MHFP: A Probabilistic Molecular Fingerprint for Big Data Settings
I stand at the seashore, alone, and start to think. There are the rushing waves mountains of molecules each stupidly minding its own business trillions apart yet forming white surf in unison.

\[ P(h_{min}(A) = h_{min}(B)) = J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]
I stand at the seashore, alone, and start to think. There are the rushes waves mountains of molecules each stupidly minding its own business trillions apart yet forming white surf in unison.

\[
P(h_{\text{min}}(A) = h_{\text{min}}(B)) = J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]
I stand at the seashore, alone, and start to think. There are the rushing waves, mountains of molecules each stupidly minding its own business, trillions apart, yet forming white surf in unison.

\[
P(h_{\min}(A) = h_{\min}(B)) = J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]
I stand at the seashore, alone, and start to think.
There are the rushing waves
mountains of molecules
each stupidly minding its own business
trillions apart
yet forming white surf in unison.

\[ P(h_{min}(A) = h_{min}(B)) = J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]
I stand at the seashore, alone, and start to think. There are the rushing waves, mountains of molecules, each stupidly minding its own business, trillions apart yet forming white surf in unison.

\[ P(h_{\text{min}}(A) = h_{\text{min}}(B)) = J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]
I stand at the seashore, alone, and start to think. There are the rushing waves mountains of molecules each stupidly minding its own business trillions apart yet forming white surf in unison.

$$P(h_{min}(A) = h_{min}(B)) = J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$
<table>
<thead>
<tr>
<th>Method</th>
<th>AUC</th>
<th>EF 1%</th>
<th>EF 5%</th>
<th>BEDROC20</th>
<th>EDROC100</th>
<th>RIE20</th>
<th>RIE100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shingling (r=2)</td>
<td>0.773</td>
<td>31.590</td>
<td>9.718</td>
<td>0.496</td>
<td>0.492</td>
<td>8.520</td>
<td>25.368</td>
</tr>
<tr>
<td>Shingling (r=3)</td>
<td>0.776</td>
<td>32.698</td>
<td>9.897</td>
<td>0.506</td>
<td>0.505</td>
<td>8.691</td>
<td>26.118</td>
</tr>
<tr>
<td>ECFP (r=2)</td>
<td>0.755</td>
<td>32.702</td>
<td>9.792</td>
<td>0.500</td>
<td>0.503</td>
<td>8.590</td>
<td>26.027</td>
</tr>
<tr>
<td>ECFP (r=3)</td>
<td>0.754</td>
<td>32.563</td>
<td>9.780</td>
<td>0.499</td>
<td>0.502</td>
<td>8.571</td>
<td>25.956</td>
</tr>
</tbody>
</table>