CASSS HOS 2017

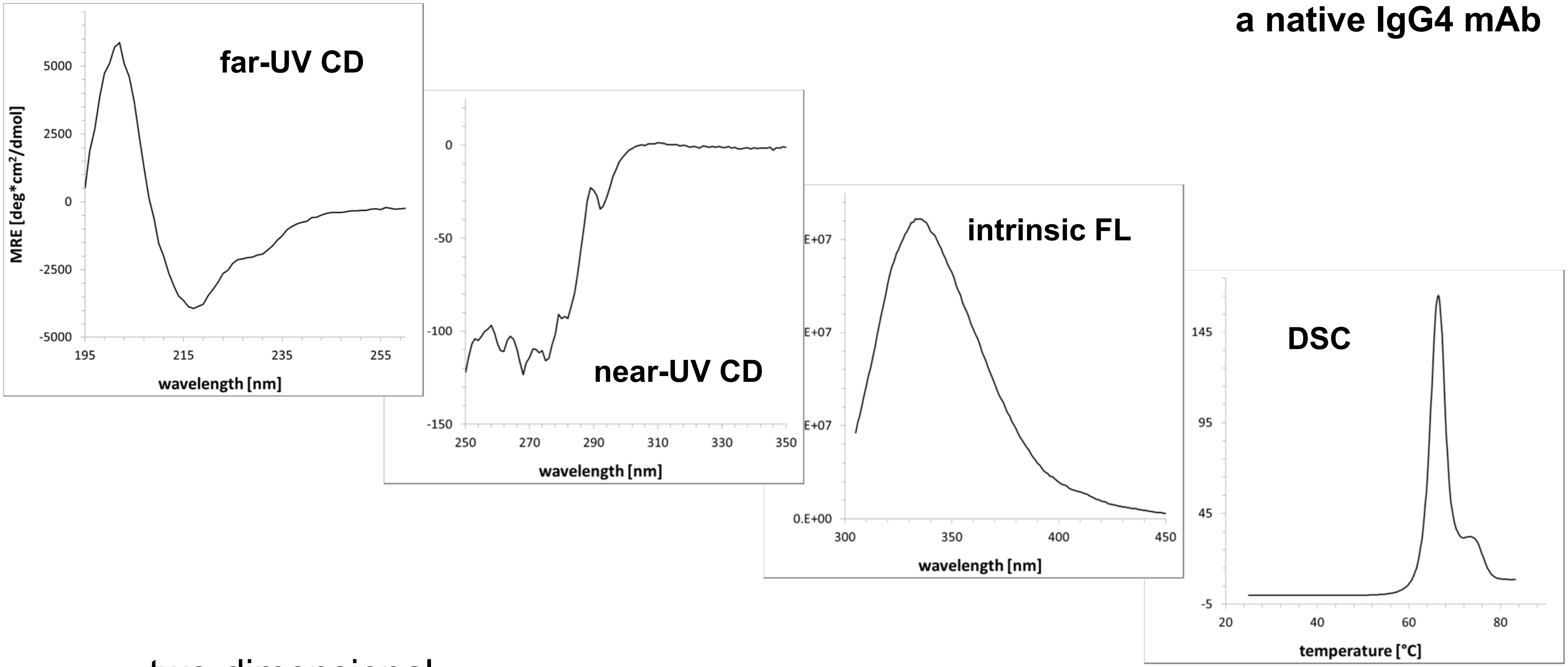
Variability estimates and data comparison with higher order structure data

Ivan L. Budyak

Biopharmaceutical Research and Development
Eli Lilly and Company

April 03-05, 2017

HOS characterization data

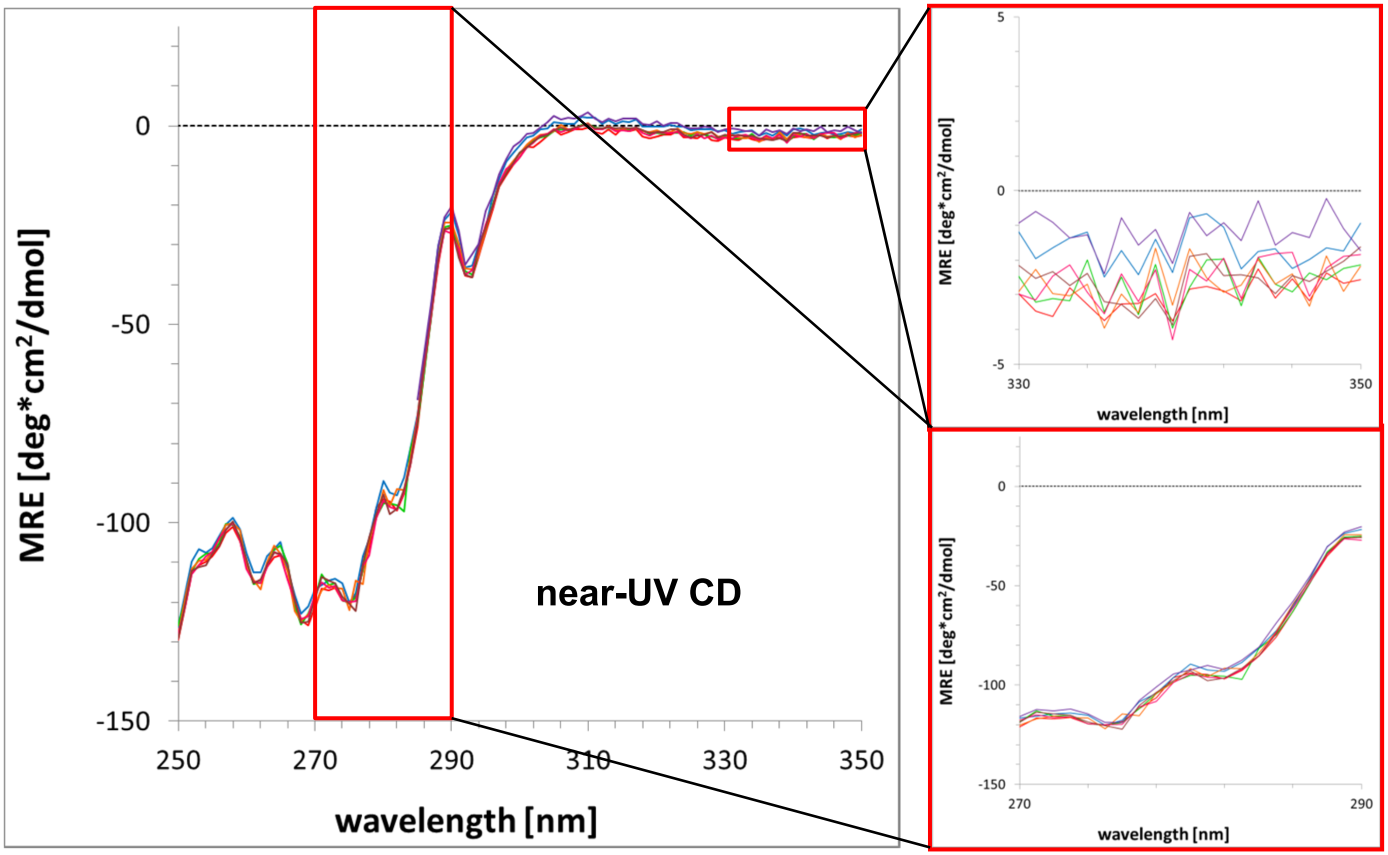


a native IgG4 mAb

- two-dimensional
- features range from a single broad maximum (FL) to many bands (vibrational spectra)
- contain regions of different signal:noise
- multi-step manipulation (e.g. blank subtraction, normalization, baseline correction, derivative calculation)

A closer look: near-UV CD

near-UV CD spectra of a native IgG4 mAb



different? same?

how much?

how significantly?

what about groups?

There are specific challenges in establishing a measure of 'similarity' for data sets in 2D+

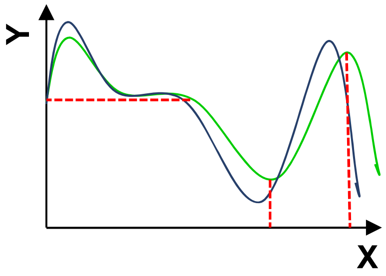
Approaches to data comparison

Potential considerations

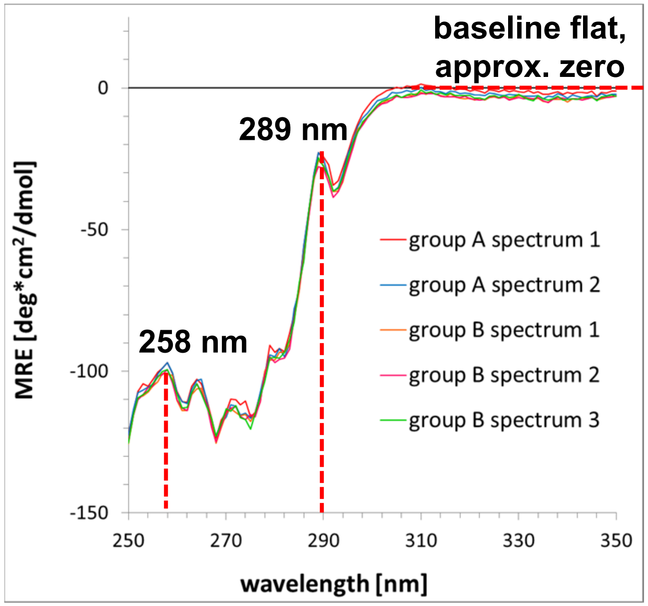
- Applicable to one-, two-, and multidimensional data?
- Can take into account method variability?
- Applicable to any number of groups of data consisting of any number of individual data?
- Robust against variations in resolution, signal intensity, and noise levels?
- Easy to interpret and implement?

What is in the toolbox?

Spectral overlay



data comparison

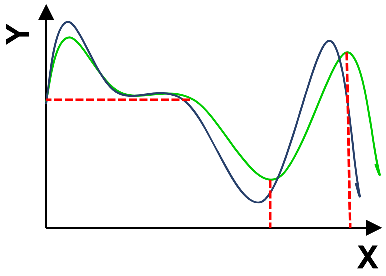


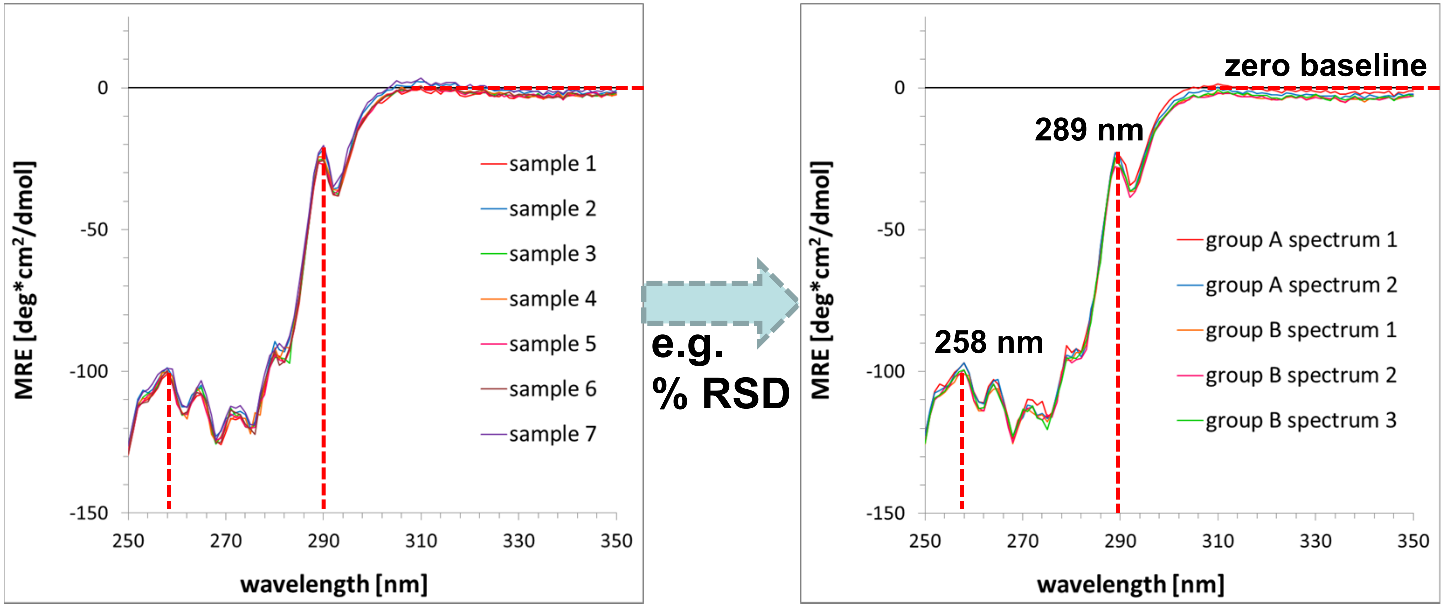
data comparison

spectral feature	mean ± SD, deg*cm^2/dmol	% RSD
289 nm band	−24.7 ± 2.1	8.4
258 nm band	−98.7 ± 1.8	1.8
baseline	N/A	N/A

- Intuitive and easy to implement
- In case the spectral feature is assigned – May link to known structural properties
- Partial spectral coverage for feature-rich spectra – What about other bands?
- Potential problems defining spectral features – When a shoulder becomes a band? What is a sufficiently flat and sufficiently zeroed baseline?
- May not be applicable to smaller data sets / groups – SD inflates at small N

Spectral overlay





variability estimate

data comparison

e.g. % RSD

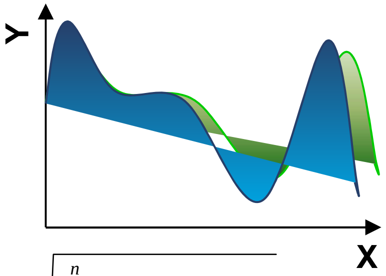
The small N limitation may be addressed by applying historical method performance data or generated variability estimates using representative samples

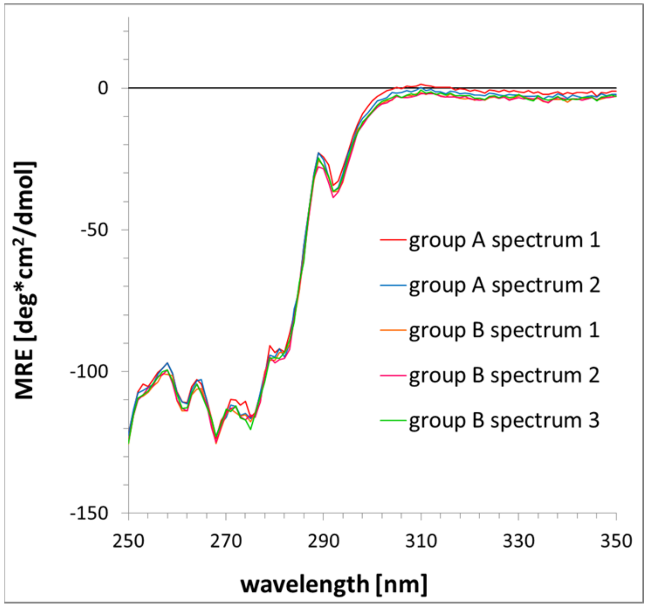
Spectral reduction

CASSS HOS 2017

d'Antonio J. *et al*. J.Pharm.Sci. (2012)
Teska B.M. *et al*. Anal.Biochem. (2013)
Dinh N.N. *et al*. Anal.Biochem. (2014)

- Correlation coefficient
- Derivative correlation coefficient
- (Modified) Area of overlap
- (Weighted) Spectral difference





$$r = \sqrt{\frac{\left(\sum_{i=1}^{n} y_{1i} y_{2i}\right)^2}{\left(\sum_{i=1}^{n} y_{1i}^2\right)\left(\sum_{i=1}^{n} y_{2i}^2\right)}}$$

correlation coefficient

	A1	**A2**	**B1**	**B2**	**B3**
A1	1.0000	0.9997	0.9994	0.9993	0.9995
A2		1.0000	0.9998	0.9997	0.9997
B1			1.0000	0.9999	0.9999
B2				1.0000	0.9998
B3					1.0000

$$wSD = \sqrt{\frac{\sum_{i=1}^{n} \left|y_{1i}\right| \left(y_{1i} - y_{2i}\right)^2}{\sum_{i=1}^{n} \left|y_{1i}\right|}}$$

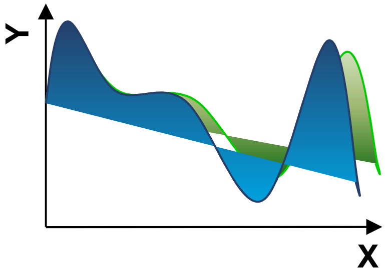
weighted spectral difference

	A1	**A2**	**B1**	**B2**	**B3**
A1	0.0000	1.6800	2.8896	2.9564	2.7146
A2		0.0000	2.0915	2.0852	2.1782
B1			0.0000	1.2978	1.5448
B2				0.0000	1.5817
B3					0.0000

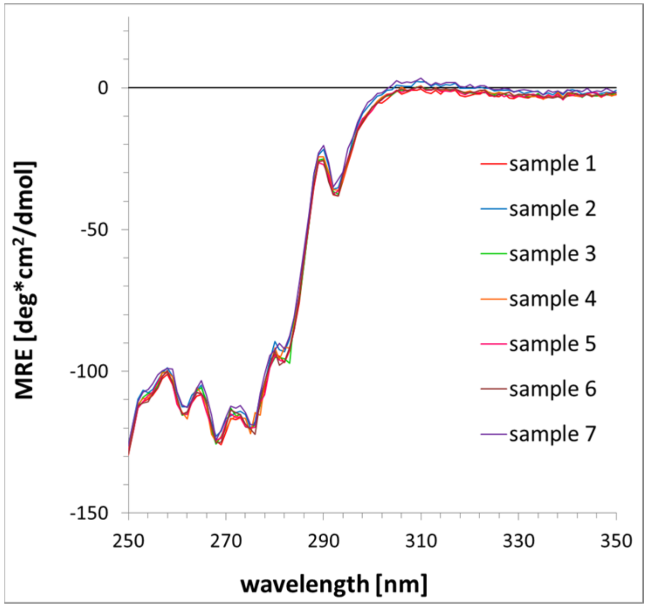
- Sensitive, yield a single number (data reduction)
- Mostly limited to pairwise comparison – What about groups?

Spectral reduction

d’Antonio J. *et al*. J.Pharm.Sci. (2012)
Teska B.M. *et al*. Anal.Biochem. (2013)
Dinh N.N. *et al*. Anal.Biochem. (2014)



$$r = \sqrt{\frac{\left(\sum_{i=1}^{n} y_{1i} y_{2i}\right)^2}{\left(\sum_{i=1}^{n} y_{1i}^2\right)\left(\sum_{i=1}^{n} y_{2i}^2\right)}}$$



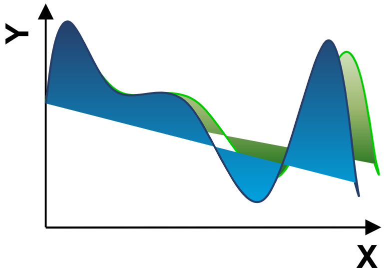
correlation coefficient

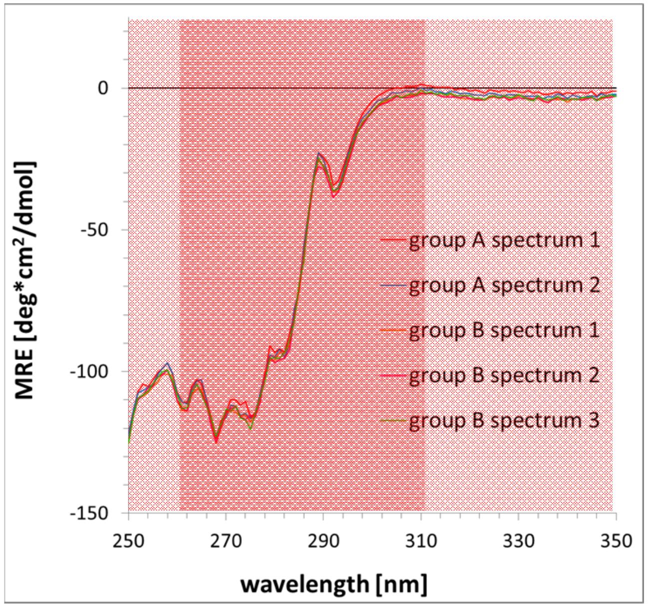
	1	2	3	4	5	6	7
1	1.0000	0.9996	0.9998	0.9998	0.9999	0.9999	0.9993
2		1.0000	0.9997	0.9997	0.9997	0.9997	0.9998
3			1.0000	0.9998	0.9999	0.9998	0.9994
4				1.0000	0.9998	0.9997	0.9995
5					1.0000	0.9999	0.9994
6						1.0000	0.9995
7							1.0000

- Sensitive, yield a single number (data reduction)
- Mostly limited to pairwise comparison – What about groups? How to include method variability?

Spectral reduction

d'Antonio J. *et al*. J.Pharm.Sci. (2012)
Teska B.M. *et al*. Anal.Biochem. (2013)
Dinh N.N. *et al*. Anal.Biochem. (2014)





correlation coefficient

260-350 nm

	A1	A2	B1	B2	B3
A1	1.0000	0.9997	0.9992	0.9990	0.9993
A2		1.0000	0.9997	0.9996	0.9997
B1			1.0000	0.9999	0.9998
B2				1.0000	0.9998
B3					1.0000

250-310 nm

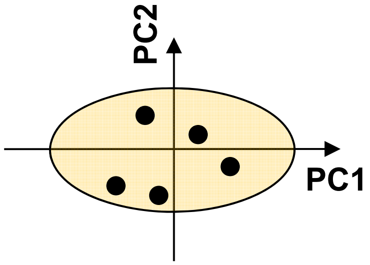
	A1	A2	B1	B2	B3
A1	1.0000	0.9998	0.9996	0.9995	0.9997
A2		1.0000	0.9998	0.9998	0.9998
B1			1.0000	0.9999	0.9999
B2				1.0000	0.9999
B3					1.0000

- Sensitive, yield a single number (data reduction)
- Mostly limited to pairwise comparison – What about groups? How to include method variability?
- The degree of (dis)similarity may depend on the magnitude, resolution, signal:noise

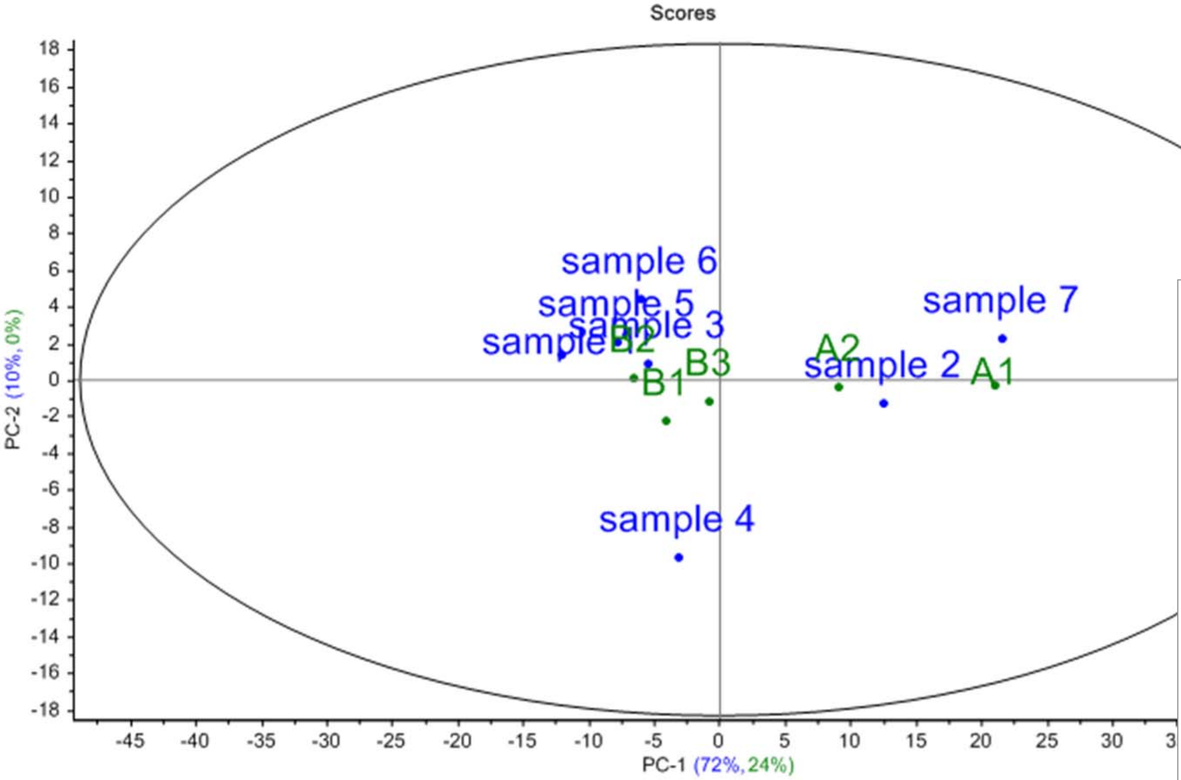
Spectral reduction – PCA

CASSS HOS 2017

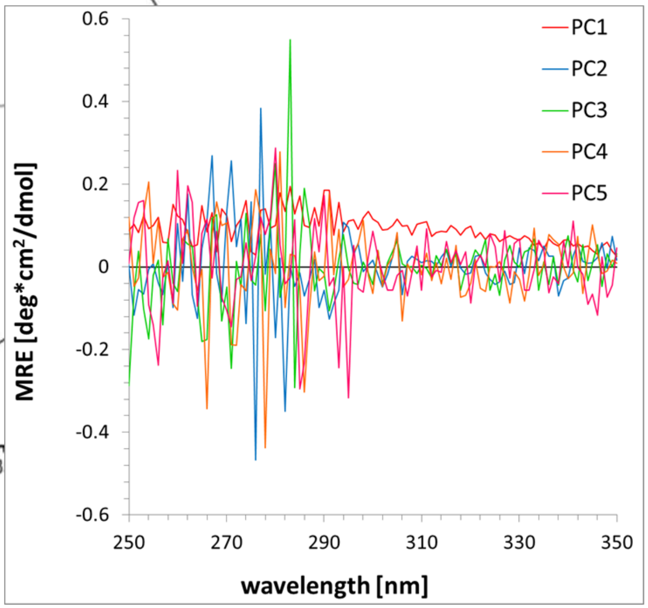
Stockdale G. *et al*. J.Pharm.Sci. (2014)
Rogstad S. *et al*. Anal.Bioanal.Chem. (2015)



principal components analysis



component loadings

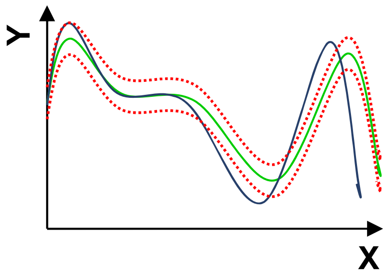


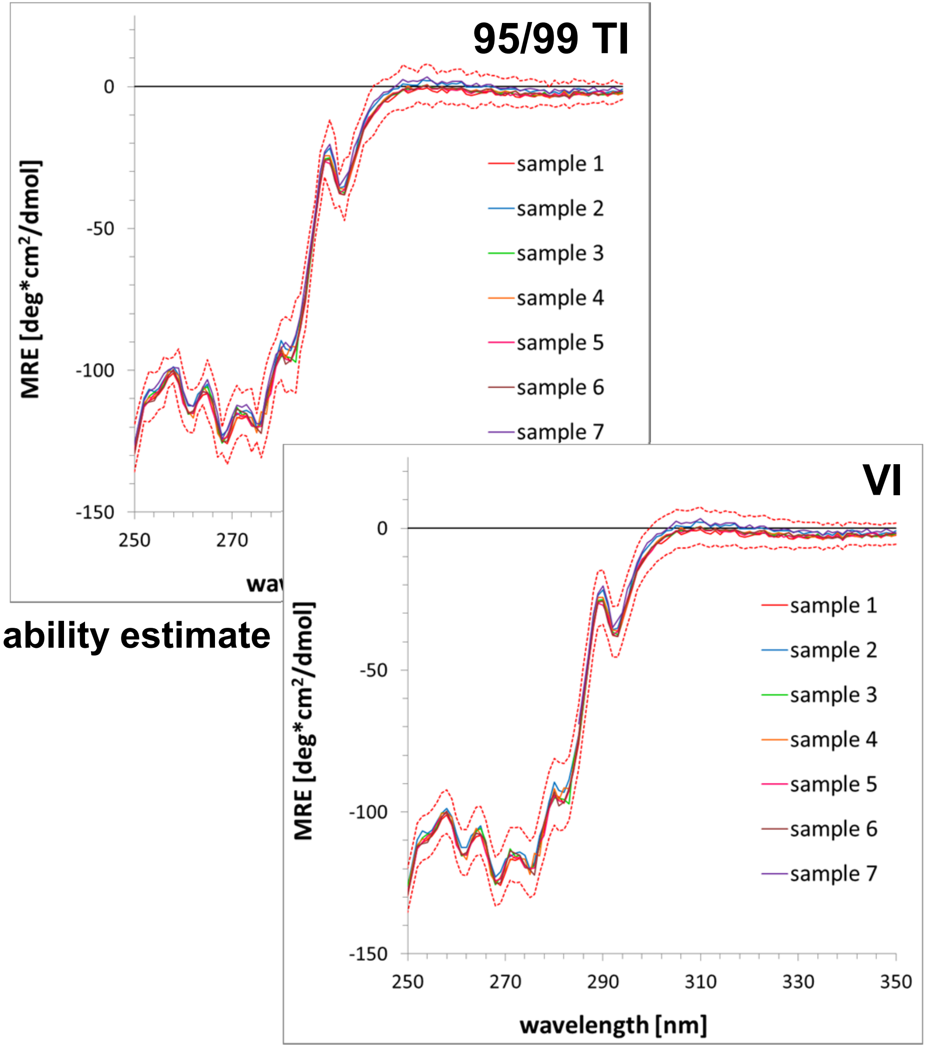
- Universal and powerful, applicable to multi-dimensional data, and groups of data
- Potentially complex interpretation, sensitivity to noise

CASSS HOS 2017

Variability-based intervals

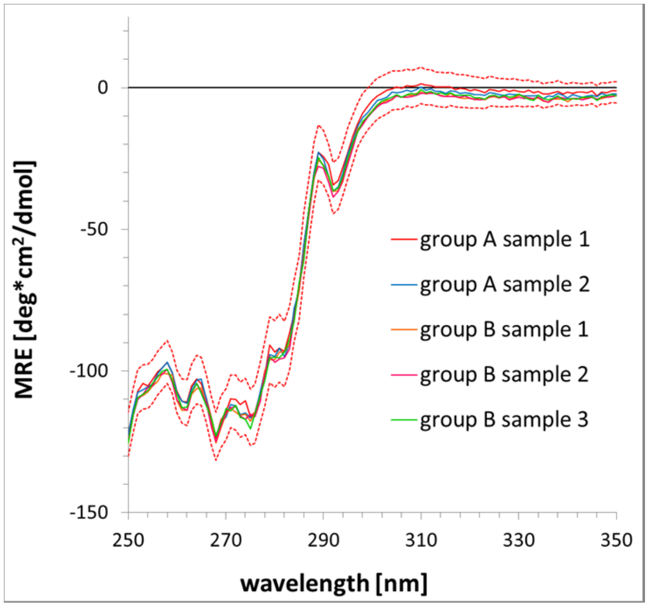
Lin J.C. *et al*. J.Pharm.Sci. (2015)
Budyak I.L. *et al*. Anal.Biochem. (2016)





variability estimate

data comparison

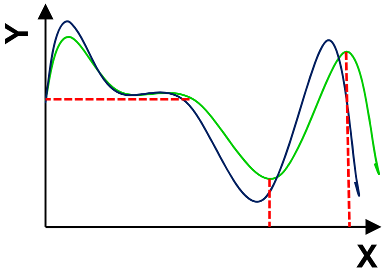


- Intuitive, variability-based, spectral region-independent, can be applied to groups of any size
- Require good control in X dimension

The HOS comparison toolbox

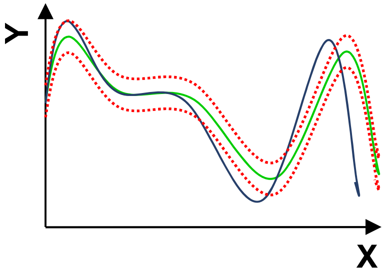
Spectral overlay

- Intuitive, easy to implement, applicable to groups
- Can link selected features to known structural properties
- Partial spectral coverage for feature-rich spectra
- Potential problems defining spectral features



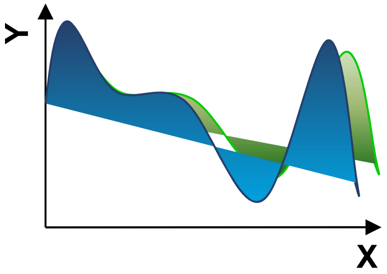
Variability-based intervals

- Intuitive, easy to implement
- Spectral region-independent, applicable to groups
- Require good control in X dimension



Spectral reduction

- Holistic, sensitive, yield a single number (data reduction)
- Many are limited to pairwise comparison
- Complex dependence on the magnitude, resolution, signal:noise
- Potentially complex interpretation



Increasing complexity: NMR

CASSS HOS 2017

Quinternet M. *et al*. J.Pharm.Biomed.Anal. (2013)
Amezcua C.A., Szabe C.M. J.Pharm.Sci. (2013)
Poppe L. *et al*. Anal.Chem. (2013)
Japelj B. *et al*. Sci.Rep. (2016)

Japelj, B. *et al.*
Biosimilar structural comparability assessment by NMR: from small proteins to monoclonal antibodies.
Sci. Rep. **6**, 32201; doi: 10.1038/srep32201 (2016).

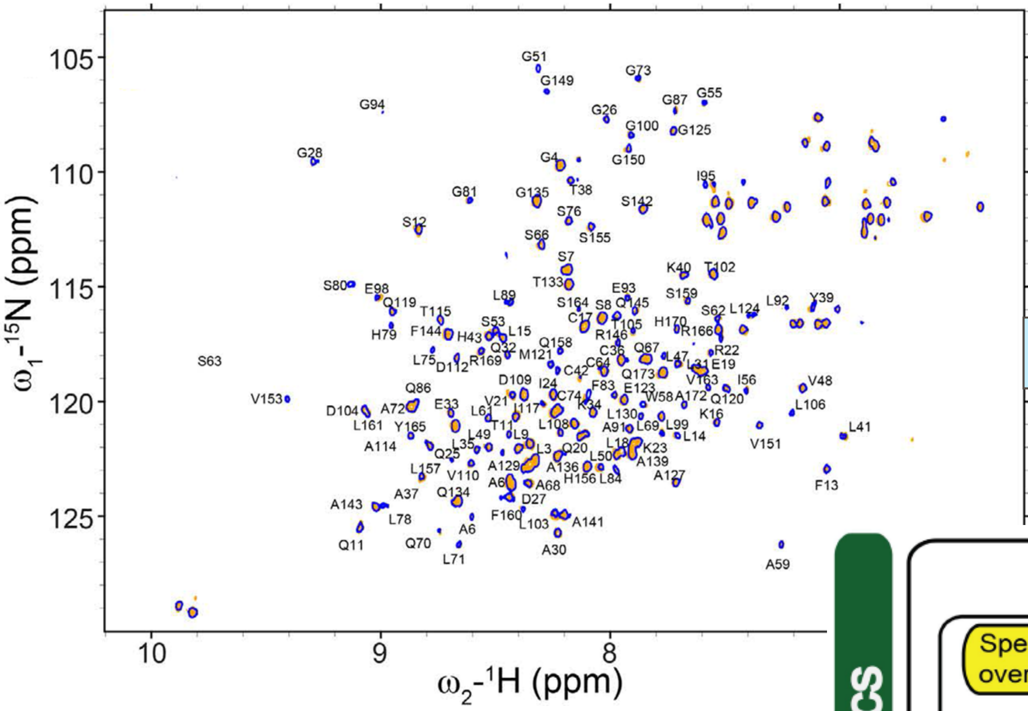
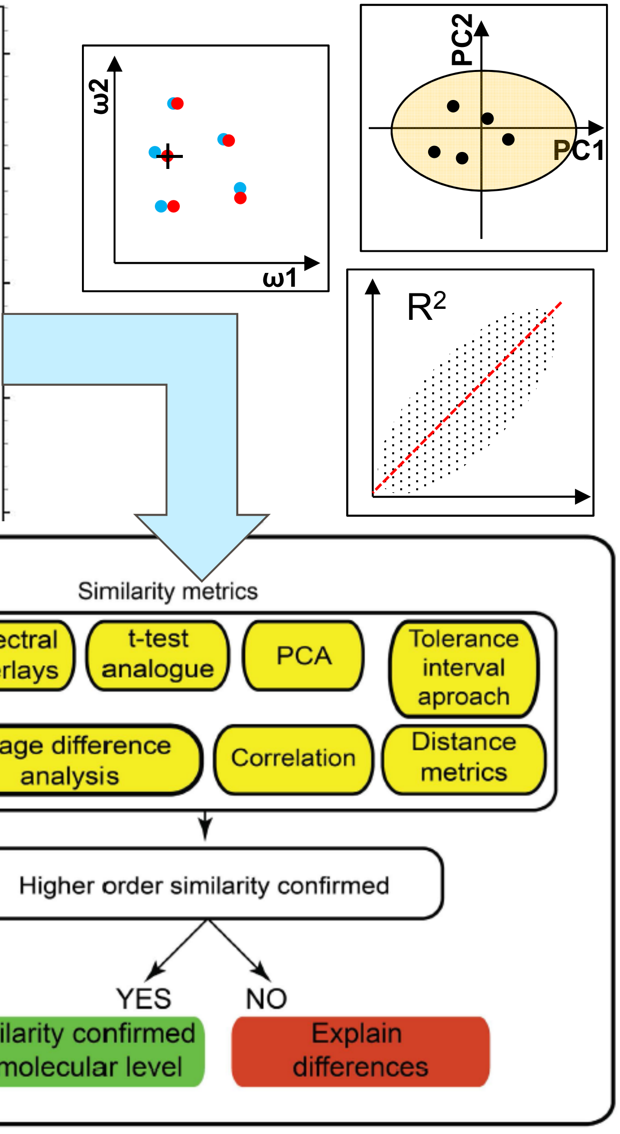


Figure 2a. Overlay of the amide fingerprint spectra of filgrastim reference product and biosimilar filgrastim product.

- The choice of similarity metrics may depend on multiple factors such as the type of spectrum, signal:noise, signal overlap *etc*.



Bioinformatics

Similarity metrics

Spectral overlays
t-test analogue
PCA
Tolerance interval aproach
Image difference analysis
Correlation
Distance metrics

Higher order similarity confirmed

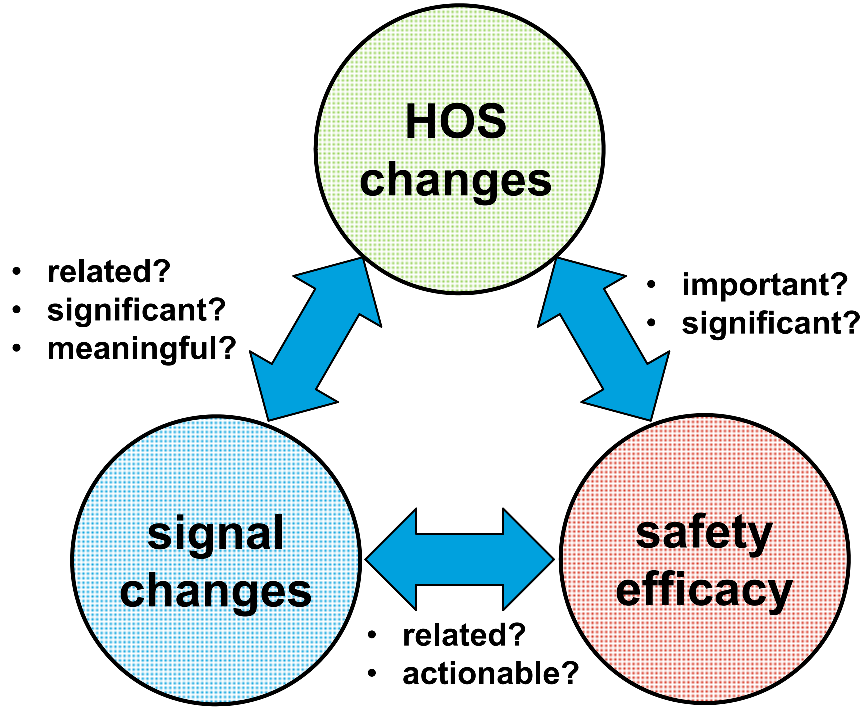
YES
NO

Similarity confirmed at molecular level
Explain differences

Figure 7 (adapted). NMR fingerprint-bioinformatics workflow.

Conclusions

- <u>Robust methods to estimate variability and compare data are critical</u> for the use of HOS data in technical decision making, particularly for complex, multi-dimensional data (e.g. NMR and LC-MS)
- Advantages and limitations exist for every approach in the HOS data comparison toolbox – <u>chose the one that fits the purpose</u>
- <u>Control strategies will continue to evolve</u> based on further elucidation of the links between the HOS changes, measured properties, and clinical data



Acknowledgements

Eli Lilly and Company

- Zhenqi ‘Pete’ Shi
- William F. Weiss IV
- Kristi L. Griffiths
- Michael R. De Felippis
- Bryan J. Harmon
- Jeffrey D. Hofer