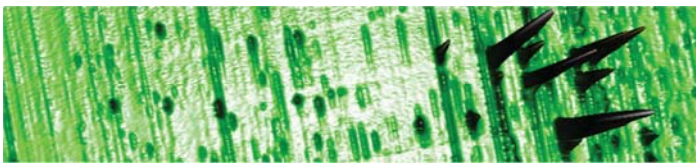
Biophysical Rationale and Quantitative Benefits of using Linear Mixed Effect Models to summarize Transitions of Peptides to Protein abundances in SRM

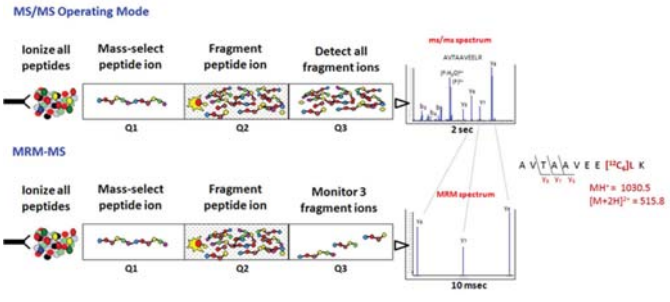


What is SRM / Field of application

- SRM
  - stands for Selective Reaction Monitoring.
  - is a mass spectrometry based method for quantitative measurement of target proteins.
- MRM™ is a trademark of AB SCIEX, but has the same principle.
- Fields of application in pharmaceutical research:
  - Pharmacodynamic: quantitation of proteins if the there is no antibody available, but target protein quite abundant (e.g. study of apoliproteins from rabbit plasma in dyslipidemic models).
  - Pharmacokinetic of therapeutic peptides.

Content

How does it work?



by Broad Institute, MA

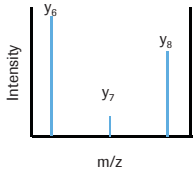
PROBLEM IN CURRENT APPLICATIONS

Biophysical Rationale

How are protein abundances determined now and what is the problem?

- It is still common to relate the relative abundance of a protein by taking the
  - Sum
  - Mean, or
  - Median

  of the peak Areas of transitions.
- Problem arises if 1-2 transition are not determined because it is
  - An outlier (e.g. by contamination of signal)
  - Below limit of Quantitation.



Transitions-Peptide-Protein Model I
Assumptions

- After trypsination, a peptide's concentration should be function of the originating proteins concentration:

  $c(\text{Peptide}) = c(\text{Protein}) * f_{\text{Tryps, p}} * \varepsilon_1 \qquad f_{\text{Tryps, p:}} \in [0,1]$
- The number of ionized peptides after Electron Spray Ionization is a function of the peptides concentration:

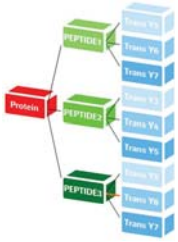
  $n(\text{Peptide}) = c(\text{Peptide}) * f_{\text{Ion, p}} * \varepsilon_2$
- The AUC of transitions over MS/MS and fragmentation is a function of the number of ES – ionized peptides passing the mass selection in MS and the fragmentation factor f_{frag}.

  $A(\text{Transition}) = n(\text{Peptide}) * f_{\text{frag, p}} * \varepsilon_3$
- Combining the three terms forms a relation between A and protein abundance:

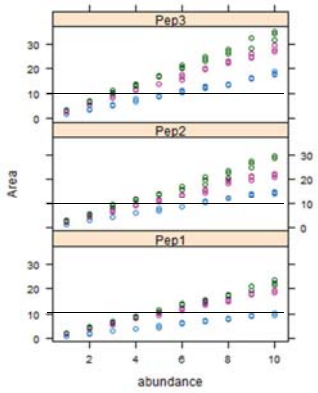
  $A(\text{Transition}) = c(\text{Protein}) * f_{\text{Tryps, p}} * f_{\text{Ion, p}} * f_{\text{frag}} \varepsilon_*$

Transitions-Peptide-Protein Model II
Assumptions

- After log-transformation, factors become addends:
  $\log(A(\text{Transition})) = \log(c(\text{Protein})) + \log(f_{Tryps,\,p} * f_{Ion,\,p}) + \log(f_{frag}) + \log(\varepsilon_*)$
  $\log(A(\text{Transition})) = \text{RelAbundanceProtein} + \text{PeptideEffect} + \text{FragEffect} + \varepsilon^*$
- Assumptions:
  - PeptideEffect ~ $N(0, \sigma_p)$
  - FragEffect ~ $N(0, \sigma_{f(p)})$
  - $\varepsilon \sim N(0, \sigma_e)$
- Example of a hierarchical, mixed effect model where a transition is nested in a peptide which is nested in protein.
- log(A(Transition)) ~ 0 + SampleID + (1| PeptideID/TransitionID)



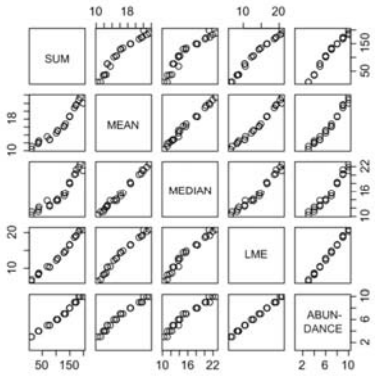
Data Simulation: Study Design



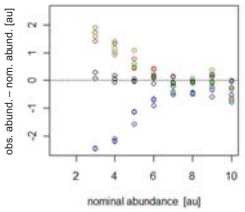
- **Data Simulation:**
  - True, nominal abundance from 1 to 10 au; protein with 3 peptides (boxes), each with 3 transitions (colors).
  - Transition Areas are computed from imaginary ionization and fragmentation factors multiplied with nominal abundance.
  - 'Below Limit of Quantitation' is simulated by setting all areas below 10 to NA.

QUANTITATIVE BENEFIT-DATA SIMULATION

Data Simulation: Mutual Relations



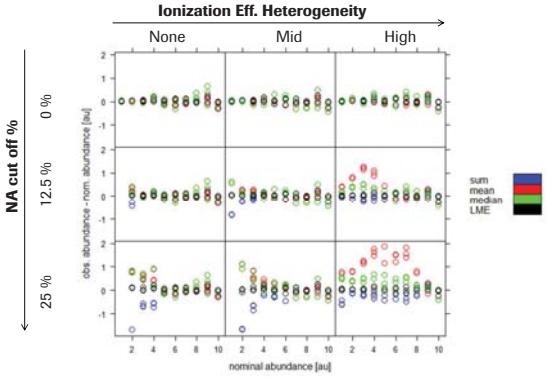
Data Simulation: Bias Plot



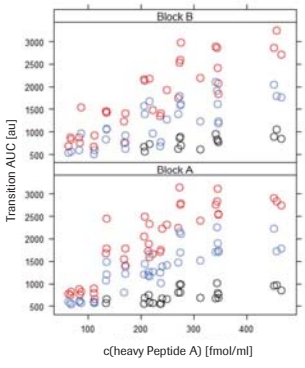
- Bias Plot (observed-nominal) ~ nominal
- At low nominal abundances (where most NA occur,)
  - Sum of transitions deviates strongly from 0 residual (> 100 % off in relative terms), is biased towards reporting lower observed abundances than true nominal abundance
  - mean, or median of transitions ALSO deviate from 0 with a positive bias.
  - **LME** (black) shows no bias and has smallest residual.

QUANTITATIVE BENEFIT – REAL DATA

Influence of NA cut off % and Ionization Efficiency Heterogeneity on Bias

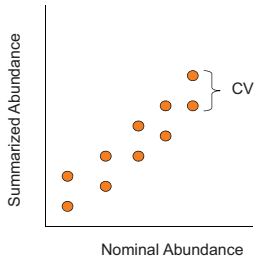


Real Data: Study Design



- Approach works best in poor data
- Spike-in of defined concentrations of "heavy-labeled" peptide A in human plasma. No interference with endogenous "light" peptide A.
- Each sample measured with three transitions (**o**, **o**, **o**) in two blocks (A & B).
- Transition **o** has low ionization efficiency and is not available in all samples.

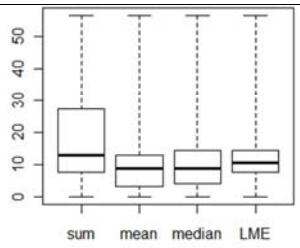
Example Batch to Batch repeatability



Real Data: Quantitative Characteristics from Spike-in

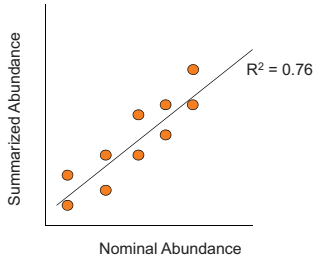
Accuracy: R^2 0.71 0.69 0.65 0.73*

Repeatability: CV



* R^2 increased to 0.77 with "Block" as random factor.

Example Accuracy: R^2



Summary & Conclusions I
How to infer protein abundance from SRM transitions

- SRM for quantitation of proteins and peptides:
  - Very useful if the there is no antibody available, but target protein quite abundant.
- Biophysical Rationale:
  - Hierarchical, mixed effect model explains best that a transition is nested in a peptide which is nested in protein.
- Data Simulation: At low nominal abundances (where most NA occur)
  - Sum of transitions deviates strongly from 0 residual, is biased towards reporting lower observed abundances than true nominal abundance.
  - mean, or median of transitions ALSO deviate from 0 report a positive bias
  - **LME** (black) shows no bias and has smallest residual.
  - The higher the percentage of NA, the more important it is to use LME.

Summary & Conclusions II

- Real Data Example:
  - Accuracy: LME is better than sum, mean or median, especially if experimental conditions (e.g. blocking factors) are considered.
  - Repeatability: LME provides more repeatability compared to sum of transitions, but is not better than mean or median

The Contributors

- Anton Belousov
- Paul Cutler
- Guillemette Duchateau-Nguyen
- Gonzalo Durán-Pacheco
- Heinz Döbeli
- Arno Friedlein
- Jens Lamerz

... And beyond.

Protein Significance Analysis in Selected Reaction Monitoring (SRM) Measurements*

Ching-Yun Chang‡, Paola Picotti§, Ruth Hüttenhain§‡‡, Viola Heinzelmann-Schwarz¶‖, Marko Jovanovic**, Ruedi Aebersold§‡‡§§, and Olga Vitek‡¶‖

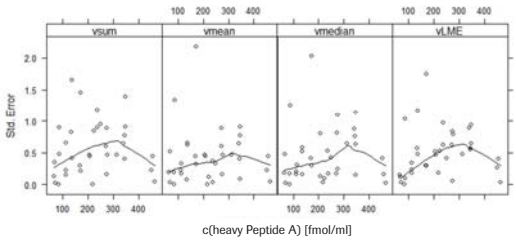
Molecular & Cellular Proteomics 11.4 10.1074/mcp.M111.014662-1

- Using LME has more advantages:
  - In a variety of experimental designs, LMEM combined with SRM have increased power compared to naïve t-tests.
  - Extensions can be made to specify heterogeneity of biological and technical variance components across features or conditions.

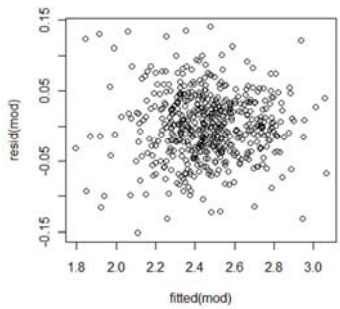
Roche

We Innovate Healthcare

Standard Error vs. Spiked in Concentration



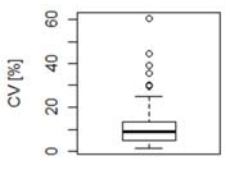
Model Diagnostics
Good characteristics of unexplained variance



Why?

Biological & Technical Feasibility
Free Peptide A in Plasma

- Result II: Repeatability of technical replicates is very good (median CV 9 %)



	CV [%]
Min	1.2
Q1	5.1
Median	8.9
Q3	13.4
Max	60.8

Acknowledgements

Pharm Res
DOI 10.1007/s11095-012-0707-7

RESEARCH PAPER

Mass Spectrometry Based Quantification of CYP Enzymes to Establish *In Vitro-In Vivo* Scaling Factors for Intestinal and Hepatic Metabolism in Beagle Dog

Ali T. Heikkinen • Arno Friedlein • Jens Lamerz • Peter Jakob • Paul Cutler • Stephen Fowler • Tara Williamson • Roberto Tolando • Thierry Lave • Neil Parrott

A. T. Heikkinen • S. Fowler • T. Lave • N. Parrott (✉)
F. Hoffmann-La Roche Ltd., Pharmaceuticals Division
Non-Clinical Safety
Grenzacherstrasse 124, B70/R130
CH-4070 Basel, Switzerland
e-mail: neil_john.parrott@roche.com

A. Friedlein • J. Lamerz • P. Jakob • P. Cutler
Translational Research Sciences, Pharmaceuticals
Division F. Hoffmann-La Roche Ltd.
Basel, Switzerland

T. Williamson
Celsis In vitro Technologies
Baltimore, Maryland, USA

R. Tolando
Celsis In vitro Technologies GmbH
Neuss, Germany