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

- 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

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

different? same?
how much?
how significantly?
what about groups?

near-UV CD spectra of a native IgG4 mAb
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

- 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

### Data Comparison

<table>
<thead>
<tr>
<th>spectral feature</th>
<th>mean ± SD, deg*cm²/dmol</th>
<th>% RSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>289 nm band</td>
<td>$-24.7 \pm 2.1$</td>
<td>8.4</td>
</tr>
<tr>
<td>258 nm band</td>
<td>$-98.7 \pm 1.8$</td>
<td>1.8</td>
</tr>
<tr>
<td>baseline</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
The small N limitation may be addressed by applying historical method performance data or generated variability estimates using representative samples.
Spectral reduction

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

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

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

\[
wSD = \frac{\sum_{i=1}^{n}|y_{1i}|(y_{1i} - y_{2i})^2}{\sum_{i=1}^{n}|y_{1i}|}
\]

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.0000</td>
<td>1.6800</td>
<td>2.8896</td>
<td>2.9564</td>
<td>2.7146</td>
</tr>
<tr>
<td>A2</td>
<td>1.0000</td>
<td>2.0915</td>
<td>2.0852</td>
<td>2.1782</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>0.0000</td>
<td>1.2978</td>
<td>1.5448</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>0.0000</td>
<td>1.5817</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Spectral reduction

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

![Graph showing MRE vs. wavelength for different samples.]

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

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>0.99996</td>
<td>0.9998</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9993</td>
</tr>
<tr>
<td>2</td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9998</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.0000</td>
<td>0.9998</td>
<td>0.9999</td>
<td>0.9998</td>
<td>0.9994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.0000</td>
<td>0.9998</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.9995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.0000</td>
<td>0.9995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Spectral reduction

- 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

### Correlation Coefficient

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A1</strong></td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9992</td>
<td>0.9996</td>
<td>0.9993</td>
</tr>
<tr>
<td><strong>A2</strong></td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9996</td>
<td>0.9997</td>
<td></td>
</tr>
<tr>
<td><strong>B1</strong></td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9999</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B2</strong></td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9998</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B3</strong></td>
<td>1.0000</td>
<td>0.9998</td>
<td>0.9999</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 260-350 nm

#### 250-310 nm
Spectral reduction – PCA

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

- 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

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.
Conclusions

- Robust methods to estimate variability and compare data are critical 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 – chose the one that fits the purpose
- Control strategies will continue to evolve 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