipid transfer protein, cruciferin**

**Peanut - conglutin, PR-10, profilin, agglutinin, oleosin, lipid transfer protein**

# Soybean - Kunitz trypsin inhibitor

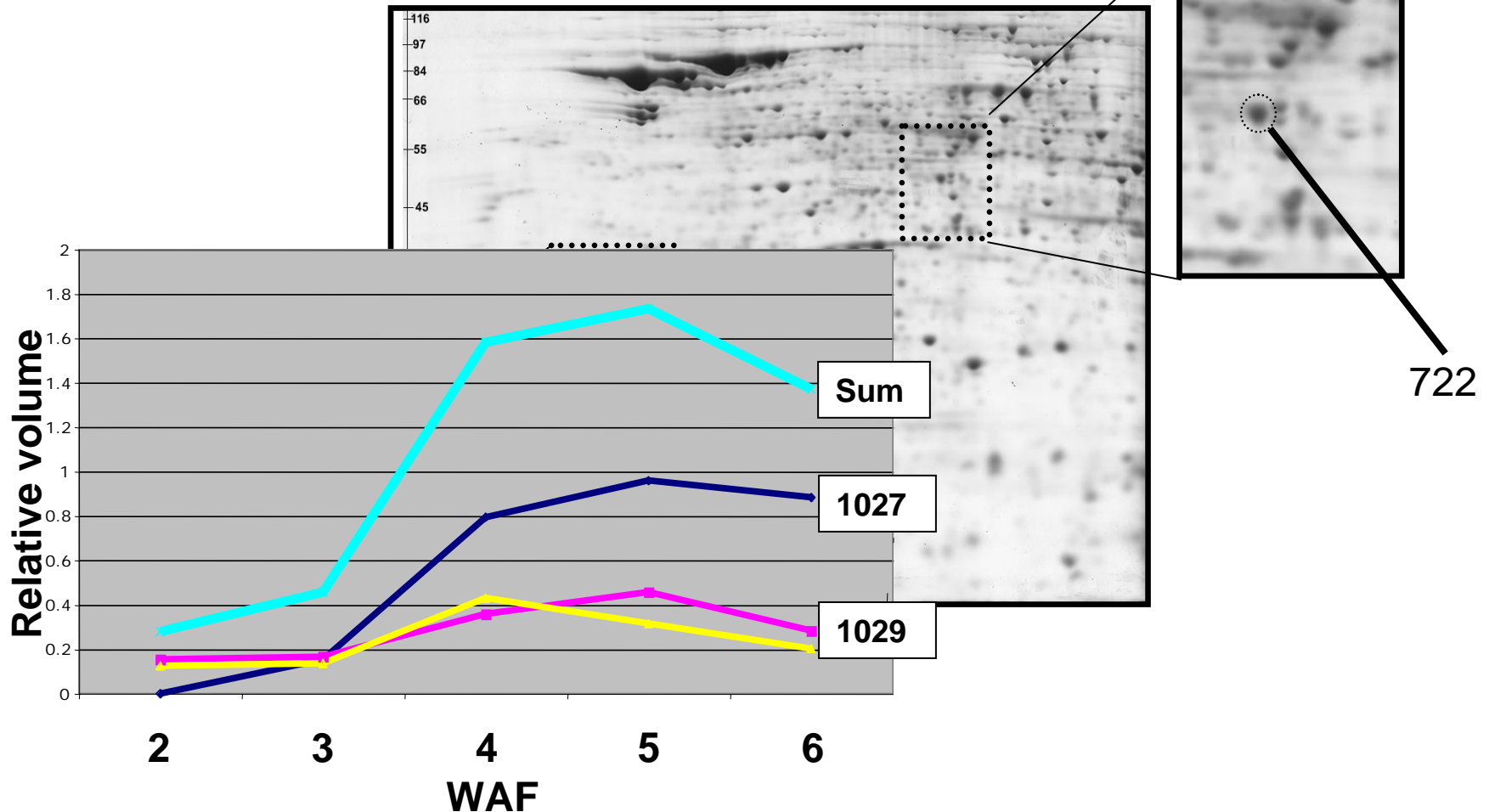
➤ Observed 4 isoforms, each by MS & MS<sup>2</sup>



# Soybean - Gly m Bd 30K

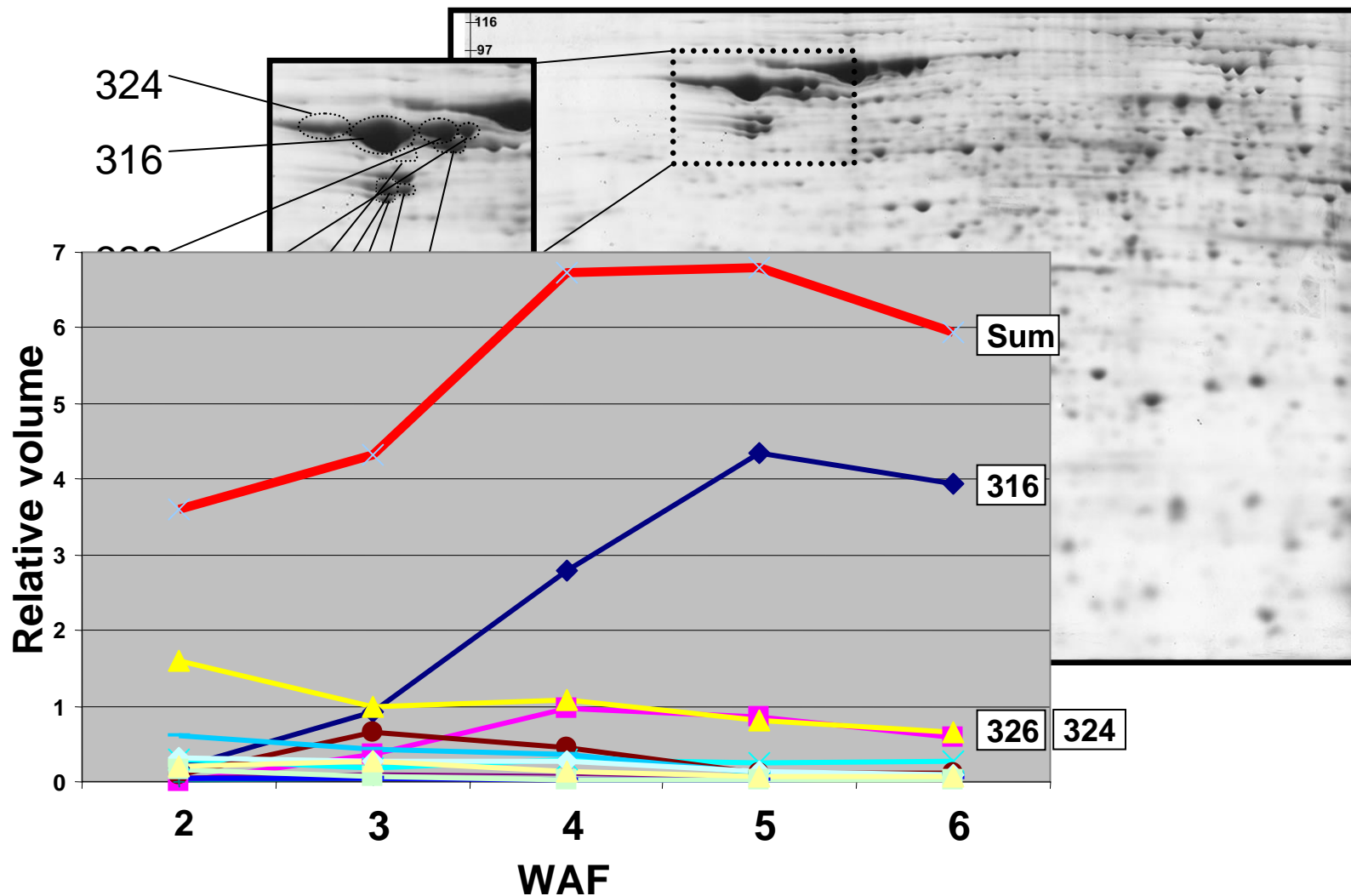
## 34 kDa maturing seed vacuolar thiol protease

➤ Observed 3 isoforms by MS<sup>2</sup>



# Soybean - $\alpha$ subunit $\beta$ -conglycinin

➤ Observed 13 isoforms, 8-both MS & MS<sup>2</sup>, 5-MS<sup>2</sup>



# Other allergens

---

## ➤ Soybean

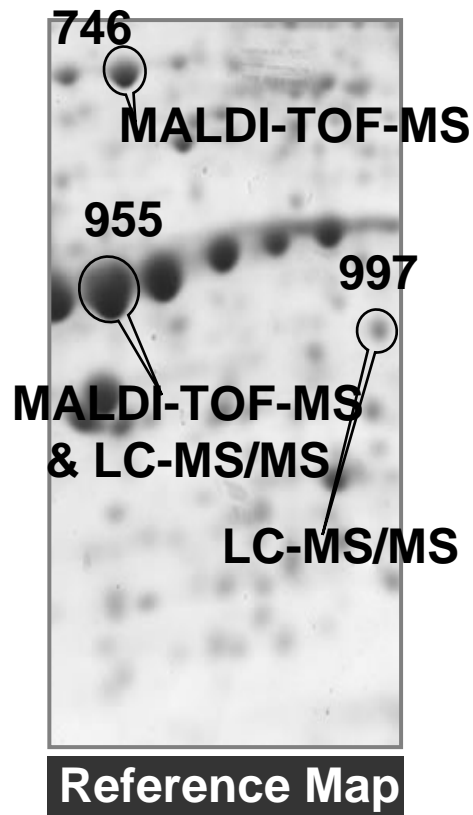
**29 spots - glycinin**

## ➤ Canola

**70 spots - cruciferin (Hajduch et al., 2006)**

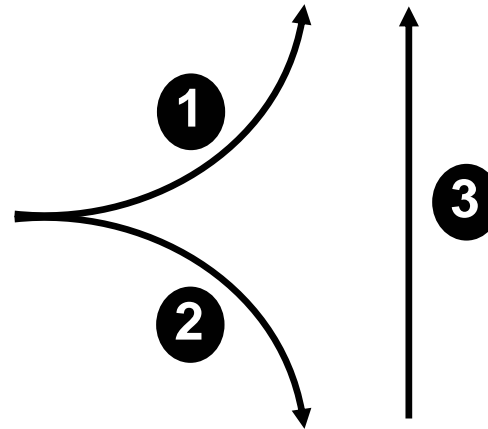
**- some of this diversity is due to post- translational modification as 20 phosphoprotein spots were mapped to cruciferin (Agrawal & Thelen, 2006)**

# Construction of “Oilseed Proteomics” website



**View Options**  
(Clickable Protein Spots)

- Quantitative expression profiles
- MALDI-TOF-MS data
- LC-MS/MS data



## Table of identified proteins

- Quantitative expression profiles
- MALDI-TOF-MS data
- LC-MS/MS data

# Proteomics of Oilseeds

HOME

ARABIDOPSIS

OILSEED RAPE

SOYBEAN

CASTOR

## SEARCH:

Search text

Google Search

☐ Web ☒ This site

## PROJECT

Personnel

Protocols

## Welcome

The following website was created to disseminate data from NSF-Young Investigator Award Plant Genome Research program-funded research project DBI-0332418 "Proteomics of Seed-filling in Oilseeds."

## Funding start date

October 1, 2003

<http://oilseedproteomics.missouri.edu>

## Links

## RELATED LINKS

National Plant Lipid Cooperation

Plant Lipid Metabolism Lab

The Arabidopsis Lipid Gene Database

Enhancing Canola through Genomics

Plant Genome Outreach Portal



are regulated in oilseeds which produce higher quantities of oil and protein. To better understand the integrated processes occurring during embryogenesis and seed-filling of select oilseeds, total proteins isolated from whole seeds at key developmental stages will be resolved, profiled and identified. Proteomic analysis of four diverse oilseeds, Arabidopsis, oilseed rape, castor and soybean will collectively provide more insight into seed-filling than analysis of any single oilseed.

## What's New

- Quantitative phosphoproteomics data of *Brassica napus* seed-filling is available under "links".
- Proteomics dataset for oilseed rape is available. However, to view results password is needed. All data will be freely available upon manuscript's acceptance for the publication.
- The new presentation format for protein functional classification using nested tables.

## Note

The website is best viewed using Internet Explorer for PC platforms and Netscape or Safari for Macintosh platforms.

02140

visitors since November 2004.



# Gel viewing option

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

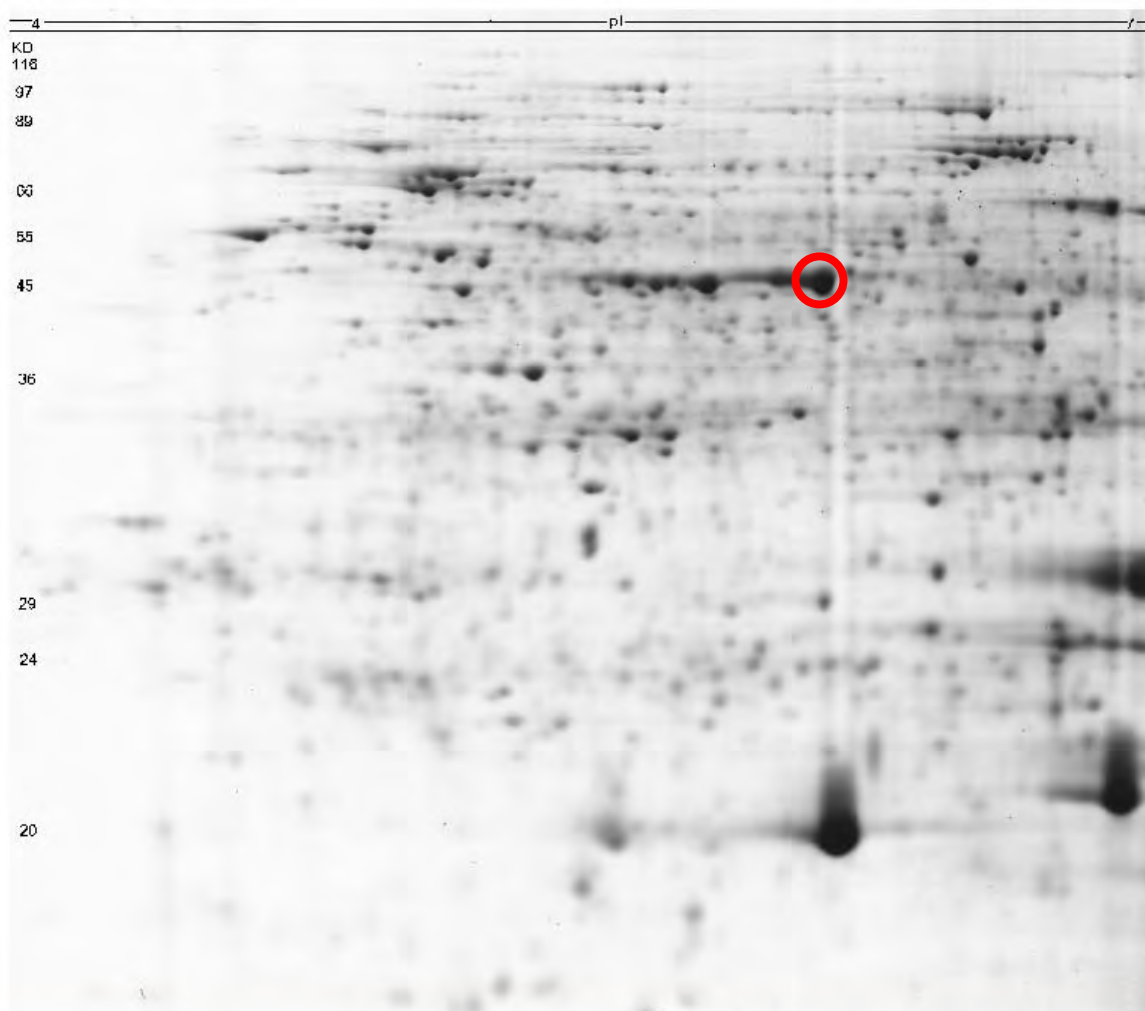
Address [http://oilseedproteomics.missouri.edu/brassicadata\\_thelen/Brassica47.html](http://oilseedproteomics.missouri.edu/brassicadata_thelen/Brassica47.html) Go Links

Google Search 117 blocked Check AutoLink AutoFill Options

## Brassica gel for pI 4-7

Spot No.: 4919

Protein ID: ribulose-1,5-bisphosphate carboxylase/oxygenase,

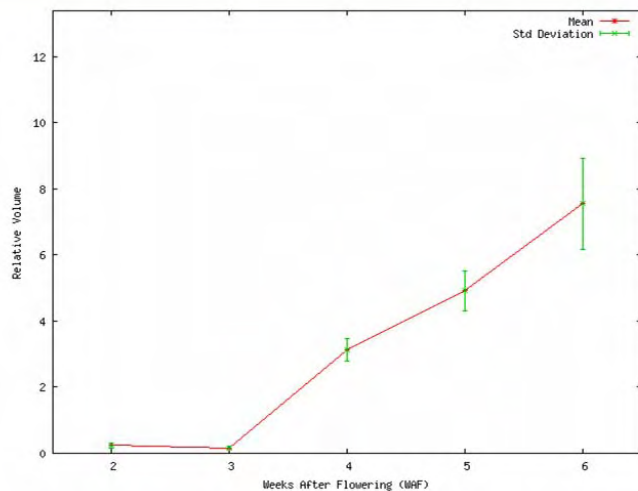




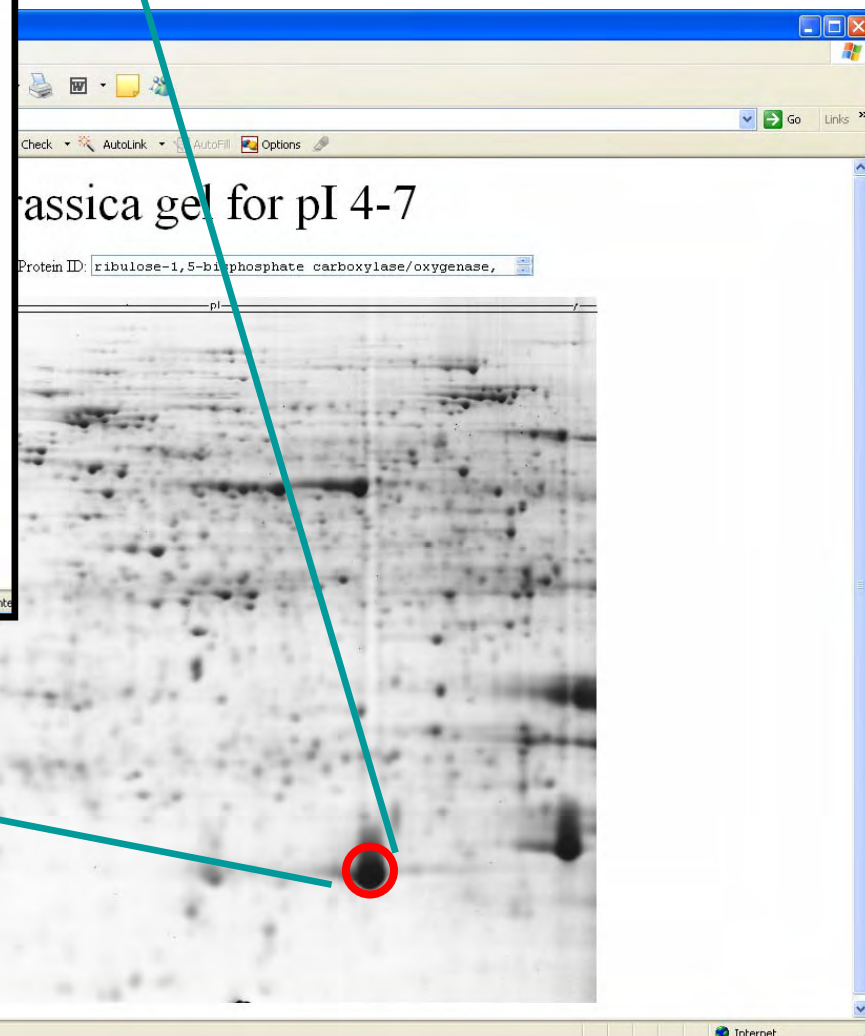
# Expression & Protein Identification page

proteomics.missouri.edu/brassicadata\_thelen/Spots/5673.html

Protein Name	Species	Database	Spot number	Accession	Theor.MW/pI	Exper.MW/pI	Cluster/Distance
cruciferin cru2/3 subunit	Bn	TIGR	5673	TC196	63.3/8.4	20.0/6.1	10/0.38
	MALDI-TOF			MS-MS			
	MOWSE score		Match/%Cov		EYS		Match/%Cov
	1.13E+05		8/61.0		320.3		4/7.9



Weeks After Flowering	2	3	4	5	6
Relative Volume (Average)	0.2329	0.1635	3.1327	4.9195	7.5498
Std. Deviation	0.1542	0.1055	0.6543	1.2152	2.7582



# Nested Table viewing option

Google Search

☐ Web ☒ This site

## PROJECT

Personnel

Protocols

Goals

Timeline

Deliverables

Publications

Links

## RELATED LINKS

National Plant Lipid Cooperation

Plant Lipid Metabolism Lab

The Arabidopsis Lipid Gene Database

Enhancing Canola through Genomics

Plant Genome Outreach Portal



## Why to investigate *B. napus* seed development?

According to 2002-03 survey ([www.barchart.com](http://www.barchart.com)) *Brassica napus* (also known as oilseed rape) is the third largest oilseed crop in the world, providing 13% of the world's supply of vegetable oil. Characterization of metabolic processes on protein level during *B. napus* seed development may uncover important correlations between proteins involved in oil production and therefore have potential high economic impact.



## Altogether, 524 identified proteins with expression profiles during seed filling are available

We applied state-of-art proteomics methodology to investigate seed filling precisely at 2, 3, 4, 5 and 6 weeks after flowering. We generated protein reference maps for proteins of pI range from 4 to 7 ([gel pI 4-7](#)) and pI 7-10 ([gel pI 3-10](#)) that contains 794 clickable protein spots with expression profiles. In total, 524 proteins were identified and classified according to their function ([Protein table](#)).

Note: Presently, password is necessary to view results. All proteomics data from this investigation will be freely available upon acceptance of the manuscript for the publication.

## Identified Proteins

- Metabolism
- Energy
- Cell growth/division
- Transcription
- Protein synthesis
- Protein destination and storage
- Transporters
- Intracellular traffic
- Cell structure
- Signal transduction
- Disease/defense
- Unclear clasification
- Unclassified
- Secondary metabolism



## Identified Proteins

- Metabolism
  - Amino Acids
  - Nitrogen and sulph
  - Nucleotides
  - Phosphate
  - Polysaccharide cat
  - Sugars and polysa
  - Lipid
  - Sterols
  - Cofactors
- Energy
- Cell growth/division
- Transcription
- Protein synthesis
- Protein destination and s
- Transporters
- Intracellular traffic
- Cell structure
- Signal transduction
- Disease/defense
- Unclear clasification
- Unclassified
- Secondary metabolism

## Identified Proteins

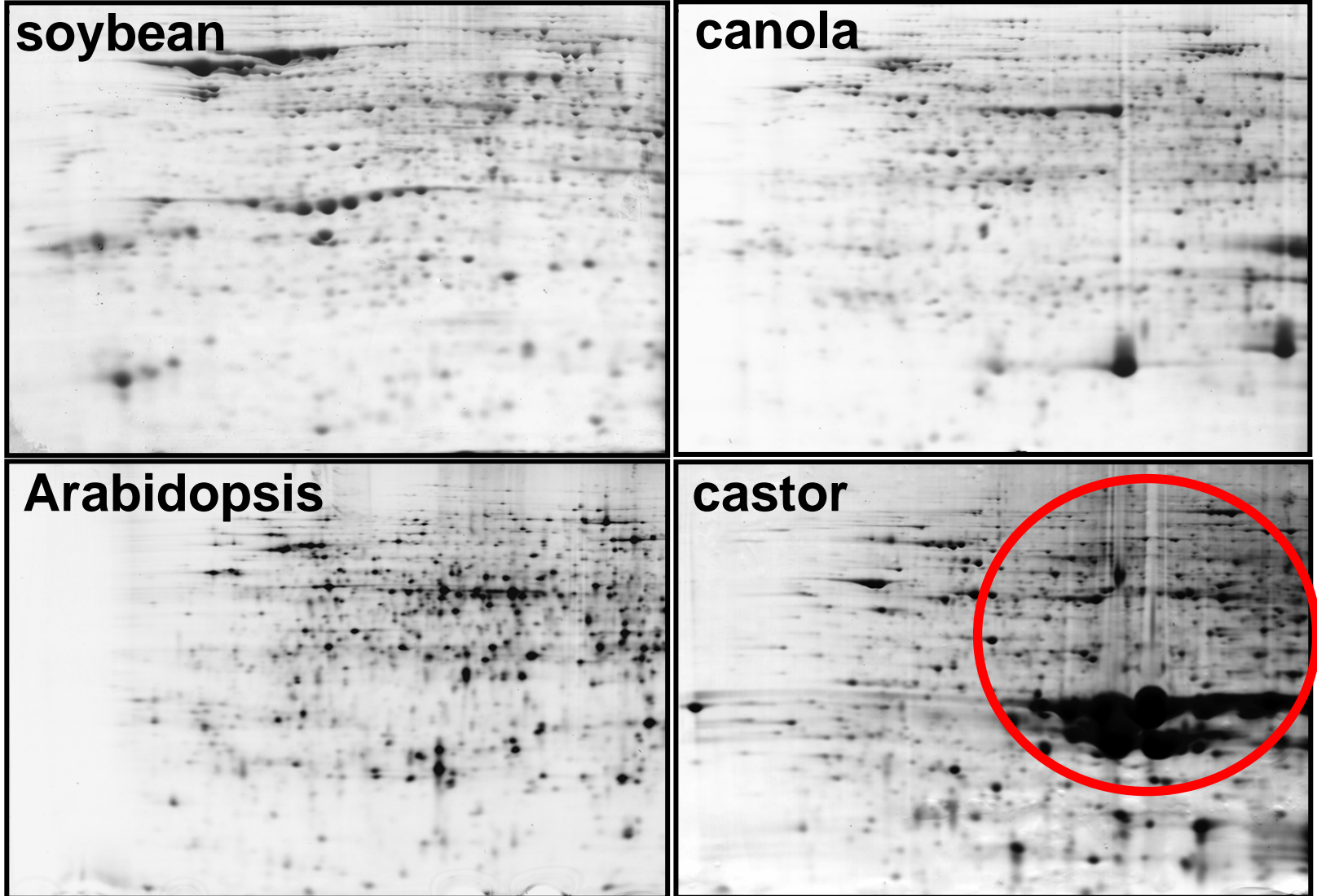
- Metabolism
  - Amino Acids
    - 3-isopropylmalate dehydrogenase
    - 3-isopropylmalate dehydrogenase
    - aconitate hydratase, cytoplasmic
      - [5809](#)
    - aminoacylase, putative / N-acyl-L-an
    - homologue to aminoacylase / N-acy
    - aspartate aminotransferase
    - aspartate aminotransferase
    - cobalamin-independent methionine
    - cobalamin-independent methionine
    - cobalamin-independent methionine
    - cobalamin-independent methionine
    - cobalamin-independent methionine
    - cobalamin-independent methionine
    - cobalamin-independent methionine
    - cysteine desulfurase
    - cysteine synthase
    - delta-1-pyrroline-5-carboxylate dehy
    - delta-1-pyrroline-5-carboxylate dehy
    - delta-1-pyrroline-5-carboxylate dehy
    - delta-1-pyrroline-5-carboxylate dehy
    - glutamate dehydrogenase (cytosolic
    - glutamate dehydrogenase (cytosolic

## Identified Proteins

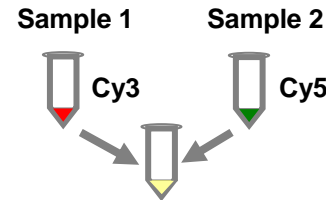
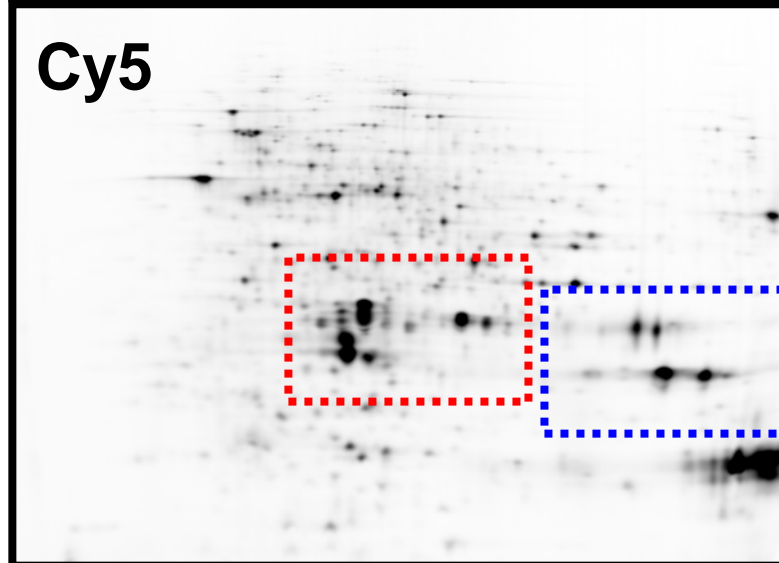
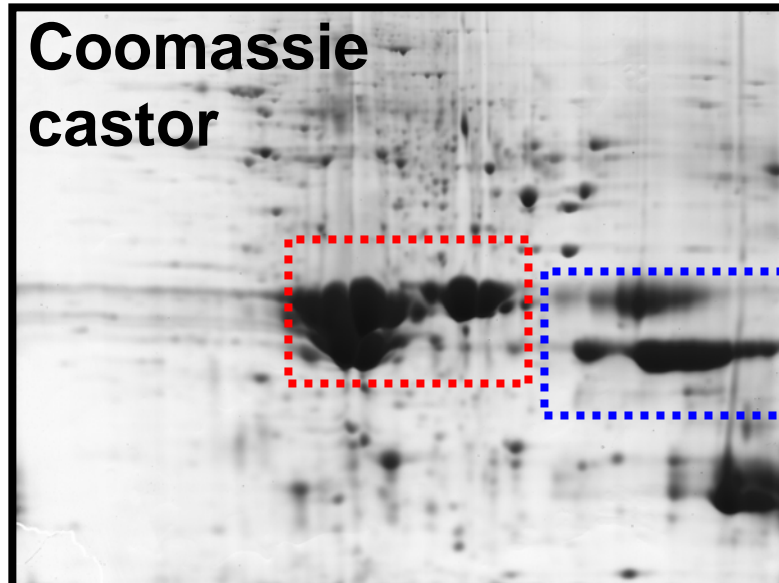
- Metabolism
- Energy
- Cell growth/d
- Transcription
- Protein synth
- Protein desti
- Transporters
- Intracellular
- Cell structure
- Signal trans
- Disease/defer
- Unclear clas
- Unclassified
- Secondary m

# Castor was a challenge due to prominent albumins in 4-7 pI range

---



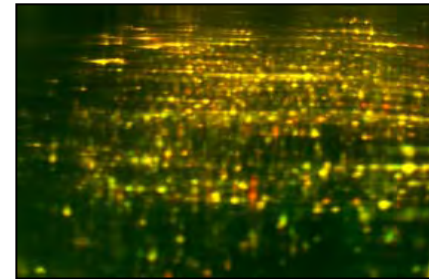
# Low Lys content for storage proteins makes DIGE superior to CBB for seed profiling



50 µg protein

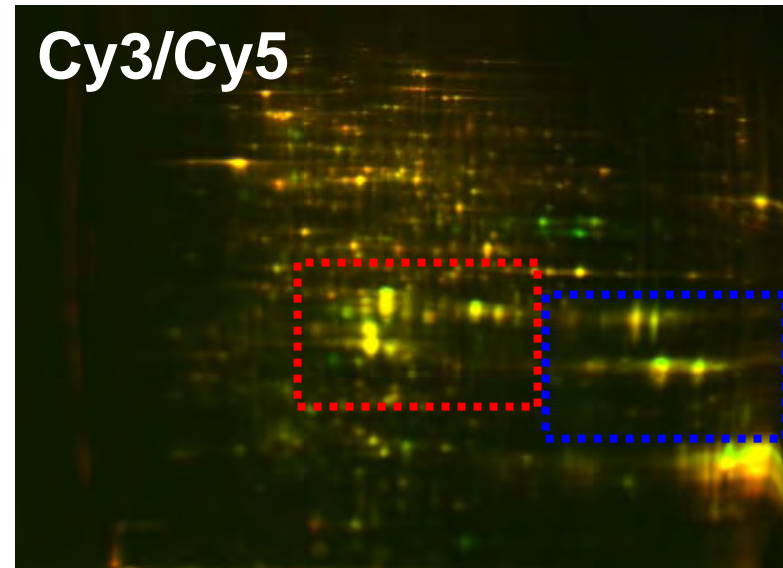
Cyanine dye label

- **Lys reactive**
- charge balanced
- 0.2 ng sensitivity



Arabidopsis  
seed

4 WAF (Cy3)/5 WAF (Cy5)





# Update/Conclusions

---

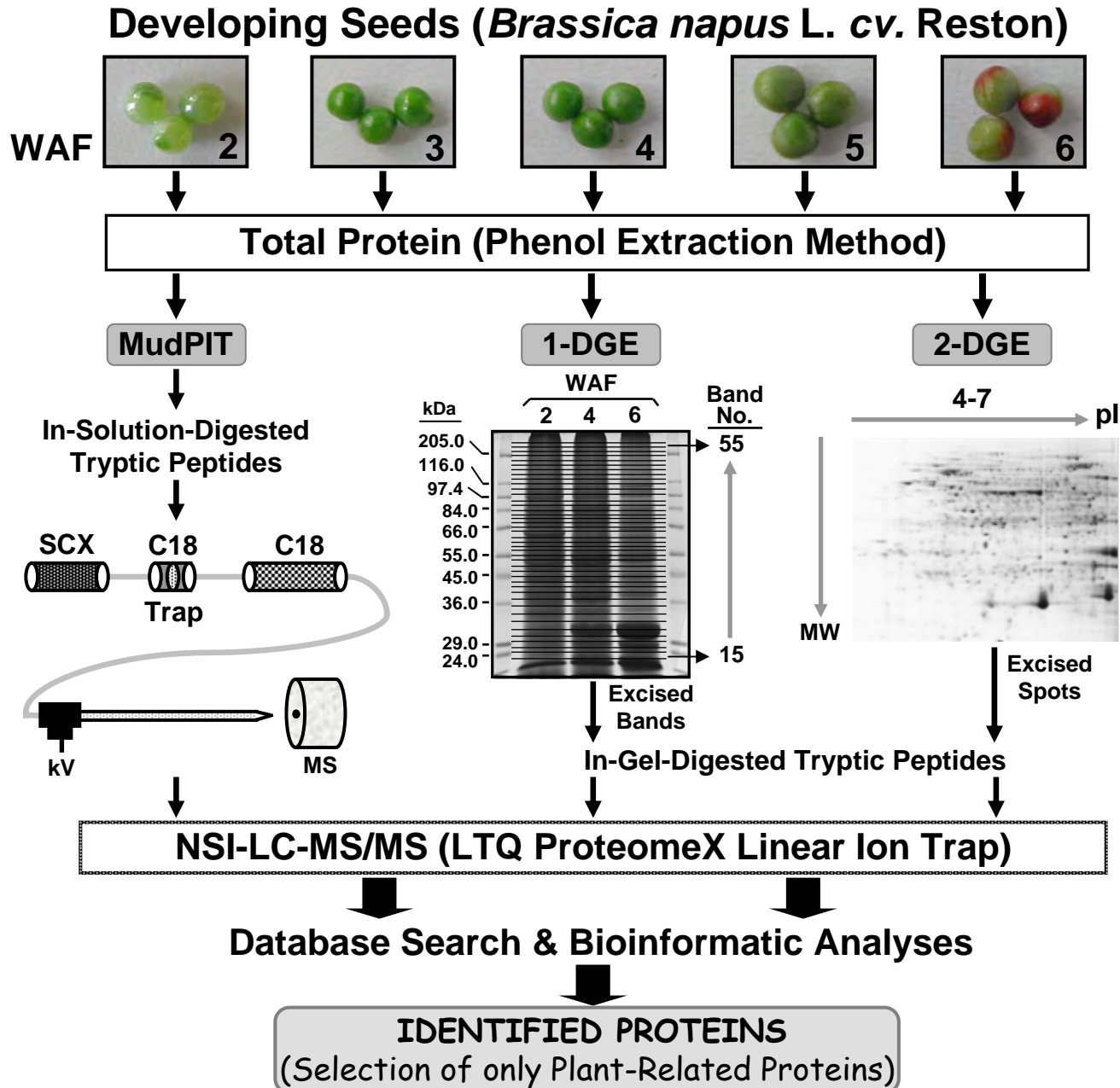
- Approx. 1400 proteins developing seed of soybean, canola, castor, and Arabidopsis were resolved and detected by 2-D Coomassie or 2-D DIGE
- 700-800 of these spots were reproducibly detected throughout seed filling (spot groups). 60-80% of these proteins were identified by LC-MS/MS
- Multiple isoforms of KTI (4), thiol protease (3),  $\alpha$  subunit  $\beta$ -conglycinin (13), glycinin (29), and cruciferin (70) detected. Phosphorylation accounts for some of this form diversity (cruciferin)
- Allergens accumulate with seed development and represent a significant portion of the protein content - KTI (5%), thiol protease (1.8%),  $\alpha$  subunit  $\beta$ -conglycinin (7%)
- Web database is open - resource for landmarking current and future allergens (e.g. gene silencing, variety studies)

# Challenges/Future work

---

- \* 1) Better quantification approach**
  - accuracy, throughput, sensitivity
  
- 2) Frequency and diversity of PTMs**
  - 11 isoforms for RuBisCO LS (single plastid gene)
  
- \* 3) Depth of proteome coverage - dynamic range**
  - SSPs represent 50-60% total proteome
  - 517 canola seed proteins
    - >98% seed protein...but <5% seed proteome

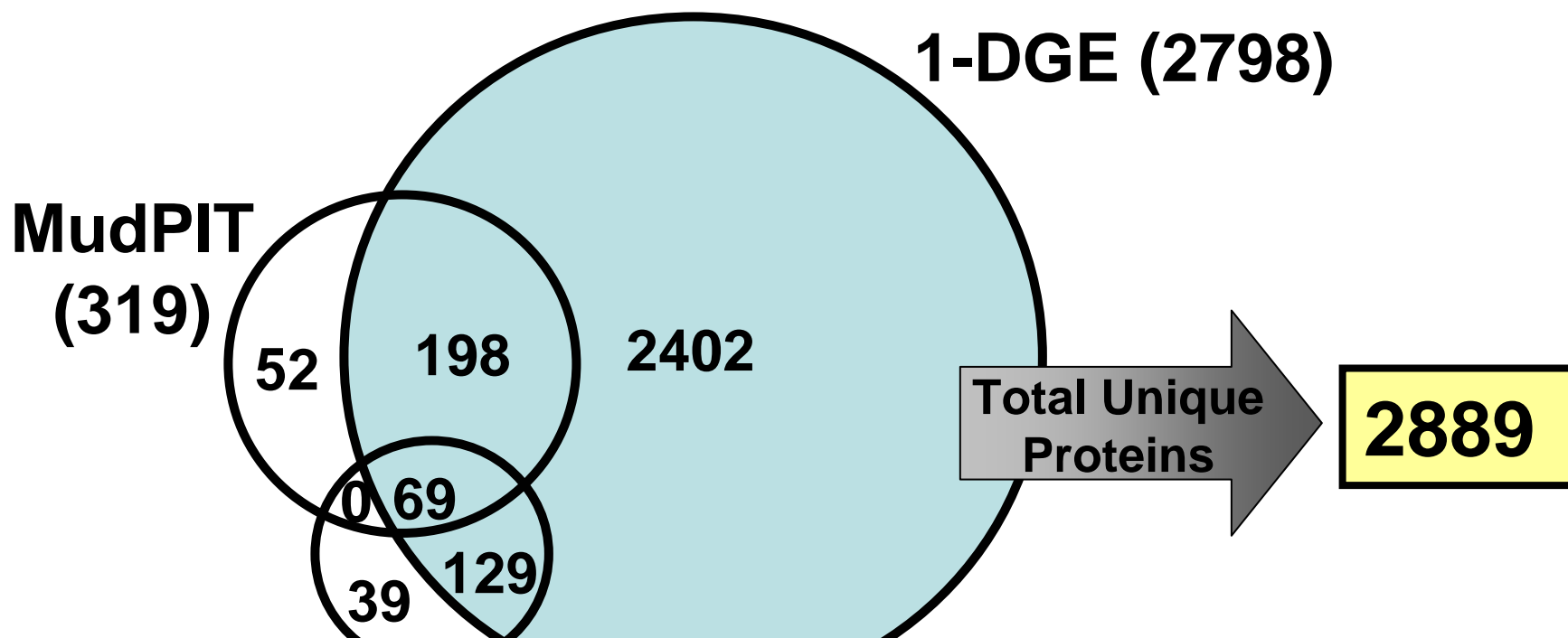
# Complementary Proteomic Approaches





# SDS-PAGE coupled to LC-MS/MS yields 10X more proteins than MudPIT and 2-DE

---



Depth of proteome issue: improved  
Can we couple “GelC-MS” to a quantitative method?

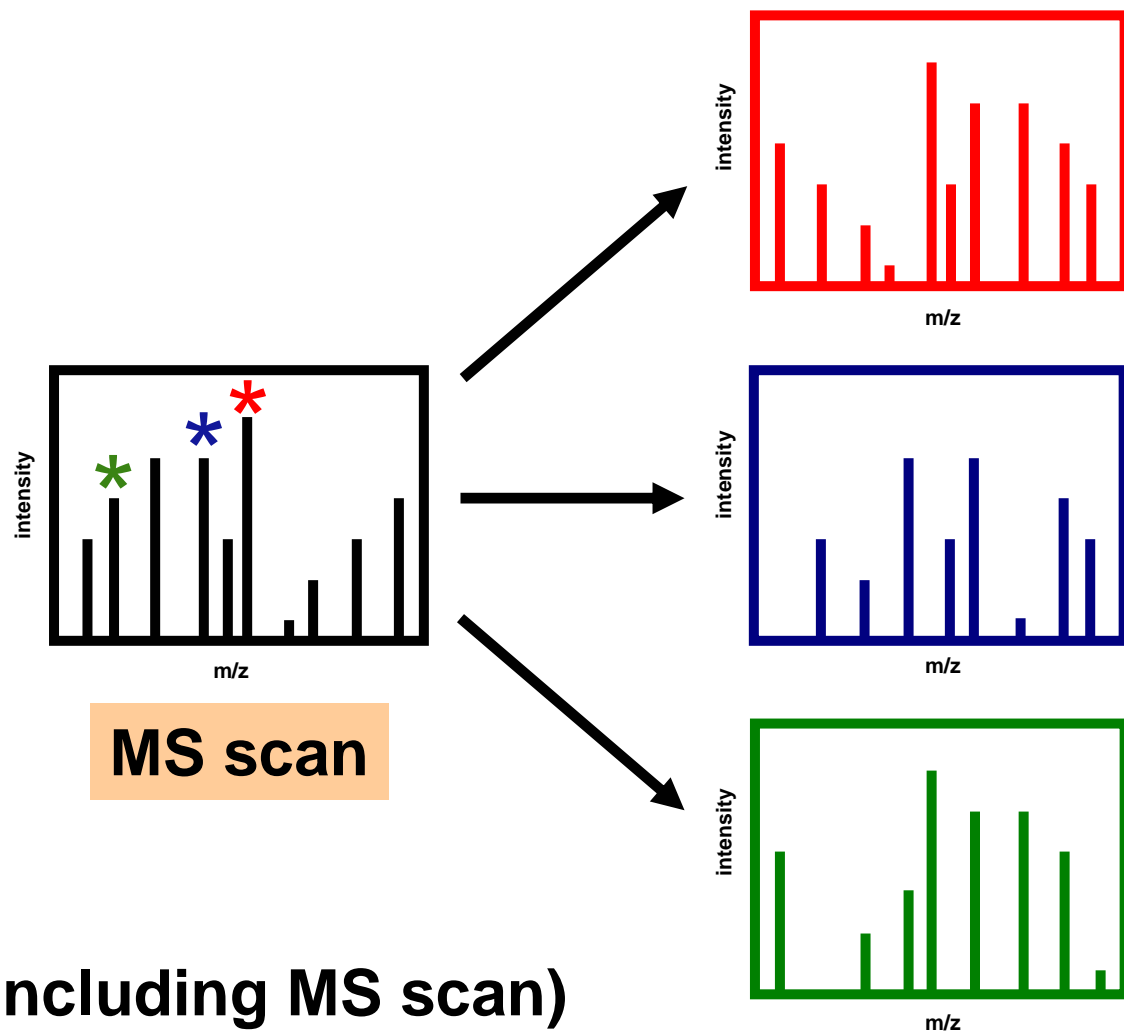
# **Label-free quantitative proteomics**

---

- spectral counting**
- peak integration**

**Protein quantification and identification are coupled unlike 2-DE ...so it is faster and cheaper to perform, if effective**

# Spectral Counting

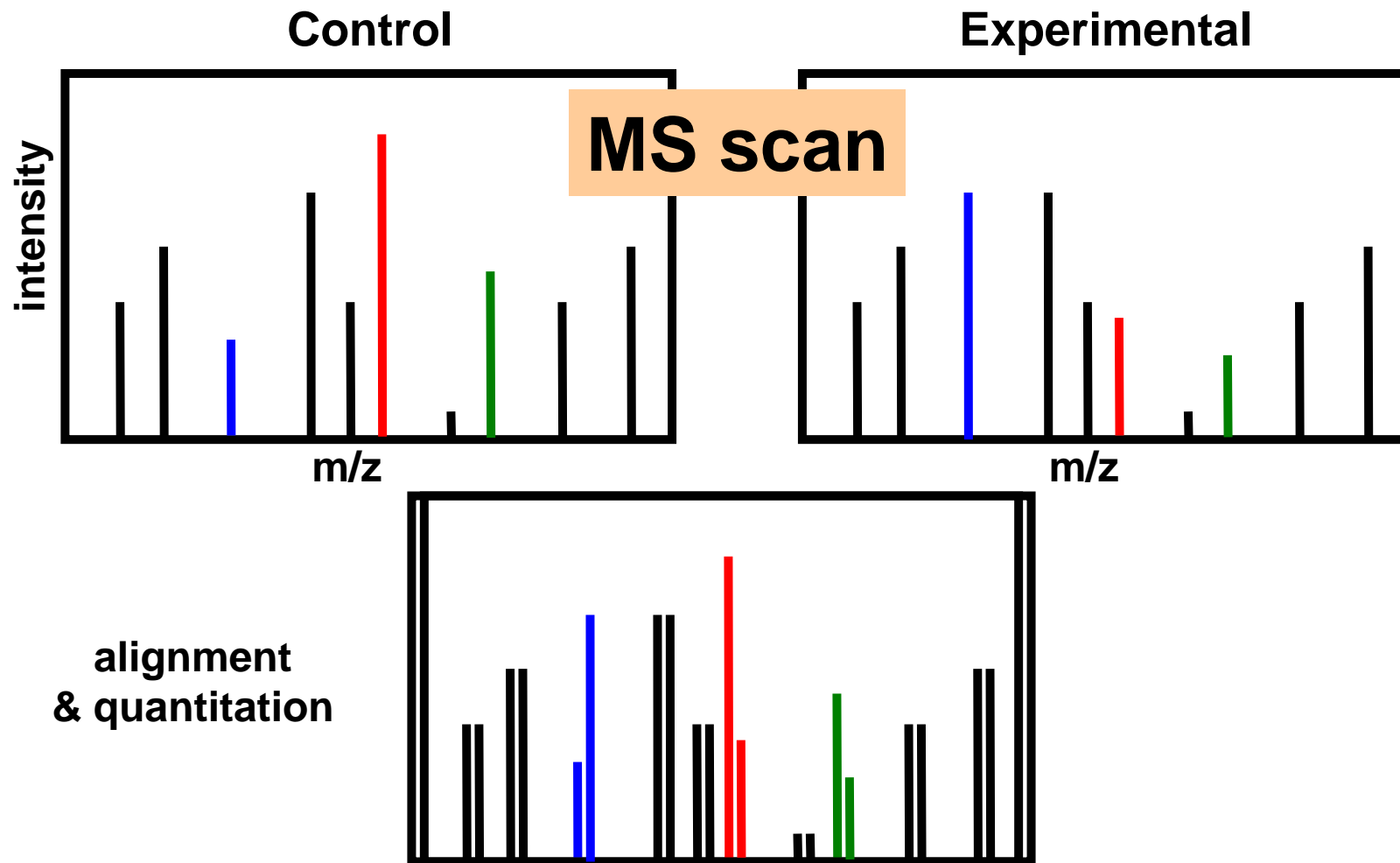


- 10 scans/sec (including MS scan)

- 2 hr C18 gradient ~ 57,600 MS/MS scans

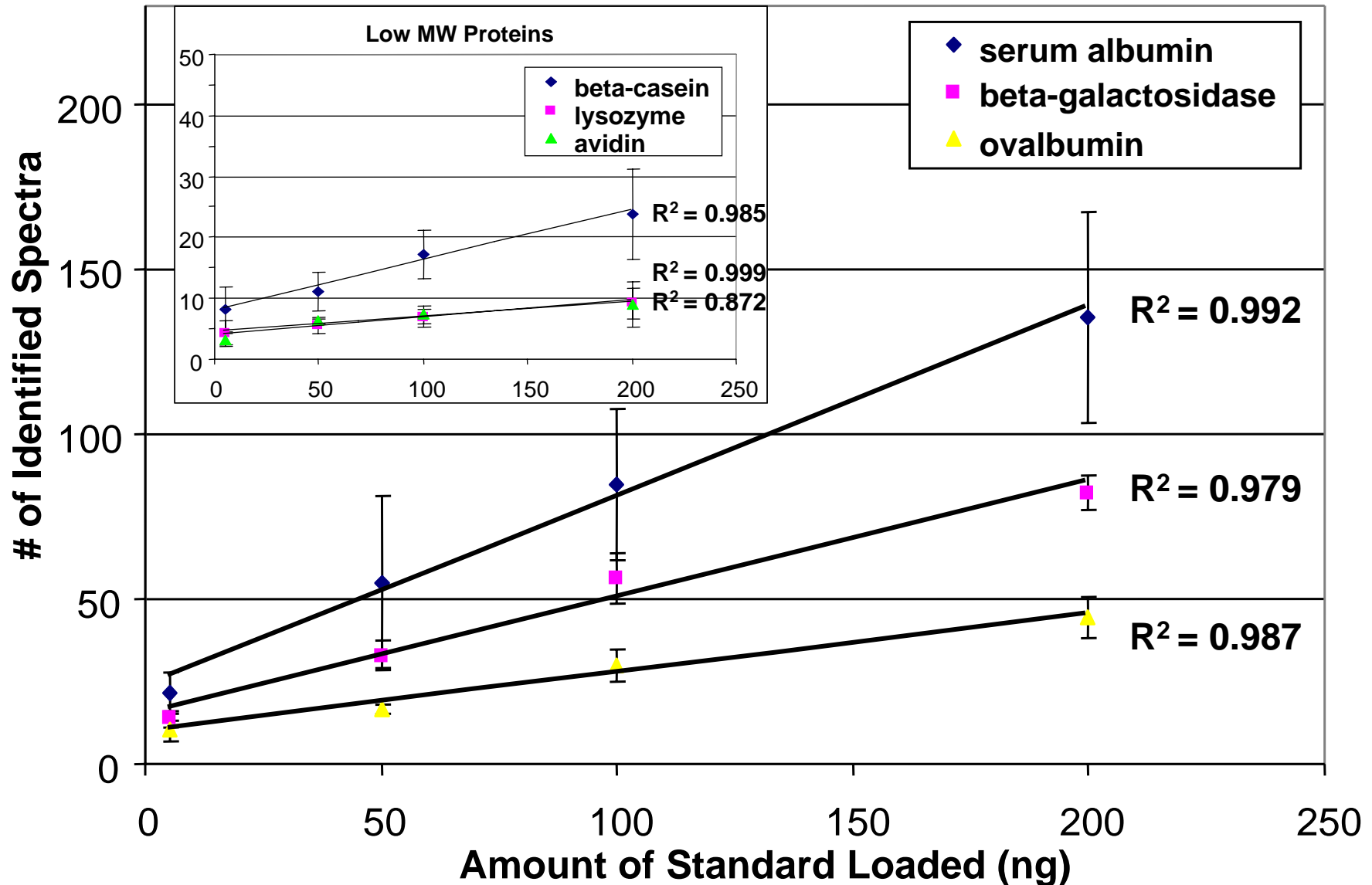
**MS/MS scans**

# Peak Integration



Ion current for each peptide is compared amongst multiple LC-MS/MS runs

# Ser. dilution of stds - spectral counting offers better dynamic range and linearity than peak integration



# **Application of label-free proteomics**

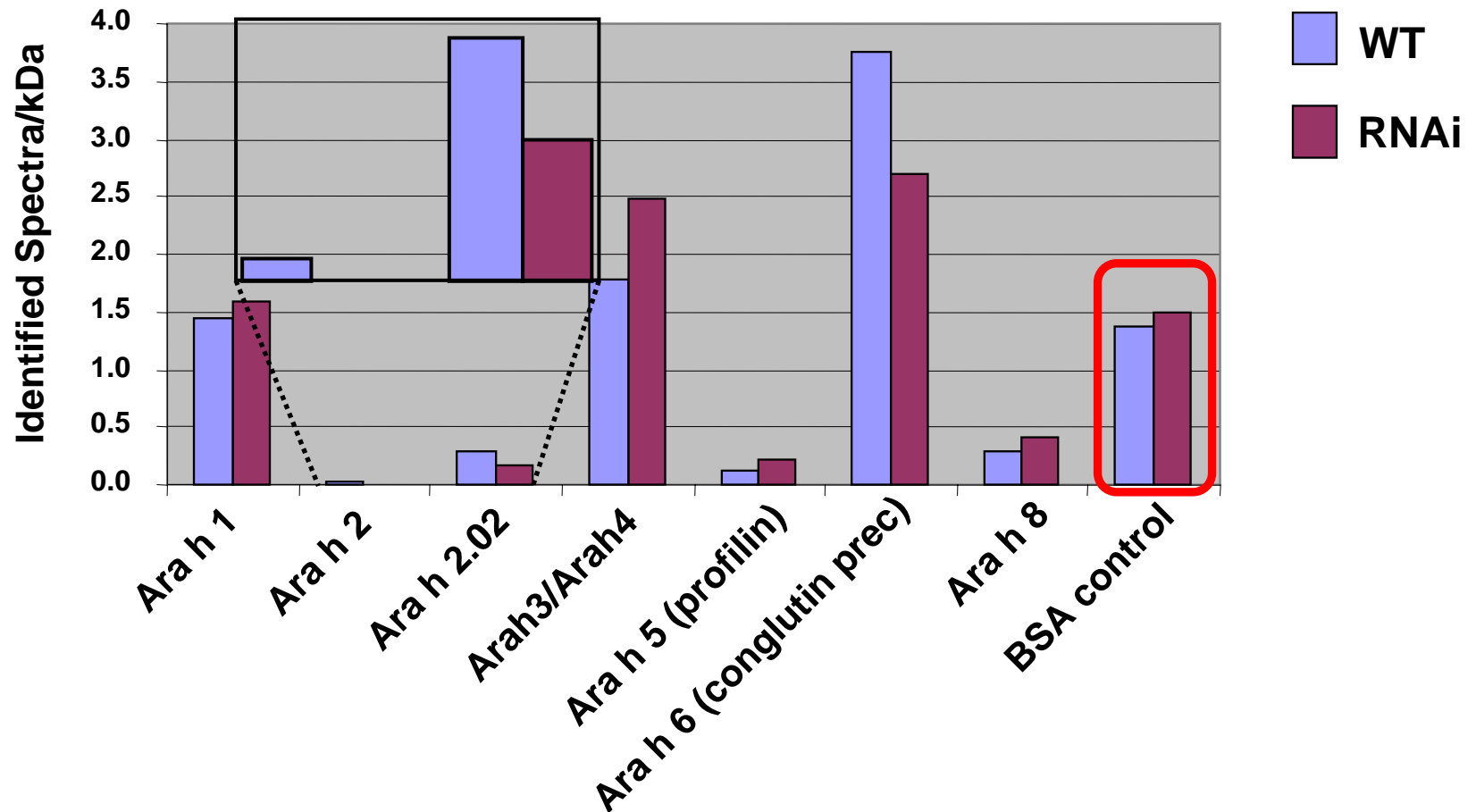
---

- **Analysis of RNAi Ara h 2 peanut lines – collaboration with Peggy Ozias-Akins (UGA-Tifton)**

# Analysis of Ara h 2 RNAi peanut lines

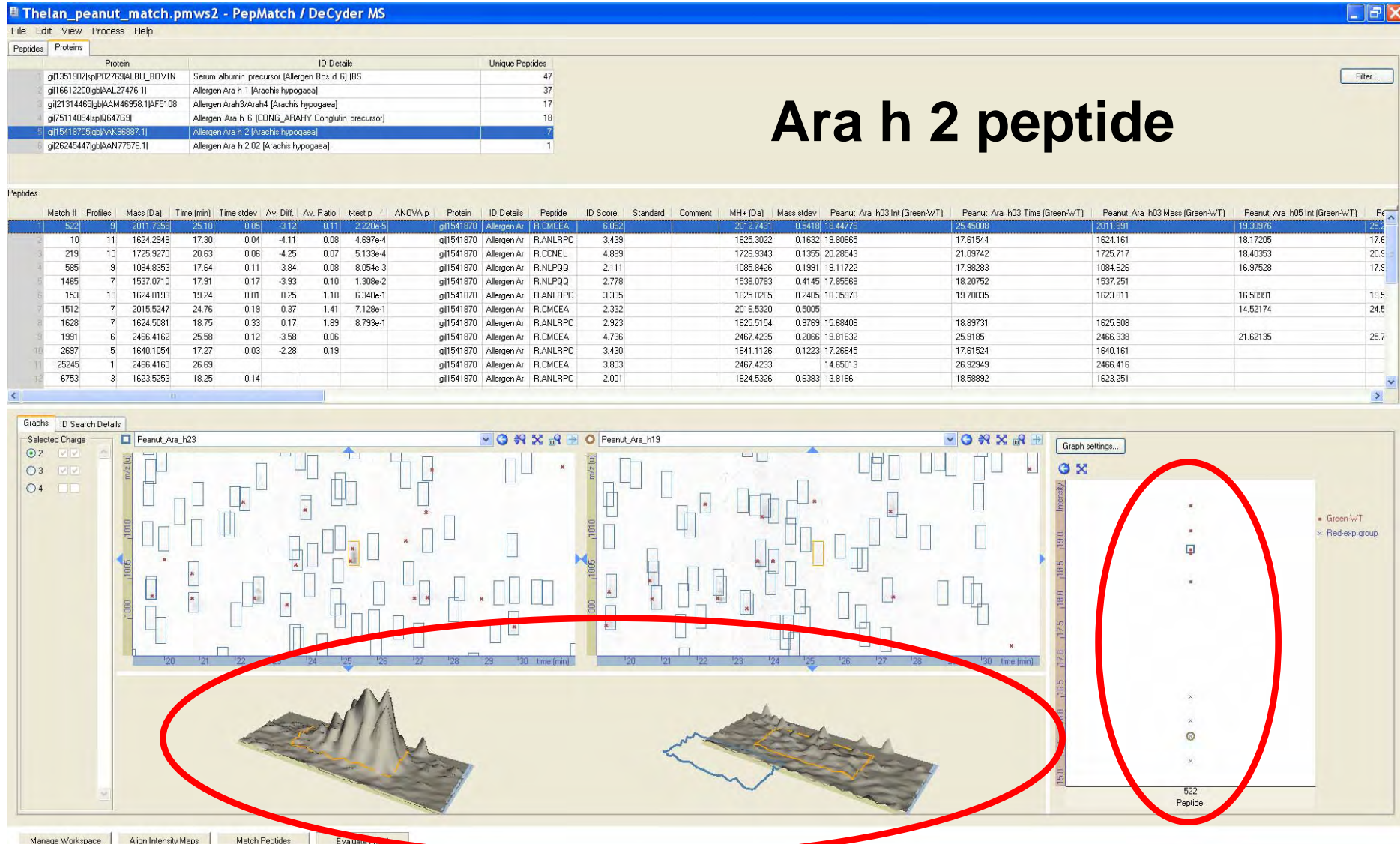
## spectral counting

### Ara h 2 RNAi vs WT allergen content



# Analysis of Ara h 2 RNAi peanut lines

## peak integration





# **Conclusions - label-free proteomics**

---

- Spectral counting offers a broader dynamic range and a more linear response than peak integration**
- When normalized for tryptic peptides (MW) spectral counting is quantitative both within and among samples**
- Spectral counting does not require any software expertise (Scaffold), whereas peak integration (DeCyderMS) is very cumbersome particularly with complex samples**
- Free software for both approaches but they are not supported. Scaffold - \$5k, DeCyderMS \$25k**