Web-based Supplementary Materials for
“Hierarchical Bayesian Modelling of Pharmacophores in Bioinformatics” by

Web Appendix A

Illustrative Example

Here we present a hypothetical example for six configurations. We start with a set of six
templates which are the individual configurations that we denote by
\[ \chi = \{1, 2, 3, 4, 5, 6\} \].

We use a scoring criteria \( G_{ab} \) for evaluating pairwise alignments between two templates \( a \) and \( b \) and we use a threshold value \( G_{min} \), which any pairwise alignment must exceed to be considered for merging. Here we suppose \( G_{min} = 0.50 \). At the first stage we evaluate all fifteen pairwise alignments involving the six starting configurations and merge pair with the highest score exceeding \( G_{min} \). Recall that we denote templates formed from, say, configurations \( a, b \) and \( c \) by \( T_{abc} \). For example, \( T_{123} \) would be the template derived from configurations 1, 2 and 3. An important point to note is that templates are derived from only the common matched points to all configurations and are formed by taking the average location of each matched point. Below we give similarity matrices for the pairwise alignments between the templates remaining at each stage of implementation.

Step 1

\[
\begin{pmatrix}
2 & 3 & 4 & 5 & 6 \\
1 & (0.85 & 0.75 & 0.40 & 0.30 & 0.20) \\
2 & 0.45 & 0.80 & 0.40 & 0.40 \\
3 & 0.74 & 0.70 & 0.80 \\
4 & 0.55 & 0.50 \\
5 & 0.40
\end{pmatrix}
\]

Step 2

\[
\begin{pmatrix}
3 & 4 & 5 & 6 \\
T_{12} & (0.60 & 0.50 & 0.65 & 0.25) \\
3 & 0.74 & 0.70 & 0.80 \\
4 & 0.55 & 0.50 \\
5 & 0.40
\end{pmatrix}
\]

Step 3

\[
\begin{pmatrix}
T_{36} & 4 & 5 \\
T_{12} & (0.60 & 0.50 & 0.65) \\
T_{36} & 0.90 & 0.62 \\
4 & 0.55
\end{pmatrix}
\]

Step 4

\[
\begin{pmatrix}
T_{346} & 5 \\
T_{12} & (0.65 & 0.65) \\
T_{346} & 0.95
\end{pmatrix}
\]

Step 5

\[ G_{123456} = 0.45 < G_{min} \]
Web Table 1: Candidate templates at each stage

<table>
<thead>
<tr>
<th>Step</th>
<th>Max Similarity</th>
<th>Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$G_{12} = 0.85$</td>
<td>$T_{12}, 3, 4, 5, 6$</td>
</tr>
<tr>
<td>2</td>
<td>$G_{36} = 0.80$</td>
<td>$T_{12}, T_{36}, 4, 5$</td>
</tr>
<tr>
<td>3</td>
<td>$G_{346} = 0.90$</td>
<td>$T_{12}, T_{346}, 5$</td>
</tr>
<tr>
<td>4</td>
<td>$G_{3456} = 0.95$</td>
<td>$T_{12}, T_{3456}$</td>
</tr>
<tr>
<td>5</td>
<td>$G_{123456} = 0.45$</td>
<td>$T_{12}, T_{3456}$</td>
</tr>
</tbody>
</table>

No new templates

Web Table 1 summarises the candidate templates output at each stage of the implementation.
At step 5, the maximum geometric mean of possible merges, (in fact, there is only one possible merge at this stage), is less than the threshold value $G_{\text{min}}$, hence no merge takes place and the algorithm terminates. Two templates are output: $T_{12}$ and $T_{3456}$.

Web Appendix B

1ATP without colouring information
Here we show the results of the 1ATP example without using chemical information in the form of colouring by atom type. The subsets of configurations found are identical to the case where colouring information is used and are shown again in Web Table 2. The configurations contributing to each template are shown in Web Figure 1.

Web Table 2: Probable templates for 1ATP

<table>
<thead>
<tr>
<th>Template #</th>
<th>Configurations</th>
<th># of common atoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2 3 6</td>
<td>35</td>
</tr>
<tr>
<td>B</td>
<td>4 5 7</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>9 10 11</td>
<td>26</td>
</tr>
</tbody>
</table>

3PTB without colouring information
In this case, using a value of $G_{\text{min}} = 0.4$ we obtain the same subsets as in the case where colouring information is used. These are shown in Web Table 3. When we use a value of $G_{\text{min}} = 0.5$ as previously, configurations 16 and 22 are not included in template B and remain as singletons. Expert opinion is that it sensible for these ligands to be included in
Web Figure 1: 1ATP ligands contributing to the templates found by the HT algorithm, without using colouring information. The template geometry is defined by the mean position of common matching atoms.
the subset as well, hence we allow the reduced threshold \( G_{\text{min}} = 0.4 \). The configurations contributing to the templates are shown in Web Figure 2.

Web Table 3: Probable templates for 3PTB

<table>
<thead>
<tr>
<th>Template #</th>
<th>Configurations</th>
<th># of common atoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19 21</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>12 14 15 16 17 18 20 22</td>
<td>9</td>
</tr>
</tbody>
</table>
Additional Figures

Colour figures of results from examples given in the main body of the paper.

Template A and Template B

Template C

Web Figure 3: 1ATP ligands contributing to the templates found by the HT algorithm. The template geometry is defined by the mean position of common matching atoms.
Web Figure 4: 3PTB ligands contributing to the templates found by the HT algorithm. The template geometry is defined by the mean position of common matching atoms.