Multi-way Analysis of 2D Liquid Chromatographic Metabolomics Data

Sarah E. G. Porter & Sarah C. Rutan

Virginia Commonwealth University, Department of Chemistry

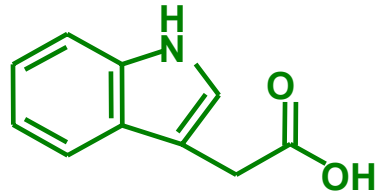
Dwight R. Stoll & Peter W. Carr

University of Minnesota, Department of Chemistry

Jerry D. Cohen

University of Minnesota, Department of Horticultural Science

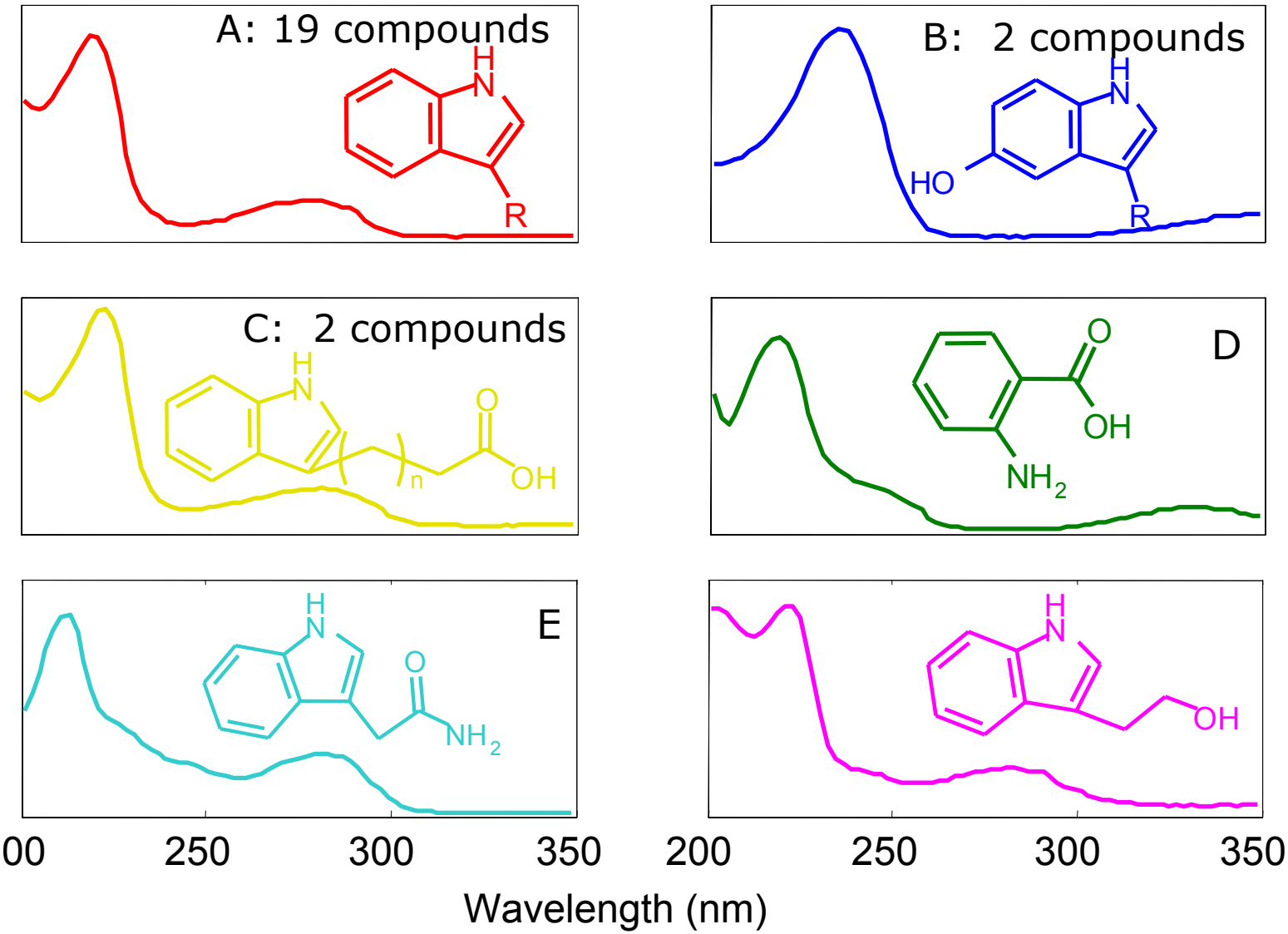
Indole-3-Acetic Acid (IAA) in Plants



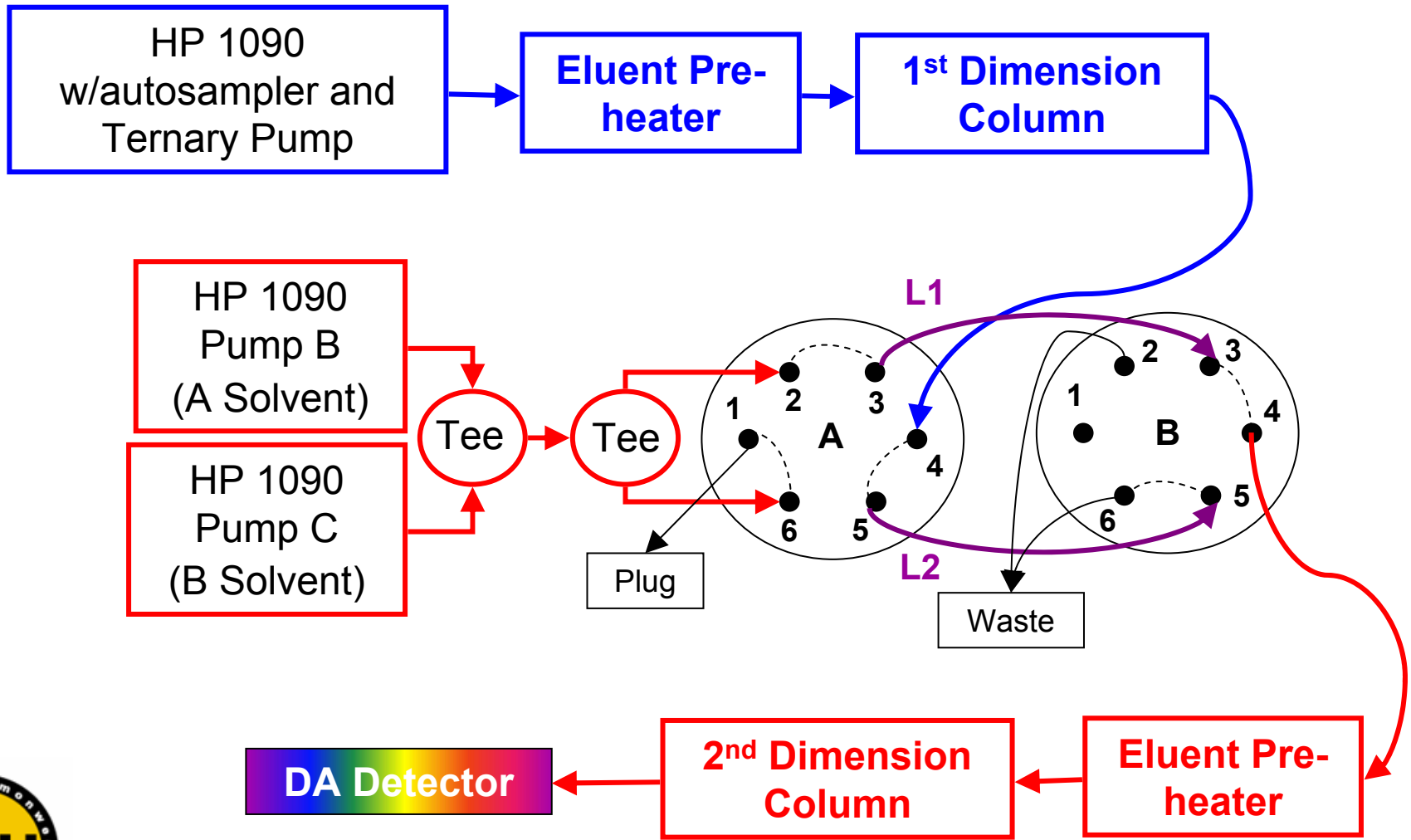
- Primary growth hormone responsible for cell division and elongation, flowering, root initiation, fruit ripening, promoting vascular tissue growth, controlling premature abscission of leaves and fruit
- Synthesized from tryptophan by a variety of pathways
- **Metabolic Profiling** – the identification and quantification of a selected group of metabolites in a biological system
- **Metabolomics** – unsupervised comparison of different biological samples to elucidate differences in metabolite levels
- **2DLC with diode array detection (DAD) was used to compare mutant and wild type maize samples to 26 indolic metabolite standards**

Wright, et al. *Science*, **1991**, *254*, 998-1000
Fiehn, O. *Plant Mol. Biol.* **2002,** *48*, 155-71

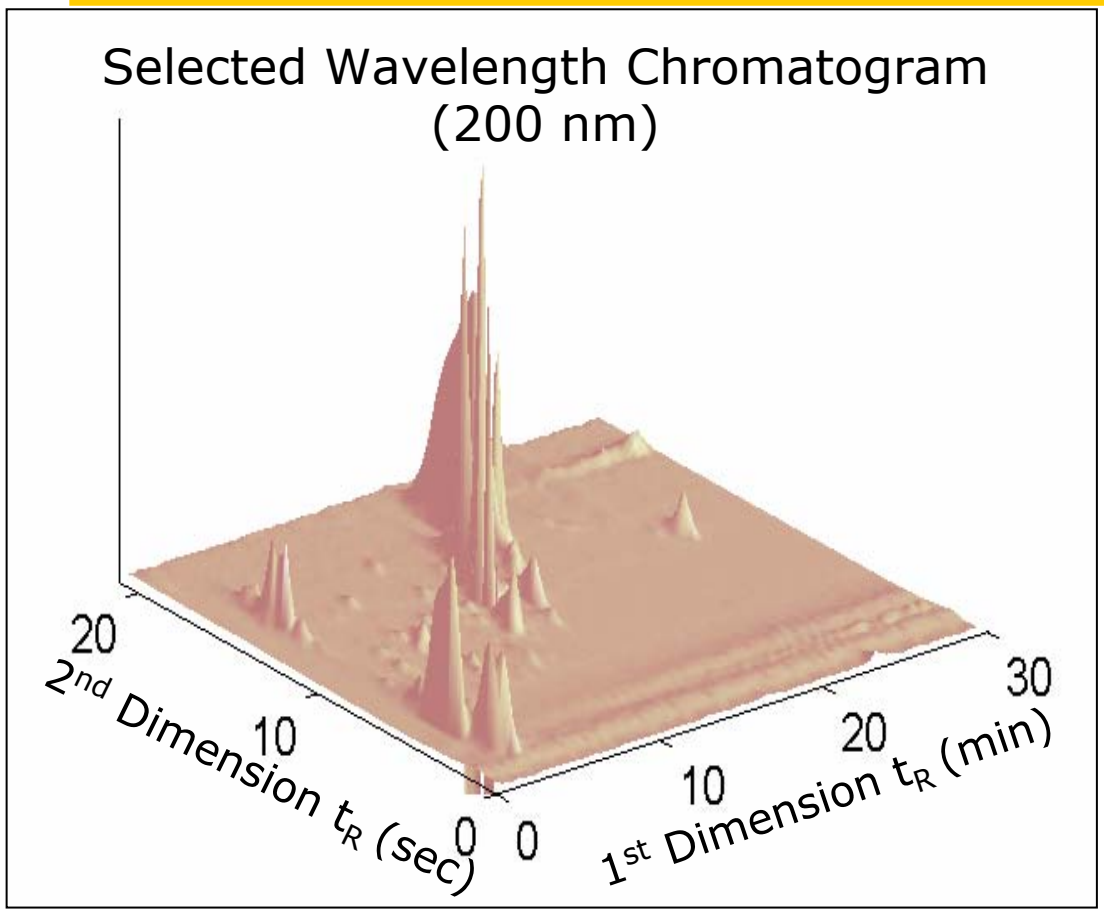
6 spectral components of 26 indolic standards

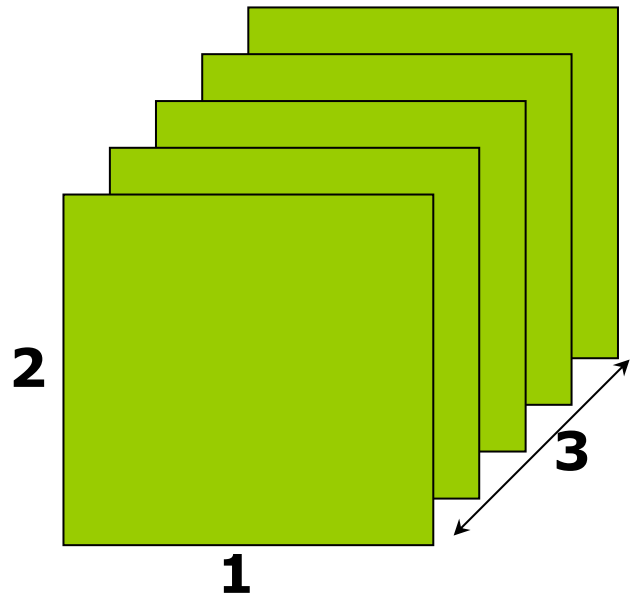


2DLC Instrumentation Capable of Gradient Elution in Both Dimensions



2DLC Data Structure – Three Way Data

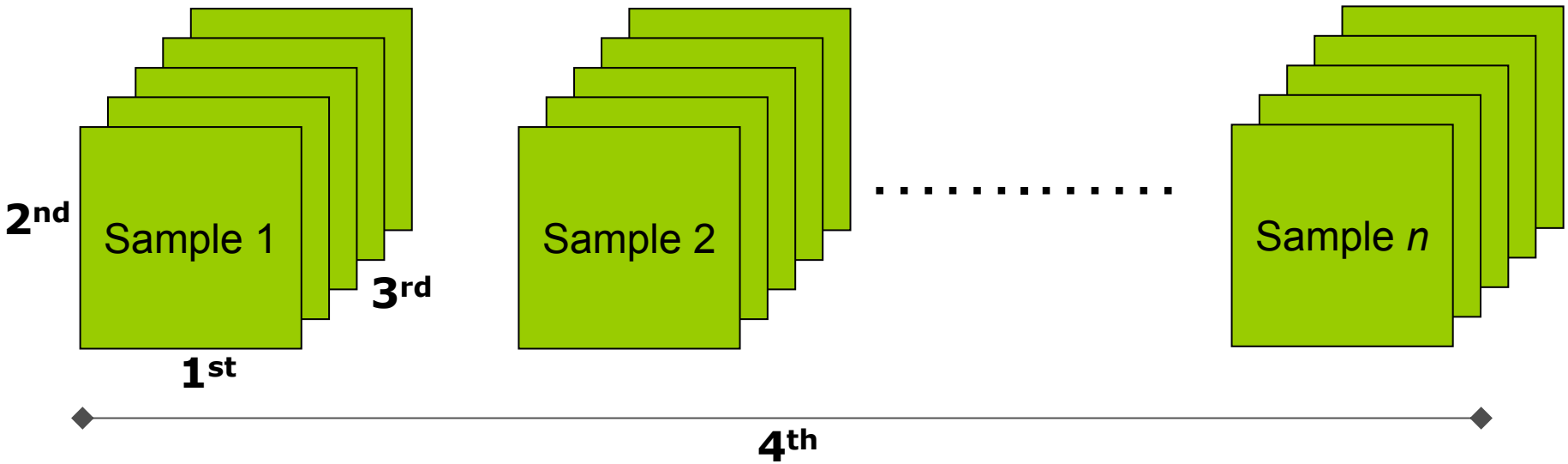




Data Dimensions:

1. 1st Dimension Retention Time (min)
2. 2nd Dimension Retention Time (sec)
3. Wavelength (nm)

Four-Way Quadrilinear Data



1st Dimension, Retention Time, minutes
2nd Dimension, Retention Time, seconds
3rd Dimension, Wavelength, nm
4th Dimension, Sample number

The instrument response of a pure component in all domains is unique, consistent, and independent of the presence of other species

Booksh, K. S. et al., *Anal. Chem.* **1994,** *66*, 2561-69.

Description of Samples

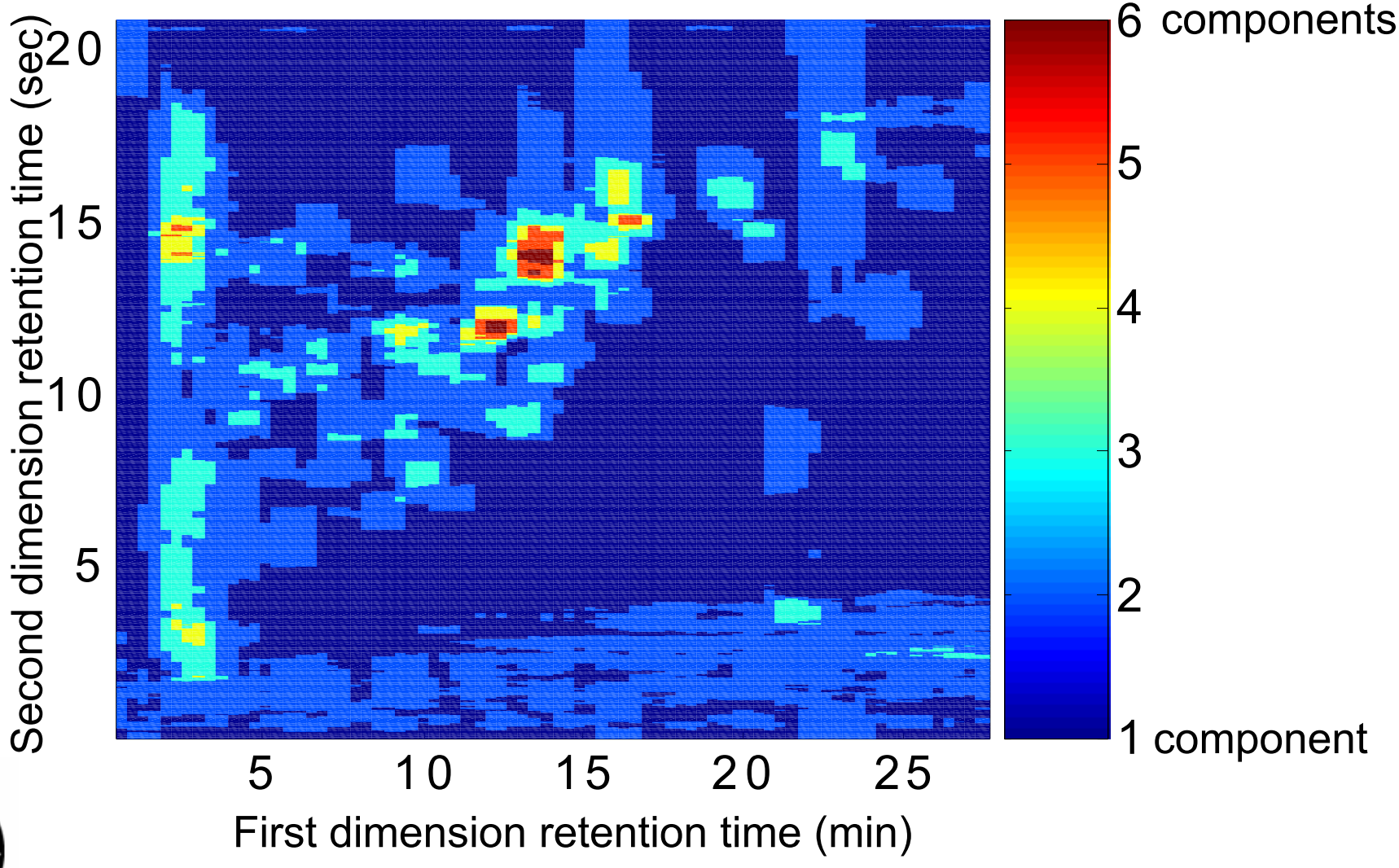
- Mobile phase blank
- Standard mixture containing 26 indoles
- Duplicate wild type maize seedling samples
- Duplicate *orp* mutant maize seedling samples
  - Lacks gene for tryptophan synthase β
  - IAA is produced via tryptophan-independent pathway

Fixed Size Image Window – Evolving Factor Analysis[a]

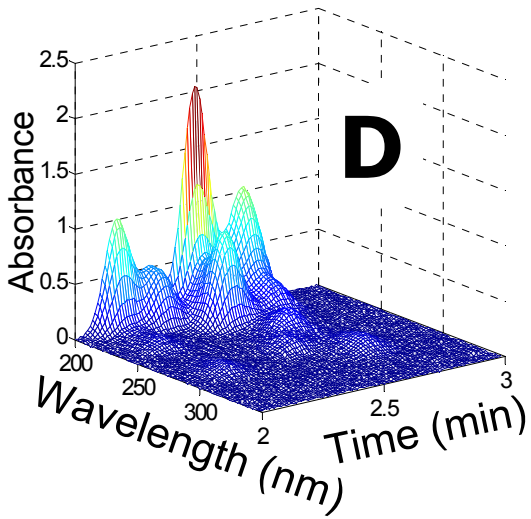
- FSIW-EFA uses sections of an image and performs factor analysis on a moving window
- Rank information is *local* – results estimate the complexity of an image for exploratory analysis
- Traditional EFA approaches would require unfolding of the three-way data set and loss of complex spatial structure
- **Can be used to select sections of data for subsequent analysis**

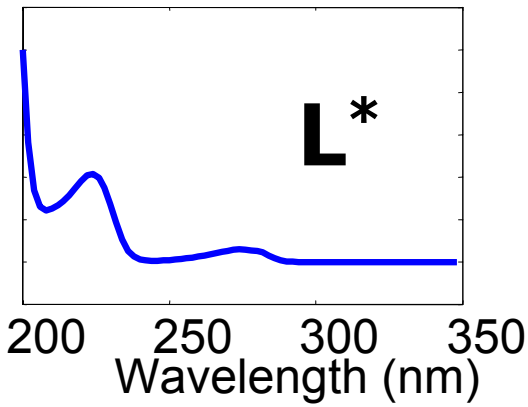
[a]de Juan, A. et al. *Chemom. Intell. Lab. Syst.*, **2005**, *77*, 2005, 64-74

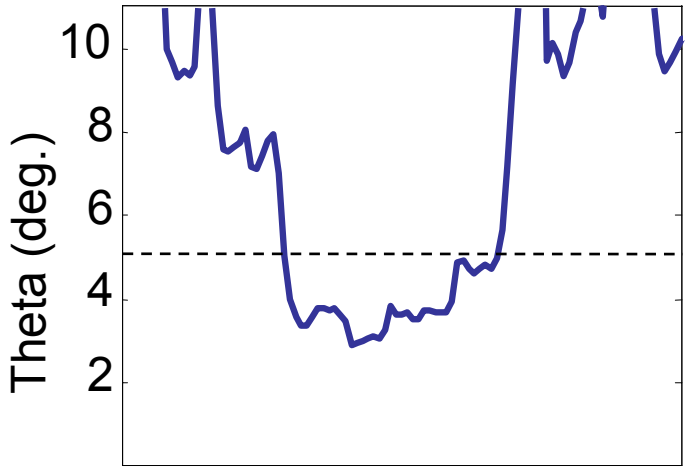
Summed Rankmap:
Standards, Wild Type, Mutant



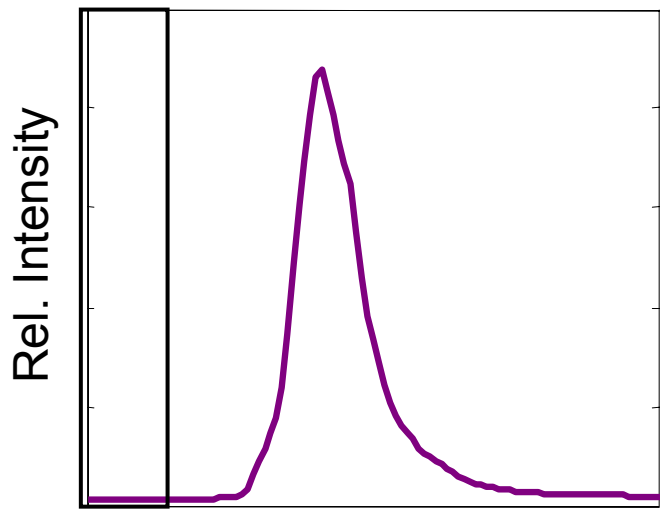
Window Target Testing Factor Analysis





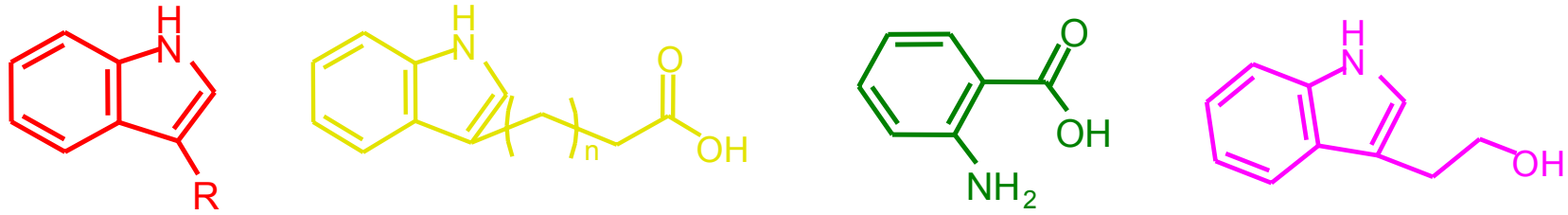


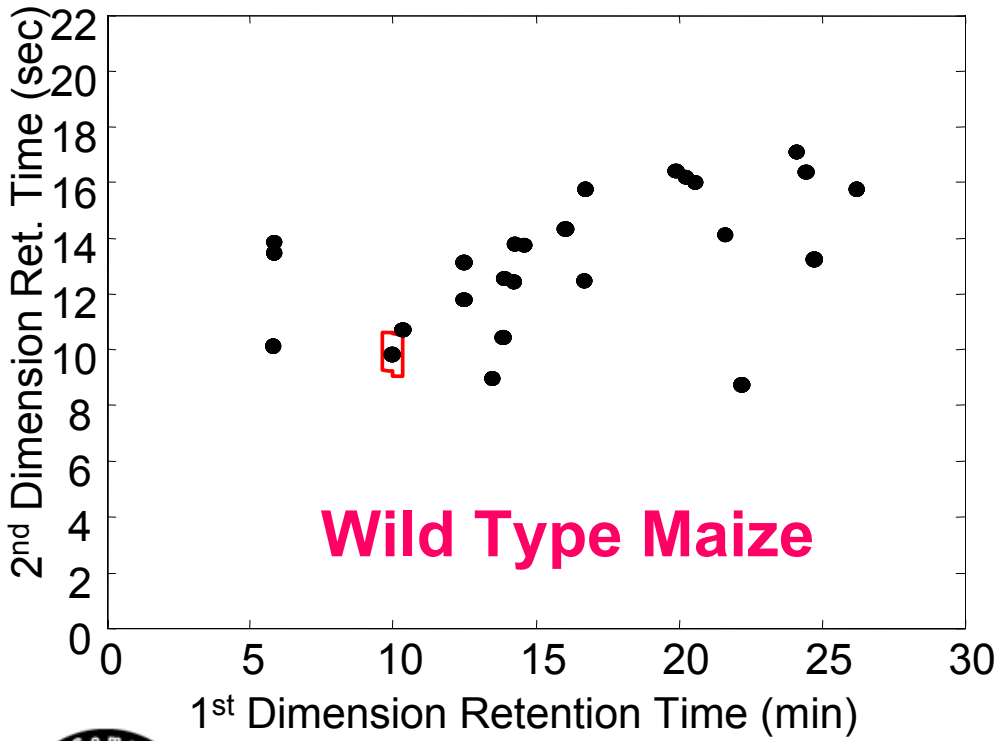
1. SVD: $\mathbf{D} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^{\mathbf{T}}$
2. Target test: $\mathbf{T} = (\tilde{\mathbf{V}}^{\mathbf{T}} \cdot \mathbf{V})^{-1} \cdot \tilde{\mathbf{V}}^{\mathbf{T}} \cdot \mathbf{L}^{*}$
3. Rotate: $\hat{\mathbf{L}}^{*} = \tilde{\mathbf{V}}^{\mathbf{T}} \cdot \mathbf{T}$
4. Correlate: $\theta = \cos^{-1}\rho$

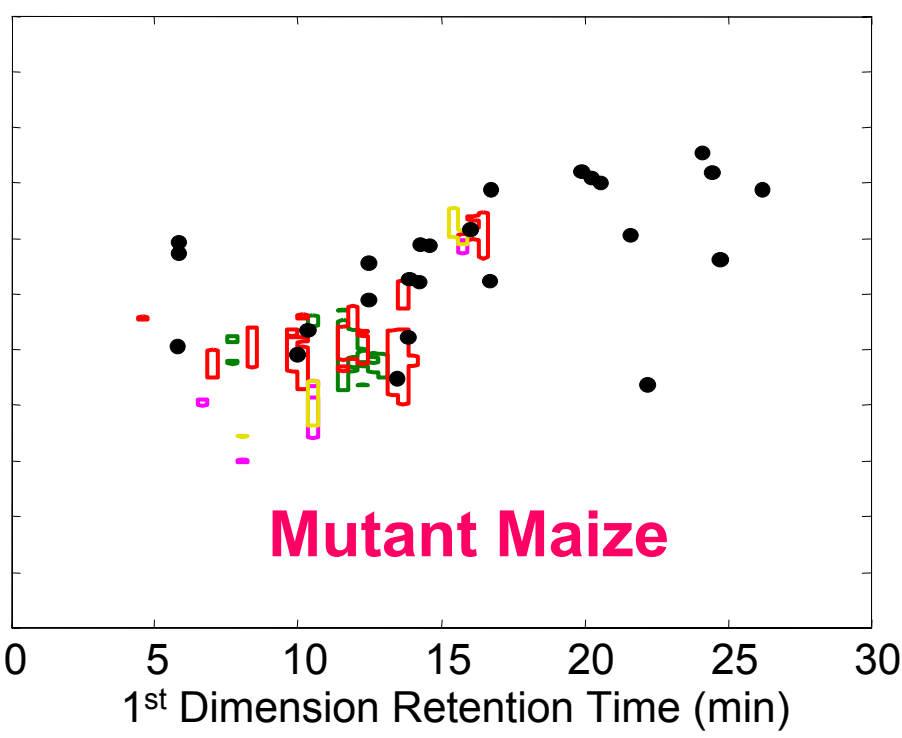


Lohnes, et al. *Anal. Chim. Acta* **1999,** *389*, 95-113

Results of Qualitative Analysis

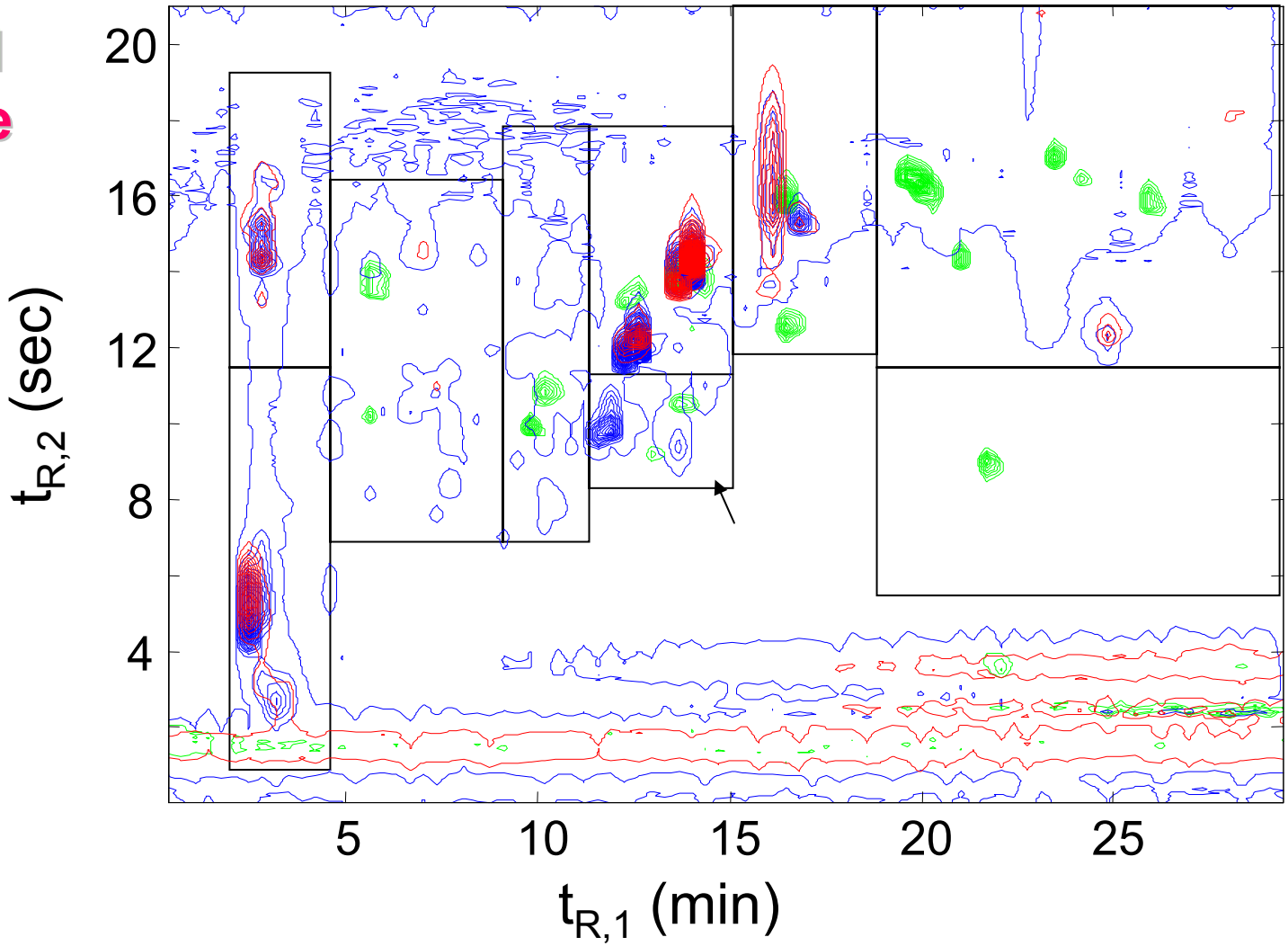






2D-Chromatograms – 220 nm

Standard
Wild-type
Mutant

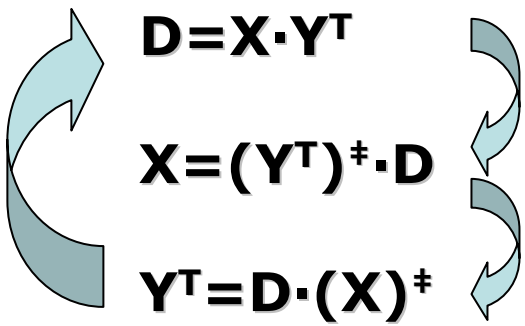


Modeling with PARAFAC and fALS

The **four-way** PARAFAC model is represented mathematically as a sum over all of the elements of each mode (where *c* is the rank of the data)[a]:

$$d_{hijk} = \sum_{f=1}^{c} w_{hf} x_{if} y_{jf} z_{kf} + e_{hijk}$$

The PARAFAC model is solved using alternating least squares (ALS)



ALS with *flexible constraints* (fALS)[b] allows the selective application of the unimodality constraint to selected components

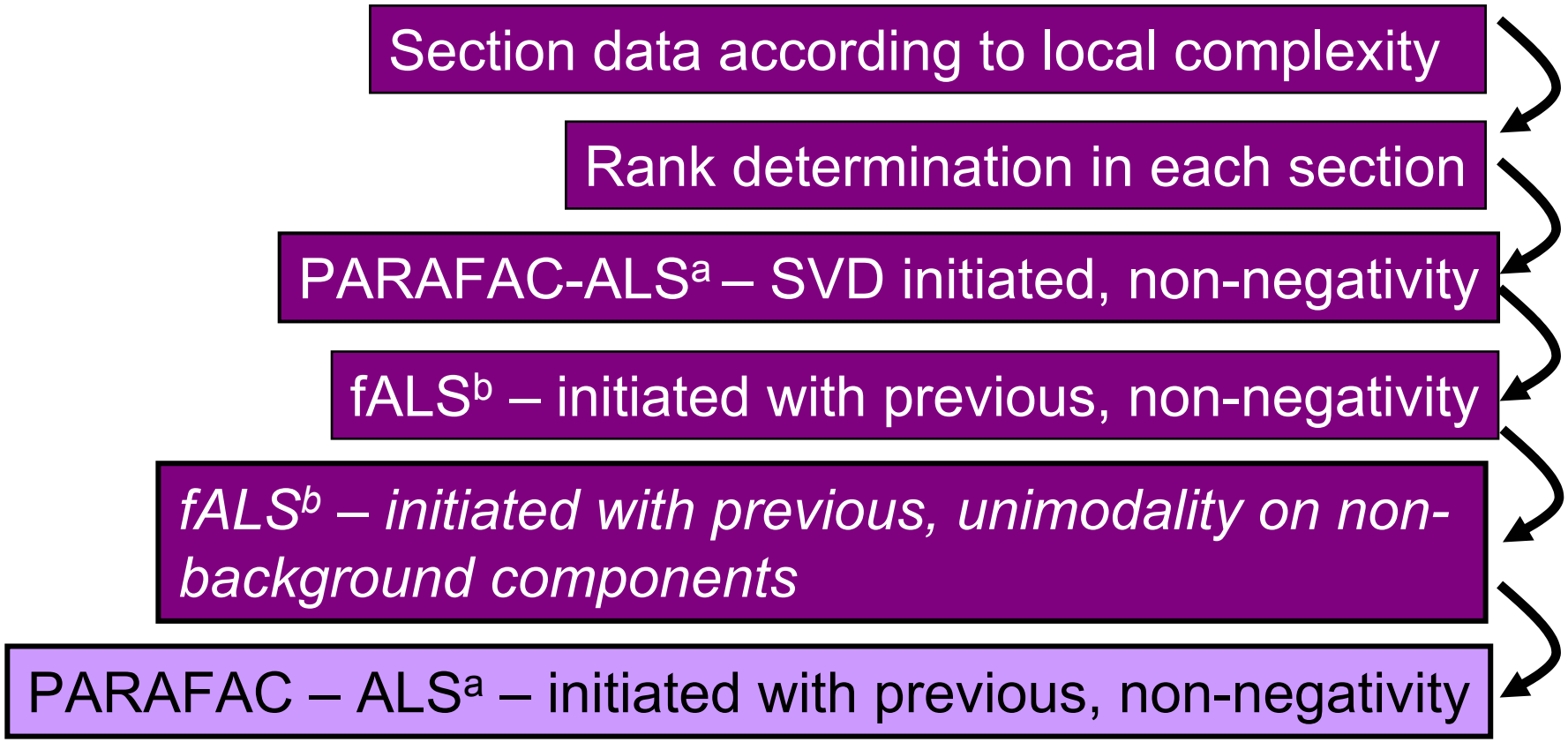
[a]Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2000,** *52*, 1-4
[b]Bezemer, E.; Rutan, S. C. *Chemom. Intell. Lab. Syst.* **2006,** *81*, 82-93

PARAFAC-ALS vs. fALS

- PARAFAC-ALS[a]
  - Multilinearity required
  - Constraints provide LS solutions with guaranteed convergence
  - Constraints must be applied to all components
- fALS[b]
  - Multilinearity optional
  - Constraints may be ad-hoc, but not LS optimal
  - Constraints can be applied to selected components

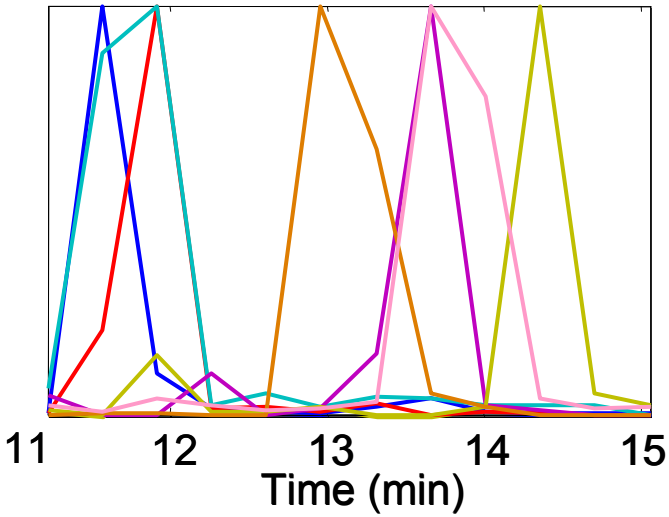
[a]Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2000,** *52*, 1-4
[b]Bezemer, E.; Rutan, S. C. *Chemom. Intell. Lab. Syst.* **2006,** *81*, 82-93

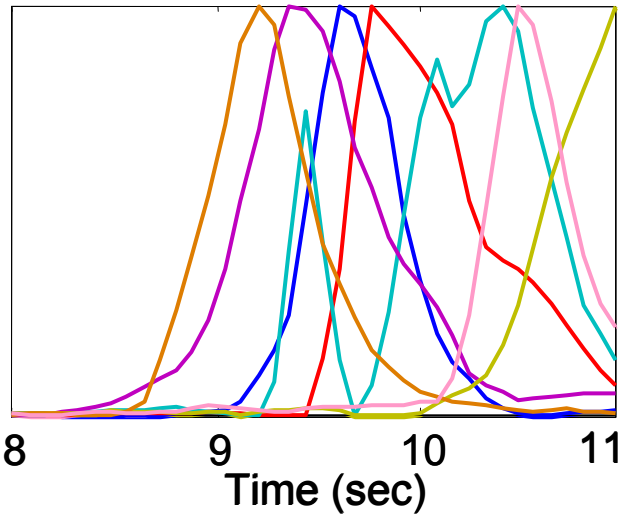
Data Analysis Procedure

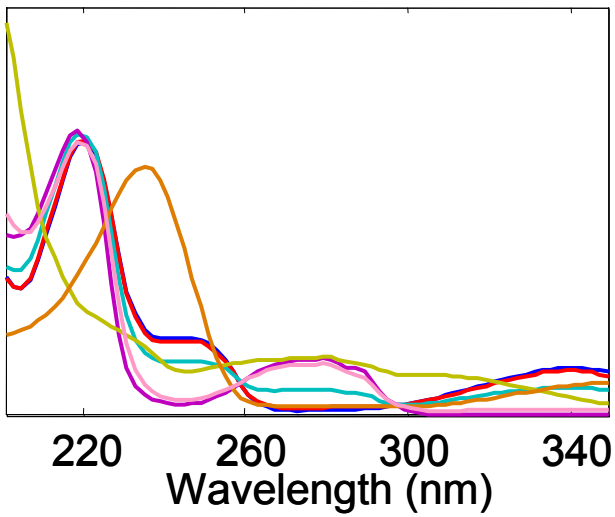


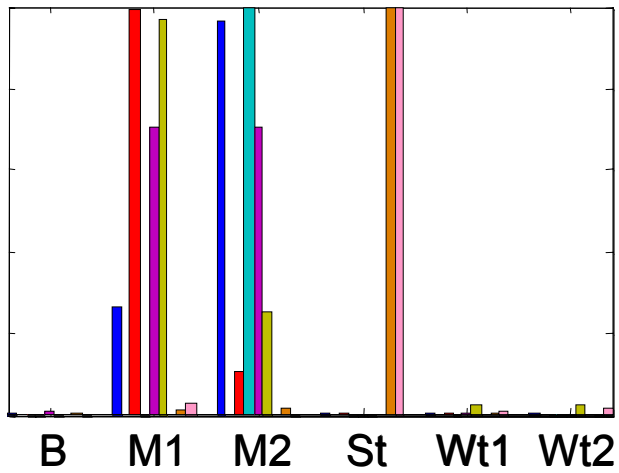
[a]Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2000,** *52*, 1-4
[b]Bezemer, E.; Rutan, S. C. *Chemom. Intell. Lab. Syst.* **2006,** *81*, 82-93

PARAFAC Results for Selected Section

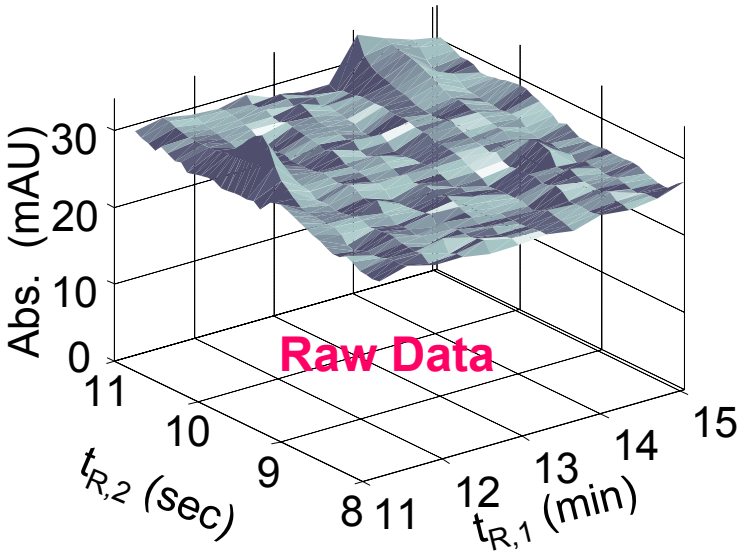


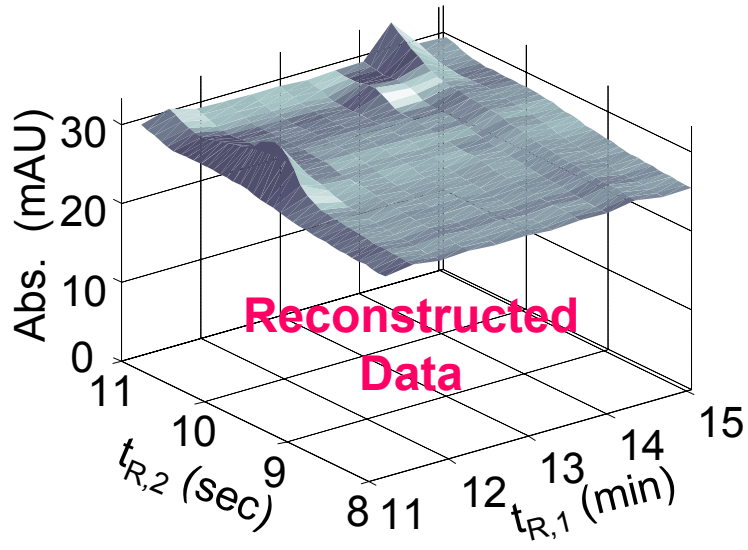


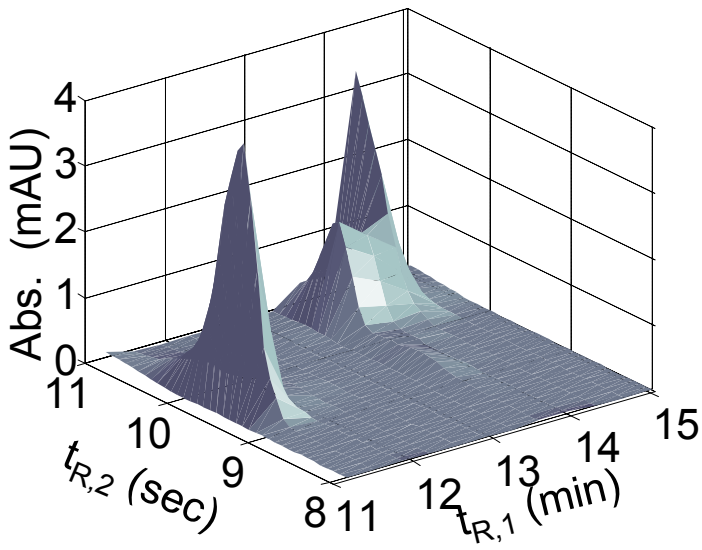




Reconstructed PARAFAC Results – Wt1

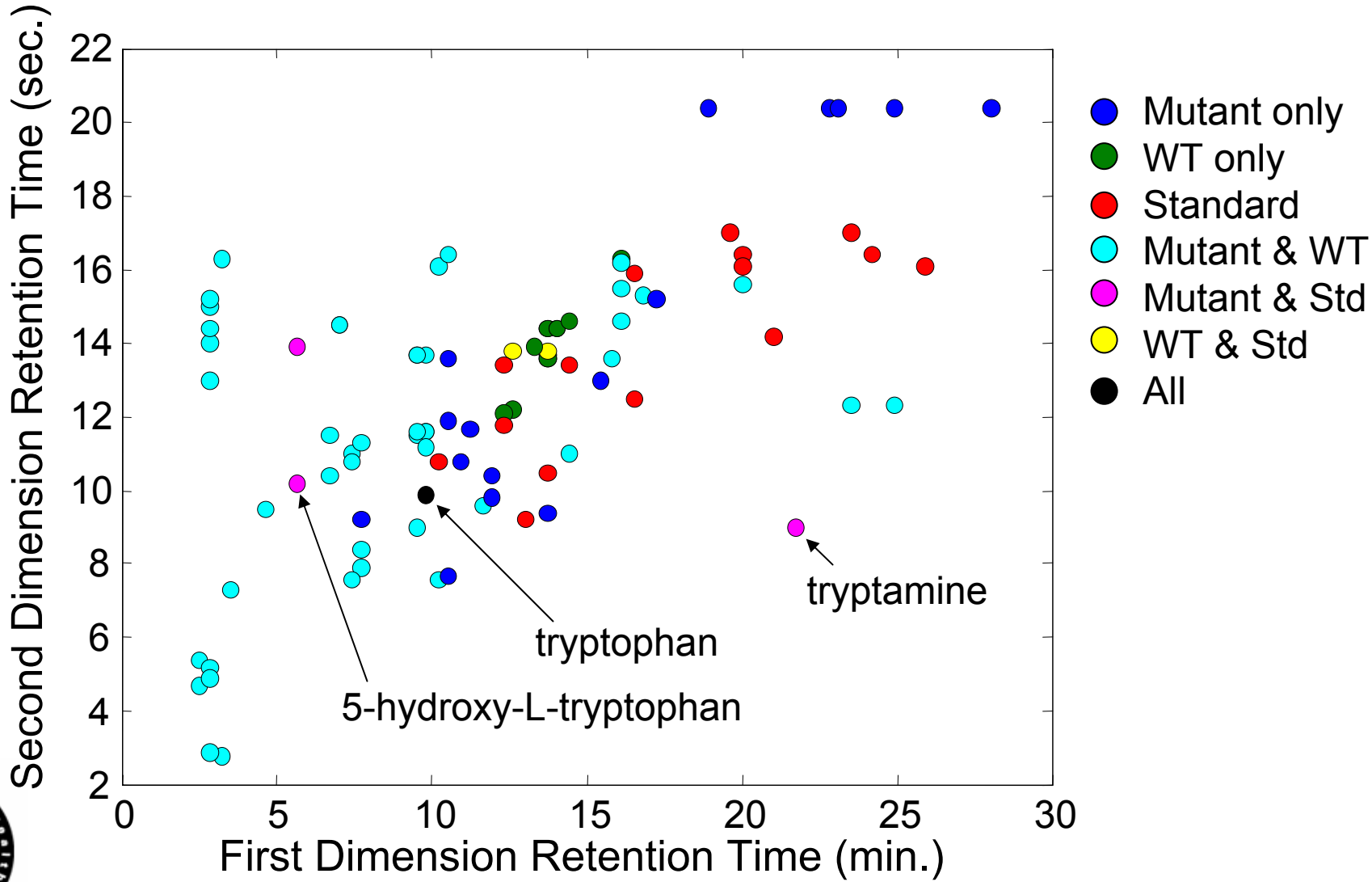






Reconstructed data with background components removed

Results of Analysis of Entire Data Set



Quantitative Results of PARAFAC Analysis

- 95 distinct chromatographic peaks were resolved
- Many peaks showed differential expression between wild type and mutant samples

	Mutant 1	Mutant 2	WT1	WT2
5-hydroxy-L-tryptophan	ND*	0.6	ND*	ND*
Tryptophan	**1.4**	**1.1**	**1.8**	**1.6**
Indole-3-acetyl-L-alanine	ND*	ND*	0.8	0.3
Tryptamine	1.9	ND*	ND*	ND*

*ND – not detected

Quantitative results are in μg of indole per gram of plant material

Selectivity in Multi-way Analysis[a]

- Messick, Kalivas, Lang (MKL)[b]
  - PARAFAC
  - *Appropriate when all components are calibrated*
  - $SEL_n = \{[(X^TX)*(Y^TY)]^{-1}\}_{nn}^{-1/2}$
- Ho, Christian, Davidson (HCD)[c]
  - GRAM
  - *Appropriate when only the target analyte is calibrated*
  - $SEL_n = \{[(X^TX)^{-1}]_{nn}[(Y^TY)^{-1}]_{nn}\}^{-1/2}$

[a]Olivieri, A. C., *Anal. Chem.* **2005**, *77*, 4936-3946
[b]Messick, N. J.; Kalivas, J. H.; Lang, P. M. *Anal. Chem.* **1996**, *68*, 1572-1579
[c]Ho, C.-N.; Christian, G. D.; Davidson, E. R., *Anal. Chem.* **1980**, *52*, 1071-1079

Selectivity in Multi-way Analysis

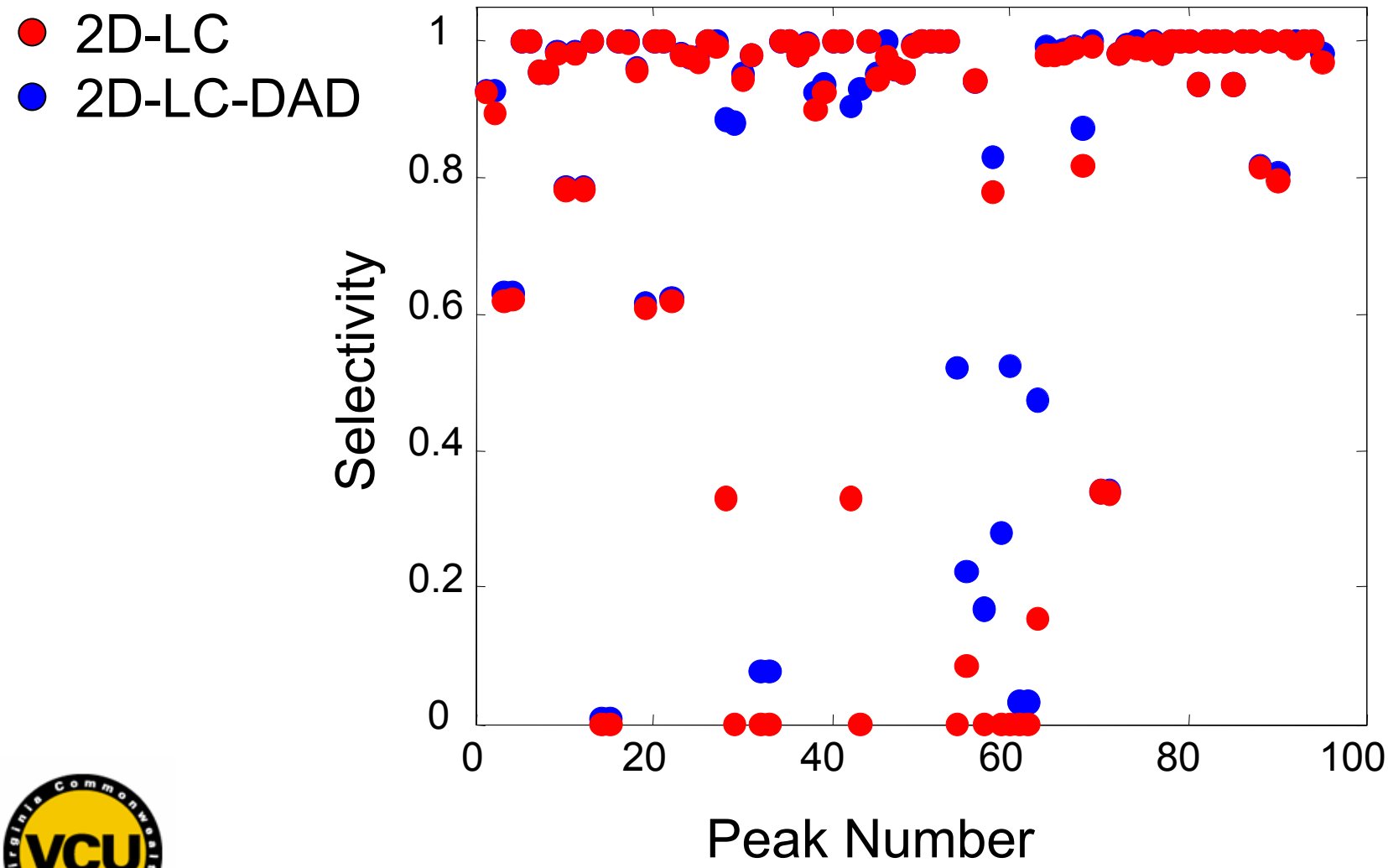
- The selectivity calculations predict the relative decrease in precision that is observed relative to that observed for a pure sample.
- Olivieri observed that for some multi-way situations, neither selectivity formulation predicted the results of Monte Carlo calculations.
  - SEL_{MKL} – upper limit of selectivity
  - SEL_{HCD} – lower limit of selectivity

Olivieri, A. C., *Anal. Chem.* **2005**, *77*, 4936-3946

Calculation of Selectivity for 2DLC vs. 2D-LC-DAD

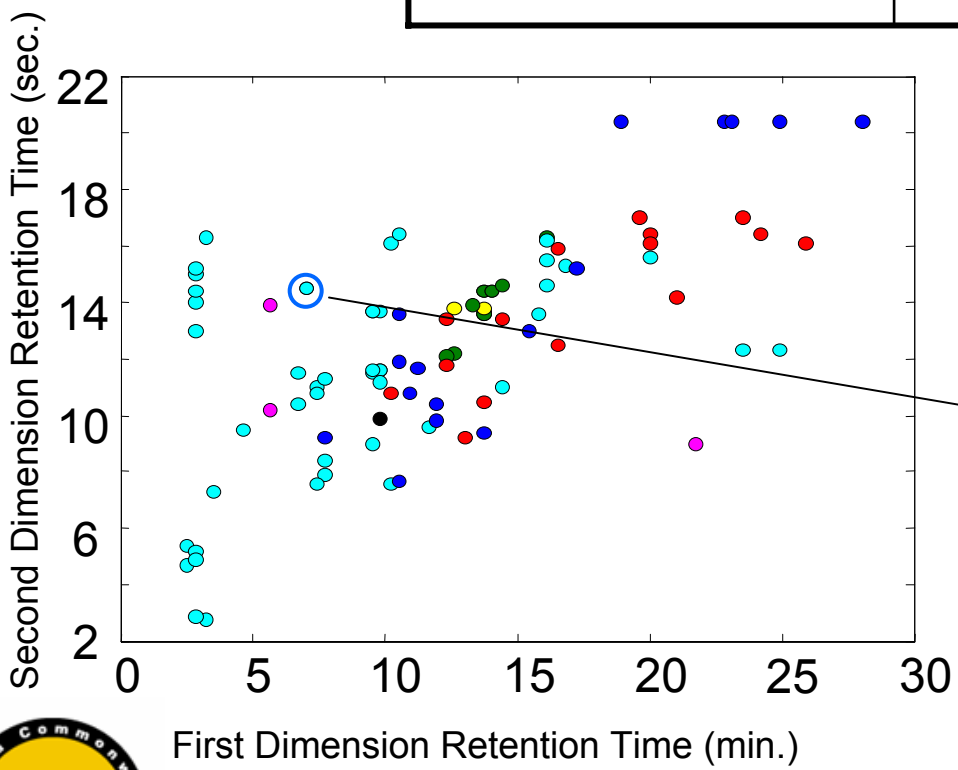
- **X**, **Y**, and **Z** simulated to approximate the real data
- **X** and **Y** – 1st and 2nd dimension retention profiles, using resolved retention times and simulated, Gaussian peaks
- **Z** – spectral profiles – as resolved by PARAFAC
- Background components omitted from the analysis

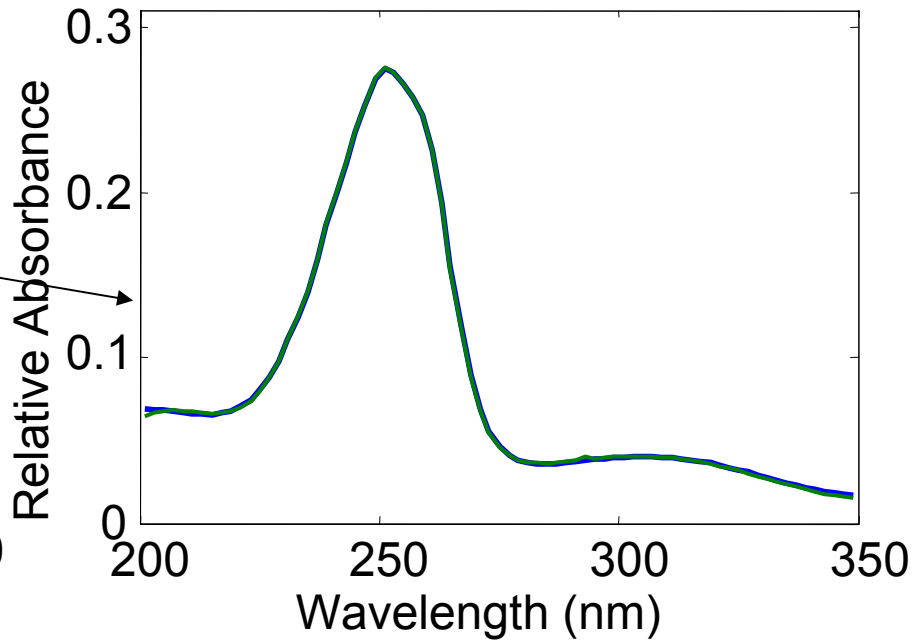
MKL Selectivity



MKL Selectivity – 2D-LC vs. 2D-LC-DAD

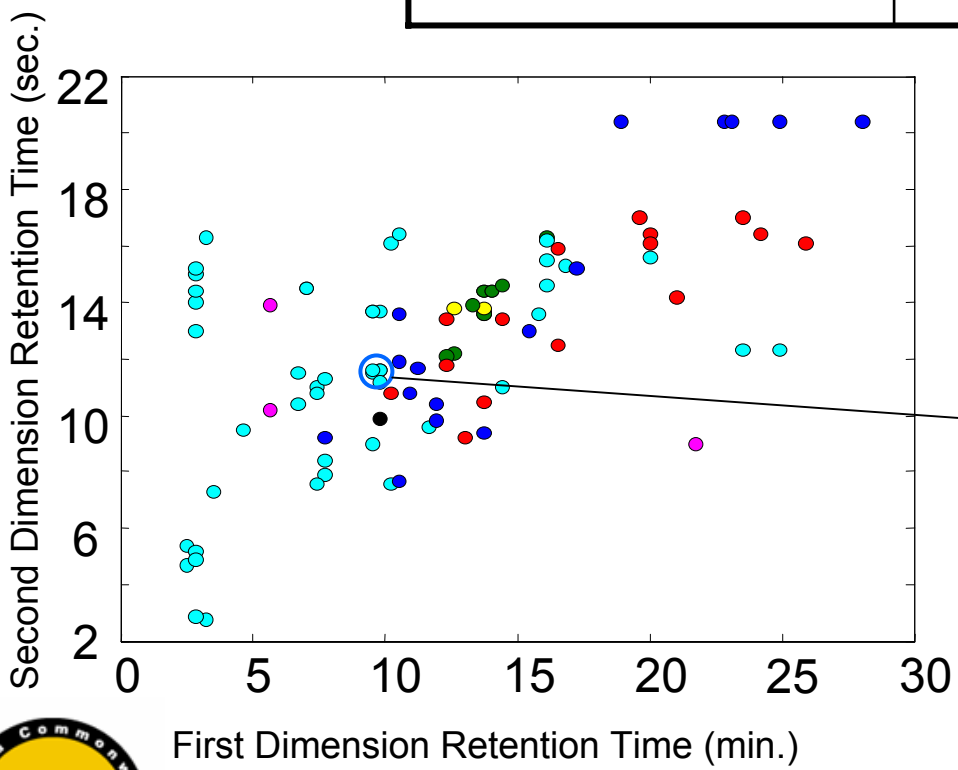
	Peak 1	Peak 2
2D-LC SEL	1×10^{-5}	1×10^{-5}
2D-LC-DAD SEL	8×10^{-3}	8×10^{-3}

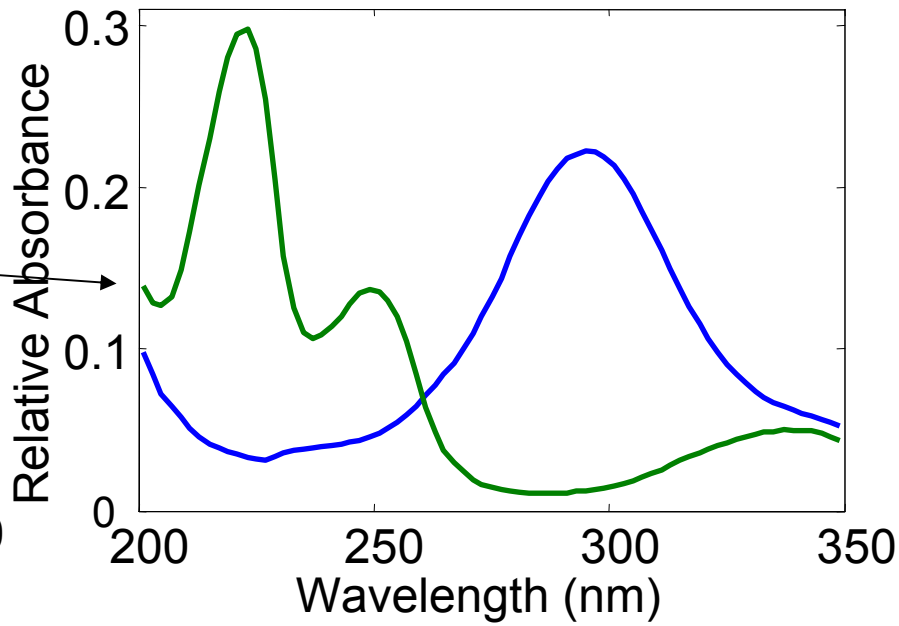




MKL Selectivity – 2D-LC vs. 2D-LC-DAD

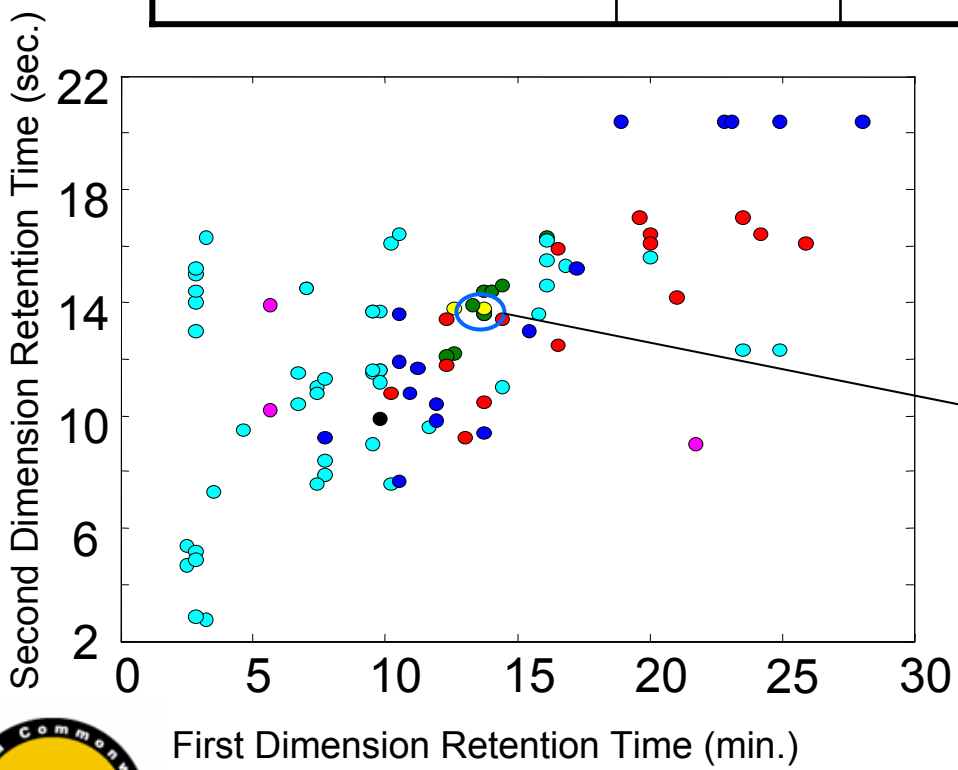
	Peak 1	Peak 2
2D-LC SEL	0.33	0.33
2D-LC-DAD SEL	0.89	0.91

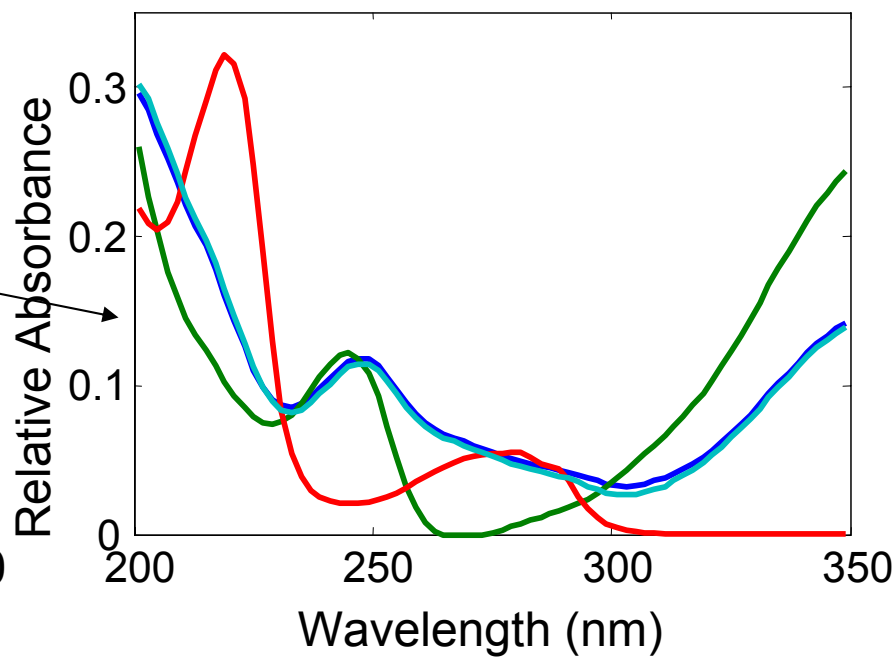




MKL Selectivity – 2D-LC vs. 2D-LC-DAD

	Peak 1	Peak 2	Peak 3	IAA-alanine
2D-LC SEL	0.33	0.33	0.98	0.16
2D-LC-DAD SEL	0.89	0.91	0.98	0.48





MKL Selectivity Comparisons

Data Dimensions	Average selectivity per component	Peak Capacity
Column 1, single wavelength	0.15	50
Column 1 + DAD	0.36	n.a.
Column 2, single wavelength	0.05	17.4
Column 2 + DAD	0.25	n.a.
2D-LC, single wavelength	0.78	870
2D-LC-DAD	0.84	n.a.

Conclusions

1. Three chemometric methods (WTTFA, PARAFAC-ALS, and fALS) have been applied to four-way quadrilinear data generated by running multiple samples with 2D-LC-DAD.
2. These methods result in a great enhancement in S/N and background suppression.
3. Several indole conjugates have been identified in mutant and wild type maize samples.
4. The indole content of the wild type and mutants are clearly differentiated.
5. The quantitative capabilities of multi-way modeling have been demonstrated.
6. Multivariate selectivity has been shown to relate to chromatographic figures of merit.

Acknowledgements

Prof. Sarah Rutan's group at Virginia Commonwealth University Department of Chemistry

Prof. Pete Carr's group at the University of Minnesota Department of Chemistry

Prof. Jerry Cohen's group at the University of Minnesota Department of Horticulture

Ms. Vibeke Svensson, Chemometrics Group, Dept. of Food Science, The Royal Veterinary and Agricultural University (Denmark)

Supelco (Discovery HS-F5 column)

ZirChrom Separations (zirconia column)

National Institutes of Health (Grant # 5R01GM054585-09) (Carr)

National Institute of Justice (Carr)

Research Corporation (Rutan)