Automated Evaluation of Chiral Screenings

PREP UK 2018
Ernst Freund, Daniel Meyer, Harald Schroeder, Trixie Wagner
Novartis Institute for BioMedical Research / AXS Separations

29/30 November 2018
Agenda

• Introduction
  • Chiral Screening Strategy and Chiral Database
  • Chiral Database Applications
• Goals and Challenges Automated Chromatogram Evaluation
• Implementation on Waters UPC2 Systems
• Summary and Outlook
Chiral Separations 2014 – 2018

Chiral requests 2014-2018

- 2014: 695
- 2015: 874
- 2016: 1039
- 2017: 1039
- 2018: 800

Turnaround times chiral requests 2014-2018

- 2014: 6
- 2015: 2
- 2016: 3
- 2017: 1
- 2018: 4

Prep separation scale distribution 2018

- >10 g: 68%
- 2-10 g: 19%
- 1-2 g: 0%
- 500 mg - 1 g: 4%
- 100-500 mg: 4%
- <100 mg: 2%

Prep separations turnaround time 2018

- 1-2 d: 38%
- 3-5 d: 37%
- 6-10 d: 5%
- 11-20 d: 0%
- >20 d: 0%

*including outsourced samples
Chiral Screening Strategy SFC First

All submitted samples → SFC Screening (std gradient) → LC Screening → Prep. SFC → Fraction QC UPC2

Prep. LC → Fraction QC Fast-LC

Discussion with chemist: Further automated screenings Outsourcing «manual» MD

5%

UPC2 - MS
10 conditions in series & Sepiatec (if needed) 24 conditions in parallel

SFC Chiral Database

define optimal screenings

90% 5% 5%
Manual Chiral Database

Low Resolution
No hit
(0)

Moderate Resolution
Partial hit
(1)

High Resolution
Full hit
(2)

Screen Nb. | AWM-nb. | Project code | Column | MeOH+NH3 | IPA+NH3 |
---|---|---|---|---|---|
1 | BS1700417979 | | Chiralpak AD | 0 | 2 |
| Chiralpak IC | 1 | 1 |
| Chiralpak ID | 1 | 2 |
| Chiralpak IG | 0 | 0 |
| Chiralpak IBN | 1 | 1 |
2 | BS17004174061 | | Chiralpak AD | 1 | 0 |
| Chiralpak IC | 0 | 0 |
| Chiralpak ID | 0 | 0 |
| Chiralpak IG | 0 | 0 |
| Chiralpak IBN | 0 | 0 |
3 | BS17004183541 | | Chiralpak AD | 2 | 0 |
| Chiralpak IC | 0 | 0 |
| Chiralpak ID | 0 | 1 |
| Chiralpak IG | 2 | 1 |
| Chiralpak IBN | 1 | 0 |
4 | BS17004183893 | | Chiralpak AD | 2 | 2 |
| Chiralpak IC | 1 | 2 |
| Chiralpak ID | 1 | 1 |
| Chiralpak IG | 2 | 0 |
| Chiralpak IBN | 0 | 0 |
5 | BS17004180441 | | Chiralpak AD | 2 | 1 |
| Chiralpak IC | 1 | 1 |
| Chiralpak ID | 1 | 2 |
| Chiralpak IG | 2 | 2 |
| Chiralpak IBN | 1 | 0 |

Why
36.9% + 29.2% = 48.9%
is correct.

Overlapping selectivity!
Optimization of Chiral Screening

Initial screening: evaluation of 442 screenings

- **Full hits per column**
  - AD: 30%
  - IC: 18%
  - ID: 17%
  - IG: 24%
  - IBN: 11%

- **Unique selectivities of columns (orthogonality)**
  - AD: 49
  - IC: 16
  - ID: 6
  - IG: 27
  - IBN: 12

- Full hit: 76%
- Partial hit: 85%
- No separation: 8%

Replace ID by OZ

Optimized screening: evaluation of 140 screenings

- **Full hits per column**
  - AD: 31%
  - IC: 23%
  - OZ: 19%
  - IC: 12%
  - IBN: 15%

- **Unique selectivities of columns (orthogonality)**
  - AD: 13
  - IC: 7
  - OZ: 6
  - IG: 6
  - IBN: 9

- Full hit: 77%
- Partial hit: 91%
- No separation: 4%

All hit rates show positive trend, selectivity broader distributed
Goals Automated Screening Evaluation

• Data driven science / Big Data
  • Make optimal use of available data
  • Generate homogeneous datasets for chiral screenings (include historical data)
  • Create data in machine readable form

• Chromatography
  • Evaluate screening chromatograms with 0/1/2 scheme
  • Allow extraction of chromatographic parameters as selectivity, resolution, ...
  • Enable «directed» screening

• Reduction of routine work
  • Setting up screenings
  • Avoid manual entries into Excel sheet
  • Automated reporting
Automated Screening Evaluation and Directed Screening in Series

Decreasing hit rate

Increasing orthogonality

Directed by e.g.
- Previous chromatogram
- Statistics
- Other correlations

Software e.g. Virscidian scores based on:
- Selectivity
- Resolution
- Peak symmetry
- Retention times
  - Stops screening as soon as full hit is identified
  - Continues with next screening

Automated report generation, transfer key data to chiral database,
create worklist for prep run

Time gain ↔ incomplete data?
Challenges Automated Screening Evaluation

- **Evaluation without background information**
  - 4 peaks
  - Ratio of peaks
  - Impurities, different masses
  - Operator compensates: e.g. perform full screening

- **Chromatographic problems**
  - No / late elution
  - Elution in next run
  - Wrong mass or mass not found?

Most of the obtained chromatograms are «normal»
Robust and conservative decision algorithm to avoid false positives
Implementation on UPC2

1. Masslynx: hardware control, raw data handling
2. Autolynx: setting up runs through external control (txt file)
3. Openlynx: automated data processing and report generation (rpt file)
4. Software for evaluation of chromatographic parameters, scoring and overall control

Data processing and export:

- Openlynx evaluation routines are used
- rpt-files are created

```
...[PEAK]

| Peak ID | 1 |
| Peak Ref | 1 |
| Time     | 4.0133 | 4.1808 |
| Peak     | 3.8991 | 1460.91 | 195251.75 |
| Intensity| 23294324.00 |
| Height   | 2201497.5000 |
| AreaAbs  | 99.56 |
| Area %Total | 49.33 |
| Width    | 0.28 |

...[PEAK]

| Peak ID | 3 |
| Peak Ref | 3 |
| Time     | 4.7451 | 5.0068 |
| Peak     | 4.5367 | -141962.16 | -307422.03 |
| Intensity| 13156310.00 |
| Height   | 2211307.5000 |
| AreaAbs  | 100.00 |
| Area %Total | 49.55 |
| Width    | 0.47 |
```

MS data for peak identification
Rpt File Analysis – Parser

Openlynx settings
- Integration parameters set to recognize maximally 10 peaks
- Other evaluation parameters adapted to give reasonable results

Software imports chromatographic data from rpt-file
- Parser allows off- and online use of data
- From rpt file imported parameters (per peak):
  - UV Peak Retention Time
  - UV Peak Start
  - UV Peak End
  - UV Peak Width
  - UV Peak Area
  - Mass confirmed (y/n)
  - MS Peak Retention Time
  - MS Peak Area

Relative areas for UV and MS are calculated
- Important: Peak width (Openlynx) > real peak width (compensated)
Evaluation Algorithm: Score 0 / 1 / 2

Initial model:
0: Rs < 1.1
1: Rs > 1.1 AND R < 1.5
2: Rs > 1.5

advanced model:
0: Rs < 1.1
1: Rs < 1.1 AND \( \alpha > 1.05 \) (e.g. early eluters)
1: Rs > 1.1 - 1.5
2: Rs > 1.5

IPA+NH3
MeOH+NH3
Distinction between 1 and 0 difficult
Evaluation Algorithm: Total Score

- Total score is calculated as a weighted sum from selectivity, resolution, retention time, peak width and peak asymmetry (empirical)

\[
\text{Total score} = w_1(x - 1) + w_2R_s + w_3(R_t) + \frac{w_4}{\text{peak width}_{av}} + \frac{w_5}{\text{peak asymmetry}_{av}}
\]
Summary Automated Screening Program

Prototype version realised with Microsoft Studio 2017 in C# and running without issues since >6 months

- Program sets up and organizes screenings
- Documents screening ratings in log file (txt)
- Generates rpt files (machine readable)
- Can stop the screening if a sufficient separation is found and continues with the next screening = directed mode
- >500 screening data sets in machine readable form generated; approx. 1000 will be added each year

GDC / Discovery Technologies / AXS Separations
Outlook

Directed Screening Program

• Introduce «human override» for wrong scores (review by exception)
• Create real database
• Include data from all screening runs into analysis of screening?
• Adapt and integrate automated reporting module

Data Evaluation

• Establish automated link to structures resp. filter (4 peaks, diastereomers, atropisomers, absolute configurations)
• Define tools for more detailed statistical analysis of screening data
• Find correlations between structural data – separation conditions

1 Sheridan et al., J. of Chromatography A (2016), 1467, 206-213.
2 Schneider et al., ChemMedChem (2018), 13, 1315-1324.
• Karin Briner
• René Wyler
• Trixie Wagner
• John Reilly
• Thomas Wolf
• Harald Schröder
• Daniel Schmid
• Daniel Meyer
• Team Separations & Analytics Basel
Additional slides
### Implementation on UPC2

1-2) Running analyses: Autolynx with text file input

<table>
<thead>
<tr>
<th>Index</th>
<th>USER FILE</th>
<th>TEXT FILE</th>
<th>NAME</th>
<th>CONDITIONS</th>
<th>INLET_FILE</th>
<th>MS_FILE</th>
<th>SAMPLE LOCATION</th>
<th>INJ_VOL</th>
<th>WAVELENGTH_A</th>
<th>MASS_A</th>
<th>MASS_B</th>
<th>PROCESS</th>
<th>PROCESS_PARAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>schroha2</td>
<td>BS-No</td>
<td>schroha2-test</td>
<td>Chiralpak IB 4.6x100mm 5um 5-55%IPA+NH3 100x4-6</td>
<td>Chiralpak IB 5um-IPA(NH3)</td>
<td>ES+ 100-1000</td>
<td>6min</td>
<td>01:31</td>
<td>5 0 0 472.2</td>
<td>OpenLynx UPC2.olp</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3) Reporting
- Openlynx generates rpt files (text) with results

| Peak ID | 1 |
| Peak Ref | 1 |
| Time | 4.0133 |
| Peak | 3.8991 |
| Intensity | 1460.91 |
| Height | 23294324.00 |
| AreaAbs | 2201497.5000 |
| Area %BP | 99.56 |
| Area %Total | 49.33 |
| Width | 0.28 |

4) Evaluation
- Software imports chromatographic data from rpt-file
- Evaluation of the separation (MS-data for peak ID)
- Documentation of the evaluation in database
- Decision, whether screening can be stopped
- Automated reporting
- Generation of next Autolynx input file
Evaluation Algorithm

1. Largest 4 UV peaks are identified (if so many exist)
2. 4 largest peaks have the correct mass (y/n)?
3. Sum of UV-area of peaks with correct mass is calculated
4. PEAKS 1 and 2 = largest UV peaks with correct mass are identified (if they exist)
5. If <2 peaks with correct mass, evaluation of the largest two UV peaks is performed if certain requirements are fulfilled (sum of UV-area, minimum UV peak area, ratio peak 1 and 2), screening is not stopped
6. Otherwise evaluation is performed with PEAK 1 and 2 (results depend on parameters)

Step 1

Step 2 (16 cases) Step 3

Step 4

Step 5

Step 6

1. Good separation (2) screening can be stopped

2. Good separation (2) screening is not stopped «too dirty»

3. Separation insufficient (1) screening is not stopped «too dirty»

4. Good separation (2) screening is not stopped
Directed Screening and Documentation

**Decision, whether screening can be stopped**
- If operator sets program into directed mode
- if score=2
- and masses of main peaks confirmed
- and peak areas above minimal threshold
- and sum of rel UV area of main peaks > 90%
- and ratio of 2 main peaks < 80:20

**Chiral Database: Tab separated text file**
- Major peak resp PEAK 1 and PEAK 2, which are evaluated are stored
- Copy – paste into e.g. Excel for automated statistics

**screening identifiers scores and chromatographic data for screening evaluation**

<table>
<thead>
<tr>
<th>PROJECT</th>
<th>BS-NUMBER</th>
<th>COLUMN</th>
<th>CO_SOLVENT</th>
<th>SCORE0_1_2</th>
<th>TOTALSCORE</th>
<th>SELECTIVITY</th>
<th>RESOLUTION</th>
<th>RT(p1)</th>
<th>MASS_CONFIRMED(p1)</th>
<th>UV_AREA(p1)</th>
<th>UV_AREA%(p1)</th>
<th>PEAK_WIDTH(p1)</th>
<th>ASYMMETRY(p1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS18004578660</td>
<td>AD IPA</td>
<td>1</td>
<td>5.07</td>
<td>1.05</td>
<td>1.29</td>
<td>3</td>
<td>1</td>
<td>4243385</td>
<td>49.34</td>
<td>0.1</td>
<td>1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>IG IPA</td>
<td>1</td>
<td>4.45</td>
<td>1.05</td>
<td>1.06</td>
<td>3.09</td>
<td>1</td>
<td>4376503</td>
<td>49.52</td>
<td>0.13</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>IC IPA</td>
<td>2</td>
<td>13.64</td>
<td>1.24</td>
<td>3.87</td>
<td>3.58</td>
<td>1</td>
<td>4564263</td>
<td>49.93</td>
<td>0.18</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>C2 IPA</td>
<td>0</td>
<td>2.79</td>
<td>1.02</td>
<td>0.52</td>
<td>2.47</td>
<td>1</td>
<td>4281074</td>
<td>50.52</td>
<td>0.1</td>
<td>3.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>IB IPA</td>
<td>2</td>
<td>6.66</td>
<td>1.09</td>
<td>1.78</td>
<td>2.93</td>
<td>1</td>
<td>4270647</td>
<td>49.59</td>
<td>0.12</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>AD MeOH</td>
<td>0</td>
<td>3.52</td>
<td>1.02</td>
<td>0.75</td>
<td>2.85</td>
<td>1</td>
<td>4343244</td>
<td>49.23</td>
<td>0.08</td>
<td>1.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>IG MeOH</td>
<td>1</td>
<td>4.89</td>
<td>1.06</td>
<td>1.23</td>
<td>3.21</td>
<td>1</td>
<td>4546562</td>
<td>49.25</td>
<td>0.13</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>IC MeOH</td>
<td>2</td>
<td>6.72</td>
<td>1.09</td>
<td>1.83</td>
<td>3.38</td>
<td>1</td>
<td>4602337</td>
<td>49.57</td>
<td>0.15</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>C2 MeOH</td>
<td>0</td>
<td>2.95</td>
<td>1.02</td>
<td>0.56</td>
<td>2.44</td>
<td>1</td>
<td>4228262</td>
<td>51.63</td>
<td>0.09</td>
<td>3.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS18004578660</td>
<td>IB MeOH</td>
<td>2</td>
<td>6.25</td>
<td>1.07</td>
<td>1.73</td>
<td>3.13</td>
<td>1</td>
<td>4425744</td>
<td>49.25</td>
<td>0.11</td>
<td>1.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Chiral screening-methods

**100x4.6 mm, 5 μm Chiralpak AD, IG, IC, IBN, Chiralcel OZ**

IPA, MeOH + 0,1% NH₃

Gradient from 5→55% co-solvent in 5.0 min at 3 mL/min, 40°C

Time: 2 h

**Pros:**
- High sensitivity
- Coupling to MS
- Reliable system

**Cons:**
- «Slow»
- only 10 conditions

---

**150x4.6 mm, 5 μm Chiralpak AD, AS, AY, IE, Chiralcel OD, ID, OJ, Whelk-O1**

IPA, EtOH, MeOH + 0,1% NH₃

Gradient from 5→55% co-solvent in 7.5 min at 3 mL/min, 40°C

Time: 40 min

**Pros:**
- Fast
- 24 conditions

**Cons:**
- Low sensitivity
- Old system, unreliable
- Laborsome manual evaluation
The «SFC First» approach

- Standardized set-up similar to achiral lab and Cambridge
- Serial screen of 5 columns and 2 co-solvents (UPC2) = 10 conditions
- Direct translation to prep conditions via calibration function
- Additional 24 conditions on Sepiatec