Mar17,2009
UAB BMG744

The best proteomics: combination of upfront reduction of sample complexity, reproducible resolution, and effective follow up

Helen Kim, Ph.D.

Dept of Pharmacology & Toxicology
University of Alabama at Birmingham
helenkim@uab.edu
205-934-3880

OUTLINE

I. **Proteomics analysis of actions of a dietary supplement, grape seed extract (GSE), in the brain;**

II. **DIGE analysis of protein differences in rat tissues at different developmental stages to determine basis for cancer risk**

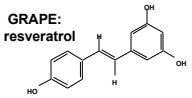
Our principal goal: to understand the molecular basis of human chronic conditions/diseases, to develop prevention or therapies.

Strategy: **a *proteomics* approach**

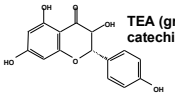
Hypothesis: Actions of "beneficial" agents such as dietary anti-oxidants in normal and disease tissue will reveal subproteomes of proteins "at risk" for disease-relevant changes.

POLYPHENOLS: similar structures among themselves, and with 17β-estradiol

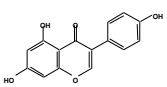
GRAPE: resveratrol



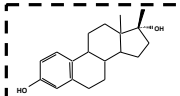
TEA (grapes): catechin



SOY: genistein



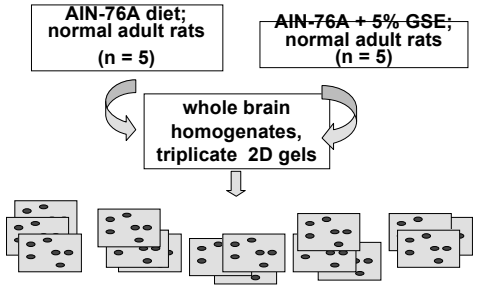
17β-estradiol



RATIONALE for predicting grape seed extract (GSE) polyphenols will have neuroprotective actions:

- **Joseph et al., 1999;**
  - **blueberry extract supplement protected against age-related cognitive impairment**
- **Pan et al., 2000;**
  - **soy isoflavones protected against ovariectomy-induced cognitive dysfunction**
- **Peng et al., 2005;**
  - **our own studies showed GSE enhanced cognition in estrogen-depleted (ovx'd) rats;**

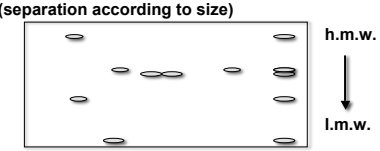
The experiment, and the dataset:



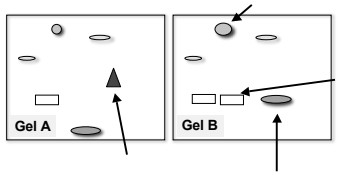
Basics of 2-D electrophoresis

- 1st dimension: isoelectric focusing on flat plastic strips *containing immobilized pH gradients (IPG)*
  (separation according to charge)

- 2nd dimension: (SDS)-polyacrylamide gel electrophoresis
  (separation according to size)

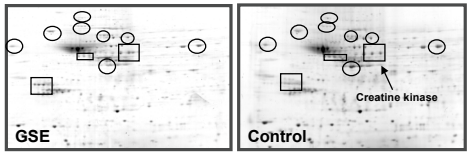


Critical part of 2-D gel proteomics: Image analysis



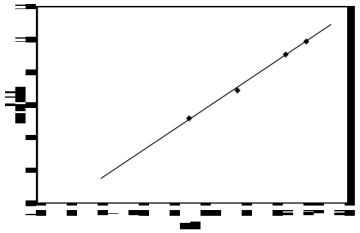
The arrows indicate the types of information *suggested* by image analysis data.

Image analysis indicated spots that differed between the two sets of gels;

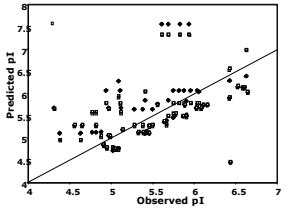


Different in intensity
Different in position (variability)

How do we know the IPG strips work?

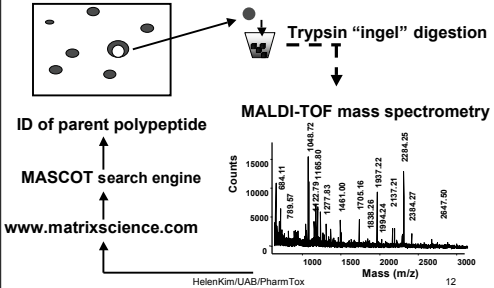


Most proteins focus near but not at their predicted isoelectric point. Why?

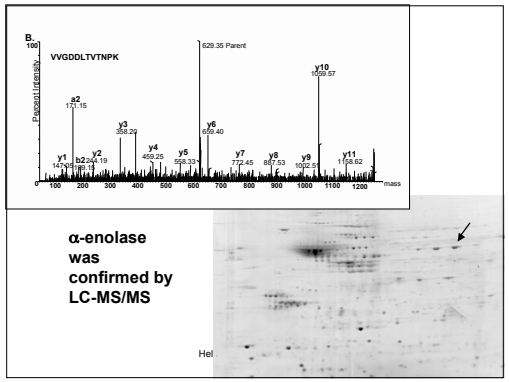


What would it mean if ALL the proteins focussed higher or lower than predicted pIs.

Standard followup procedure once "significant" spots are determined by image and statistical analysis:
Peptide mass fingerprint analysis by MALDI-TOF MS

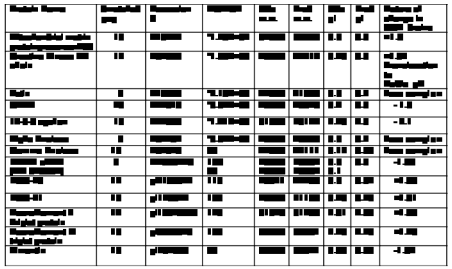


(Kim et al., in Tollefsboll, Methods in Aging Research, in press)



α-enolase was confirmed by LC-MS/MS

Database of protein differences in GSE vs CT brains



(Deshane et al., 2004. J. Agric. Food Chem.)

Initial conclusion: GSE is neuroprotective, since its effects on proteins are counter to the directions of change for the same proteins in disease.

Mitochondrial matrix protein precursor HSP-60 ↑
14-3-3 epsilon protein ↓
Heat shock cognate 70 ↑
Heat shock cognate 71 ↑
Neurofilament triplet protein L ↑
Neurofilament triplet protein M ↑
GFAP ↓
Vimentin ↓
Creatine kinase Brain isoform ↑
Gamma enolase
Alpha enolase
Actin
RIKEN cDNA NM 025944_

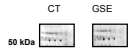
GREEN: folding/ stress response
PURPLE : energy
BLUE: new
ORANGE: cytoskeleton

(Schonberger et al., 2001)
(Tilleman et al., 2002a)
(Tilleman et al., 2002b)

(Deshane et al., 2004. J. Agric. Food Chem.)

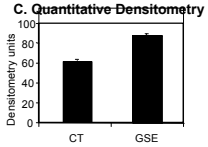
Western blot analysis of 2D gel image and statistical analysis

A. Stained gel for HSP-60



B. Western Blots

C. Quantitative Densitometry



Don't discount the power of conventional/"old" methods..

Validation of protein identifications and quantitations

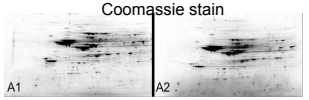
Protein	LC-MS/MS	Western blot
CK-BB	+	+
Hsp60	+	+
GFAP	--	
Actin	--	+
NFL-M	+	--
α-enolase	+	--
γ-enolase	+	--
Hsc70	--	--
Hsc71	--	--
14-3-3e	--	+
NFL-L	--	+

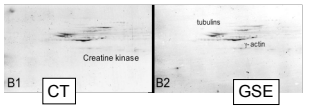
Interesting note: Only for spots from 2D gels, journals accept PMF data, without sequence data, because of the information from the 2D gel (mass, pI).

(Kim et al., 2005, in Luo and Packer, Oxidative Stress & Neurodegeneration)

Proteomics can indicate protein oxidations: an increasingly important area of study in chronic diseases

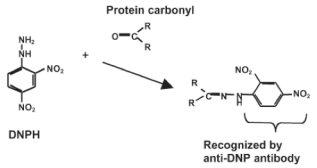
Coomassie stain



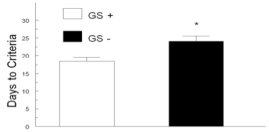


Western blot with anti-DNP

To study protein oxidations, must understand chemistry involved, make use of biological reagents



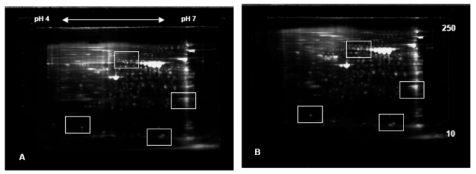
The strongest validation of proteomic data can come from completely different (orthogonal) experimental procedures that usually test for function.



In this case, GSE fed to young female rats enhanced learning and memory over those fed control diets. This data was consistent with GSE having neuroprotective actions.

(Peng et al., 2005)

DIGE (difference gel electrophoresis): powerful, but quality control issues need to be addressed to ensure robust data.



Issue addressed here: does the same sample bind the two cy-dyes, cy3 and cy5, equivalently?

How do you deal with running multiple samples on 2D gels as objectively as possible

Gel #	1	2	3	4	5
Cy 2	Internal Standard	Internal Standard	Internal Standard	Internal Standard	Internal Standard
Cy 3	Day 21	Day 50	Day 21	Day 50	Day 21
Cy 5	Day 50	Day 21	Day 50	Day 21	Day 50

Grid assigns random pairs of samples per gel;
In this experiment, the different days were swapped to make sure that there wasn't preferential dye binding of one day over the other.

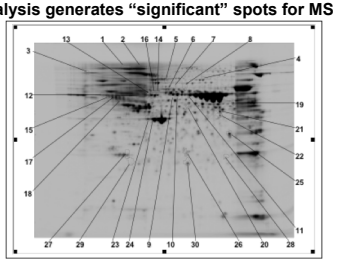
Real data from a DIGE experiment

Table 2. Fold change, significance and FDR of proteins that were significantly different between days 21 and 50

Master Spot	Figure 4 spot number	Protein identification	Fold change, 50/21d	T-Test (p value)	FDR	Pred. Mass. (Da)	Mature mass (Da)	Obs. Mass (Da)	Pred. pI	Mature pI	Obs. pI
162	1	alpha-1-inhibitor III	1.2925	0.0024	0.0500	165,038	161,076	132,210	5.70	5.87	4.97
162	1	procollagen, type XIV, alpha 1 (predicted), isoform CRA_a	1.2925	0.0024	0.0500	192,494	*	132,210	5.08	*	4.97
163	2	similar to Collagen alpha-1(VI) chain precursor	1.0945	0.0003	0.0333	108,738	*	132,210	5.21	*	5.14
215	3	alpha-1-inhibitor III	2.8150	0.0004	0.0252	165,038	161,076	122,753	5.70	5.87	4.31
261		PREDICTED: similar to Dynein, heavy chain at 89D CG1942-PA	1.2706	0.0107	0.1045	438,140	532,252	117,590	5.87	6.06	5.41
293		heat shock protein 4	1.4585	0.0101	0.1058	93,997	94,057	110,035	5.13	5.13	5.46
302		heat shock protein 4	1.3868	0.0145	0.1289	93,997	94,057	109,179	5.13	5.13	5.41
309		similar to alpha glucosidase,2 alpha neutral subunit	0.8315	0.0392	0.2249	109,300	*	108,753	5.76	*	6.11
310	4	similar to alpha glucosidase,2 alpha neutral subunit	0.8153	0.0047	0.0704	109,300	*	109,179	5.76	*	6.16
344	5	PREDICTED: similar to ribosome binding protein 1 isoform 1	1.4174	0.0082	0.0955	109,049	*	103,388	5.79	*	5.56

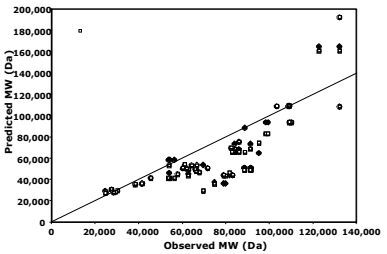
(from Kim et al., J. Proteome Res, 2008)

Image analysis generates "significant" spots for MS analysis



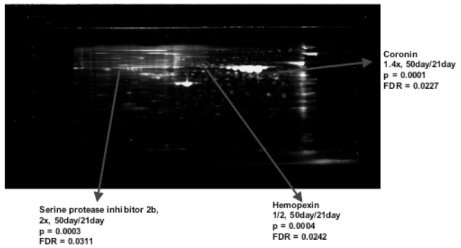
On this gel, half the spots didn't contain enough protein for MS analysis; solution?

Analysis of observed mw over predicted mw: many polypeptides run at mw higher than predicted.

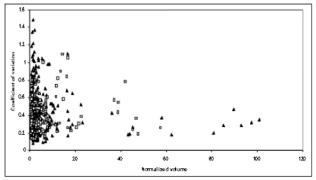


What would it mean if every single spot ran at a higher mw than predicted?

Some protein markers for rat development



Coefficient of variance: statistical tool that assesses the extent of variation over the range of samples:

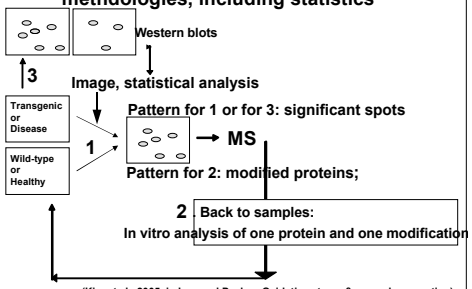


Bottom line here: there was no difference in coefficient of variance between the different groups, thus the differences were due to the true means. (Kim et al., J Prot Res, 2008)

Take home lessons, part I

- **Proteomics suggests GSE has pleiotropic effects in the brain:**
  - **gene expression/protein turnover;**
    - **(Deshane et al., J. Agric. Food Chem.,2004)**
    - **How would you follow up?**
  - **protein oxidations;**
  - **Informatics helps you relate your data to the rest of the world; in the case of actions of GSE, we know that many of the proteins affected in normal brain are also differentially expressed in AD brain.**
    - ***What will make the GSE effects consistent with neuroprotection?***
    - ***How do we PROVE that any one effect of GSE actually prevents neuronal impairment?***

Optimal data analysis: Iterative process may start with proteomics but can/should include other methdologies, including statistics



(Kim et al., 2005, in Luo and Packer, Oxidative stress & neurodegeneration)

Take home lessons, part II

- **DIGE: powerful, but room for lots of quality control;**
- **In the case of the data presented, these were differences in protein abundance between two developmental ages in female rat tissues;**
- **How do we know which are important in the actions of a carcinogen which is given at the later age?**
- **How do we know which cell type these occurred in, and minimize blood protein (albumin) differences?**