Supplementary Information

UGT74B1 from *Arabidopsis thaliana* as a versatile biocatalyst for the synthesis of desulfoglycosinolates

Sami Marroun†, Sabine Montaut‡, Stéphanie Marquès†, Pierre Lafite, Gaël Coadou, Patrick Rollin, Guillaume Jousset, Marie Schuler, Arnaud Tatibouët, Hassan Oulyadi and Richard Daniellou*

‡ Normandie Univ, COBRA, UMR 6014 et FR 3038; Univ Rouen; INSA Rouen; CNRS, IRCOF, 1 rue Tesnière, 76821 Mont Saint Aignan Cedex, France
‡ Department of Chemistry and Biochemistry, Biomolecular Sciences Programme, Laurentian University, 935 Ramsey Lake Road, Sudbury, ON P3E 2C6, Canada
† Université d’Orléans, CNRS, ICOA, UMR 7311, Orléans, France.

**Table S1:** Percentages of sequence identities for UGT74B1 and its templates.

**Table S2:** Percentages of sequence identities for UGT74B1 and its templates in the active site region.

**Figure S1:** Sequence alignment for UGT74B1 and its closest homologues.

**Figure S2:** *¹H* and *¹³C* NMR Spectra of compound 7.

**Figure S3:** HRMS Spectrum of compound 7.

**Figure S4:** *¹H* and *¹³C* NMR Spectra of compound 9.

**Figure S5:** HRMS Spectrum of compound 9.

**Figure S6:** *¹H* and *¹³C* NMR Spectra of compound 10.

**Figure S7:** HRMS Spectrum of compound 10.
Table S1: Percentages of sequence identities obtained by a dual structure-sequence multiple sequence alignment (green) and structural similarity scores (blue) obtained using CATH-SSAP for UGT74B1 and its templates. The maximum possible structural similarity score is 100.

<table>
<thead>
<tr>
<th></th>
<th>74B1</th>
<th>2VCE</th>
<th>3HBF</th>
<th>2C1Z</th>
<th>2PQ6</th>
<th>4WHM</th>
<th>2ACW</th>
</tr>
</thead>
<tbody>
<tr>
<td>74B1</td>
<td>100</td>
<td>85.65</td>
<td>87.53</td>
<td>88.12</td>
<td>86.55</td>
<td>86.61</td>
<td>84.47</td>
</tr>
<tr>
<td>2VCE</td>
<td>25.83</td>
<td>100</td>
<td>83.7</td>
<td>84.01</td>
<td>84.99</td>
<td>82.56</td>
<td>87.09</td>
</tr>
<tr>
<td>3HBF</td>
<td>23.26</td>
<td>24.58</td>
<td>100</td>
<td>91.09</td>
<td>85.14</td>
<td>87.88</td>
<td>83.17</td>
</tr>
<tr>
<td>2C1Z</td>
<td>23.48</td>
<td>24.79</td>
<td>48.03</td>
<td>100</td>
<td>85.82</td>
<td>89.27</td>
<td>83.46</td>
</tr>
<tr>
<td>2PQ6</td>
<td>25.93</td>
<td>28.01</td>
<td>23.86</td>
<td>22.82</td>
<td>100</td>
<td>84.4</td>
<td>83.53</td>
</tr>
<tr>
<td>4WHM</td>
<td>26.09</td>
<td>22.71</td>
<td>39.65</td>
<td>42.98</td>
<td>21.99</td>
<td>100</td>
<td>82.99</td>
</tr>
<tr>
<td>2ACW</td>
<td>22.15</td>
<td>29.79</td>
<td>23.44</td>
<td>22.8</td>
<td>24.27</td>
<td>20.86</td>
<td>100</td>
</tr>
</tbody>
</table>

Table S2: Percentages of sequence identities obtained by a dual structure-sequence multiple sequence alignment for UGT74B1 and its templates in the active site region.
Figure S1 (continued): Sequence alignment for UGT74B1 and its closest homologues using the TCOFFEE accurate mode. Secondary structures are highlighted in green for sheets and in yellow for helices. Secondary structure assignment for UGT74B1 is based on the homology model obtained using the MODELLER program.
Figure S2: $^1$H (top) and $^{13}$C (bottom) NMR Spectra of compound 7.
Figure S3: HRMS Spectrum of compound 7.
Figure S4: $^1$H (top) and $^{13}$C (bottom) NMR Spectra of compound 9.
Figure S5: HRMS Spectrum of compound 9.
Figure S6: $^1$H (top) and $^{13}$C (bottom) NMR Spectra of compound 10.
Figure S7: HRMS Spectrum of compound 10.