PROFILING OF THE KNOWN-UNKNOWN PASSIFLORA COMPLEMENT BY LIQUID CHROMATOGRAPHY - ION MOBILITY - MASS SPECTROMETRY

Michael McCullagh¹, Jeff Goshawk¹, Russell J Mortishire-Smith¹, Cintia AM Pereira², Janete H Yariwake³, Johannes PC Vissers¹
¹ Waters Corporation, Wilmslow, United Kingdom, ² Centro Universitario Central Paulista - UNICEP, Sao Carlos, SP, Brazil, ³ Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil

OVERVIEW

- LC-IM-MS (retention time, collisional cross section (CCS) and m/z) determination of phytochemical composition and variance of Passiflora
- Multivariate analysis based LC-IM-MS data processing strategies revealed IM species specific variant differentiation
- Variant ‘known-unknown’ libraries identified unique medicinal plant species
- Predictive informatics strategies increased specificity of the identified ‘known-unknown’ isomeric complement

INTRODUCTION

C-Glycosyl flavonoids can be used as markers in the quality control of Passiflora phytomedicines. Several studies have focused on fingerprint analysis, quantification or identification of flavonoids in Passiflora using LC-MS and the principles of utilizing the combined specificity of LC-IM-MS for ‘known-unknown’ isomer profiling of Passiflora species have been demonstrated. However, structural elucidation of identified flavonoids responsible for phytochemical activity is still required. Hence, the application of LC-MS methods to profile flavonoid markers has significantly increased. Here, the analysis of Passiflora extracts been performed to generate ‘known-unknown’ speciation profiles. This approach can be combined with historical profiling of product ion identifications. Moreover, experimentally attained information was combined with CCS prediction to perform retention time independent elucidation of known flavonoids.

Figure 1. (A) P. edulis (PE), P. alata (PA), P. caerulea (PC) and P. incarnata (PI), (B) Genome size evolution in Passiflora [1], (C) Passiflora edulis chloroplast genome map [2] and (D) graphical synopsis of the experimental LC-IM-MS workflow.
**METHODS**

**Sample preparation**
Voucher specimens of all plant materials were dried at 35°C for 48 h, powdered and ground. Only particles between 0.5 - 1.0 mm in size were utilized for ethanol/water extraction, followed by SPE sample clean-up [3,4].

**IM-MS conditions**
- **MS:** Synapt G2-S/IM enabled oaToF MS
- **Mode:** ESI -ve
- **Desol./source:** 600°C/160°C
- **Calibrant:** MajorMix (power fit)
  - 150 - 1082 m/z (131 - 322 Å²)

**LC-conditions**
- **LC:** AQUITY I-Class system
- **Column:** 100 mm x 2.1 mm, 1.7 µm BEH C18 (45°C)
- **Mobile phase:** 0.1% formic acid in H₂O (A) and 0.1% formic acid in CH₃CN (B)
- **Gradient:** 1-80% B in 15 min @ 0.75 mL/min

**Informatics**
The data were acquired using MassLynx, which was further processed using Progenesis QI and UNIFI. Libraries, including retention time, TWCCSN₂, m/z, and MS/MS information were created with development software. TWCCSN₂ predictions were performed using a machine-learning approach and internally acquired TWCCSN₂ measurements to fit an appropriate model using the XGBoost algorithm. Multivariate analysis was conducted with SIMCA-P+ and Spotfire. LC-IM-MS data interpretation was performed with UNIFI and Progenesis QI.

---

Figure 2. (A) Variant distribution of the IM separated/detected LC-IM-MS complement of P. edulis, P. alata, P. caerulea and P. incarnata, (B), hierarchical clustering results of non-IM enabled processing of the LC-IM-MS data, and (C) PCA of IM enabled processing of the same data set. Color: blue (Px_MSE) = non-IM enable processing, purple (Px_HDMSE) = IM enabled processing.
RESULTS

Profiling and differentiation Passiflora variants
Species-variant diversity is demonstrated in Figure 2. The A panel shows the distribution of the 255 detected IM resolved isobaric/isomeric components across the species, suggesting that differentiation is feasible solely based on the IM separated components. The MVA results show that this is feasible as well by analyzing only the t and m/z domains, panel B, or by considering all dimensions during peak detecting, panel C, using the complete data matrix.

Known-unknowns
Species common components were identified based on precursor m/z, t, and CCS, as well as CID M/MS using a sample co-detection approach across a technical and biological replicates (n = 12 in total) from which ‘known-unknown’ libraries were constructed. Next, these libraries were using for screening using 4 ppm, 0.2 min, and 1% tolerances, respectively.

Detection characteristics examples are demonstrated in Figure 3, illustrating a detection dynamic range of ~ 3 orders, A panel, the commonality of P. edulis with the other Passiflora variants, B panel, and the identification of an ‘known-unknown’ using this strategy, C panel, ultimately leading to faster identification of novel structures and reduced FDR.

Figure 3. (A) Feature abundance distribution of the combined replicate LC-IM-MS detection results of P. edulis, (B) intersection of the ‘known-unknown’ LC-IM-MS complement of P. alata vs. other Passiflora variants (P. edulis, P. caerulea and P. incarnata) and (C) ‘known-unknown’ identification P. alata screened against P. caerulea, illustrating detection accuracy (top), extracted ion chromatogram (bottom left) and retention and drift time resolved MS1 and MSn spectra (bottom right), respectively.

Figure 4. Chemical structures of 6-C/8-C glycoside and tentatively identified flavonoids.

Figure 5. (A) Structure and fragmentation pathway of 6-C-glycosides isoorientin (R = OH) and isovitexin (R = H). The superscripts indicate bonds that can be broken under CID conditions and (B) ‘known-known’ retention and drift time resolved Passiflora flavonoid MSn product ion spectra example excerpt from Table 1 listed tentative identifications. Superscripts indicate bonds that can be broken under CID conditions [5].

Figure 6. (A) Mass measurement error and (B) TWCCSN2 prediction error (right) distributions for known, literature reported Passiflora flavonoids (RMSE values were determined for the compounds as summarized in Table 1 and that were detected in at least 2 out of 3 technical LC-IM-MS measurement replicates).
CSS prediction supported identification

A number of isomeric glucosides and flavonoids, shown in Figure 4 and Table 1, were targeted to attain the characteristics of the identification strategy and annotation of the 'known-unknown' complement. Example CID MS/MS spectra are shown in Figure 5 to allow for conformation of identification correctness based on fragmentation patterns [4].

On average, 10 to 25 isobaric/isomeric components per variant were detected using precursor m/z based screening. The number of candidate spectra was further reduced by ∼50% by including \(^{10}\text{WCCS}_{N2}\) prediction into the schema and using it as screening parameter, Table 1. Final annotation was achieved by confirming the presence and ratio of characteristic glucoside and flavonoids product ions. The observed mass measurement and \(^{10}\text{WCCS}_{N2}\) prediction errors are shown in Figure 6.

Apart from three compounds, which were either detected in a single Passiflora variant, not detected, or assignment was ambiguous, the majority of the target compounds could be annotated as 'known'.

Both 'known-unknowns' and 'knows' can also be quantitatively characterized as demonstrated in Figure 7, which shows the abundance distribution profile together with an annotated CID MS/MS fragmentation spectrum of the target flavonoids of interest, suggesting a relative higher concentration in one of the Passiflora variants.

<table>
<thead>
<tr>
<th>Compound</th>
<th>(\Delta\text{ppm}^*)</th>
<th>observed (t_r) (min)(**)</th>
<th>(#\text{detected isomers})(^1)</th>
<th>(^{10}\text{WCCS}_{N2}) (Å)</th>
<th>predicted(^2)</th>
<th>observed(^2)</th>
<th>variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apigenin</td>
<td>4.5</td>
<td>8.90</td>
<td>34</td>
<td>18</td>
<td>15</td>
<td>37</td>
<td>155.5 (0.6)</td>
</tr>
<tr>
<td>Hyperoside</td>
<td>3.0</td>
<td>8.50</td>
<td>18</td>
<td>8</td>
<td>6</td>
<td>10</td>
<td>199.3 (0.0)</td>
</tr>
<tr>
<td>Isoorientin</td>
<td>3.2</td>
<td>7.97</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>22</td>
<td>198.5 (1.5)</td>
</tr>
<tr>
<td>Isoorientin 2(\alpha)-glucopyranoside</td>
<td>3.7</td>
<td>7.36</td>
<td>21</td>
<td>17</td>
<td>17</td>
<td>13</td>
<td>232.0 (4.3)</td>
</tr>
<tr>
<td>Isosaparin</td>
<td>2.3</td>
<td>7.15</td>
<td>28</td>
<td>13</td>
<td>19</td>
<td>25</td>
<td>235.8 (0.7)</td>
</tr>
<tr>
<td>Isocosaparin 2(\beta)-glucoside</td>
<td>5.2</td>
<td>8.79</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>235.2 (6.2)</td>
</tr>
<tr>
<td>Isoisoflavone</td>
<td>3.5</td>
<td>8.11</td>
<td>19</td>
<td>7</td>
<td>23</td>
<td>21</td>
<td>226.4 (1.3)</td>
</tr>
<tr>
<td>Isotannin</td>
<td>3.0</td>
<td>8.54</td>
<td>22</td>
<td>10</td>
<td>16</td>
<td>25</td>
<td>196.6 (1.4)</td>
</tr>
<tr>
<td>Isotannin 2(\alpha)-glucopyranoside</td>
<td>2.1</td>
<td>7.23</td>
<td>27</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>233.8 (3.3)</td>
</tr>
<tr>
<td>Lucenin-2(\alpha)-glucopyranoside</td>
<td>3.0</td>
<td>7.00</td>
<td>19</td>
<td>14</td>
<td>17</td>
<td>13</td>
<td>231.9 (1.5)</td>
</tr>
<tr>
<td>Luteolin</td>
<td>2.7</td>
<td>9.60</td>
<td>37</td>
<td>26</td>
<td>14</td>
<td>37</td>
<td>158.2 (0.4)</td>
</tr>
<tr>
<td>Luteolin 6-C-fucoside 1</td>
<td>3.3</td>
<td>10.48</td>
<td>22</td>
<td>20</td>
<td>32</td>
<td>50</td>
<td>197.0 (0.9)</td>
</tr>
<tr>
<td>Luteolin 6-C-fucoside 2</td>
<td>2.4</td>
<td>9.75</td>
<td>14</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>197.0 (2.6)</td>
</tr>
<tr>
<td>Orientin</td>
<td>2.9</td>
<td>8.03</td>
<td>14</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>193.2 (0.9)</td>
</tr>
<tr>
<td>Saparoside</td>
<td>2.6</td>
<td>7.83</td>
<td>28</td>
<td>13</td>
<td>19</td>
<td>25</td>
<td>236.9 (0.8)</td>
</tr>
<tr>
<td>Schaftoside</td>
<td>1.9</td>
<td>8.33</td>
<td>19</td>
<td>7</td>
<td>23</td>
<td>21</td>
<td>226.4 (0.8)</td>
</tr>
<tr>
<td>Swertisin</td>
<td>2.5</td>
<td>8.99 (7.55)</td>
<td>6</td>
<td>8</td>
<td>20</td>
<td>12</td>
<td>201.4 (3.2)</td>
</tr>
<tr>
<td>Vicenin-2</td>
<td>2.6</td>
<td>7.49</td>
<td>27</td>
<td>14</td>
<td>8</td>
<td>31</td>
<td>231.5 (1.9)</td>
</tr>
<tr>
<td>Vitexin</td>
<td>3.4</td>
<td>8.67</td>
<td>10</td>
<td>5</td>
<td>28</td>
<td>15</td>
<td>192.2 (0.1)</td>
</tr>
<tr>
<td>Vitexin 2-xiloside</td>
<td>4.4</td>
<td>8.54</td>
<td>19</td>
<td>7</td>
<td>24</td>
<td>21</td>
<td>223.6 (3.4)</td>
</tr>
<tr>
<td>Vitexin 2-O-rhamnoside</td>
<td>2.5</td>
<td>8.67</td>
<td>23</td>
<td>13</td>
<td>10</td>
<td>24</td>
<td>226.3 (0.7)</td>
</tr>
<tr>
<td>Vitexin 4-O-rhamnoside</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>228.6</td>
</tr>
</tbody>
</table>

Table 1. \(^1\) compound (adduct specific) used to train a machine-learning \(^{10}\text{WCCS}_{N2}\) prediction model; \(^*\) root-mean-square error (RMSE) replicate injections and Passiflora variant samples; \(**\) detected by predicted \(^{10}\text{WCCS}_{N2}\) and CID MS/MS fragmentation interpretation; \(^4\) ambiguous detection; \(^1\) # (tentative) isomers detected determined by using 5 ppm, 10% \(^{10}\text{WCCS}_{N2}\) (post-detection) and 10 min \(t_r\) screening tolerances; \(^5\) mean absolute error (MAE) vs. observed \(^{10}\text{WCCS}_{N2}\) in parenthesis; -- = not detected.
CONCLUSION

- **Passiflora** ‘known-unknown’ consensus libraries were successfully created, validated and applied in qualitative and quantitative screening applications.
- CCS and retention time both contribute significantly to identification correctness with CCS illustrating reduced matrix effect dependency.
- Species profiling and characterization solely based on IM separation has been demonstrated and applied.
- The application of MS2 spectra, retention and drift time aligned ‘known-unknown’ spectra, and libraries affords detection, identification and FDR reduction of both unknowns and (expected) known analytes.
- The conceptual use of multi-variant matrix libraries including identification strategies that use predictive AI based methods and CID based MS/MS fragmentation knowledge for faster data analysis and FDR reduction have been demonstrated.

References