Intramolecular hydrogen bond expectations in medicinal chemistry

Fabrizio Giordanetto\textsuperscript{a}, Christian Tyrchan\textsuperscript{b}, Johan Ulander\textsuperscript{c}

\textsuperscript{a}D.E. Shaw Research, 120W 45th Street, New York, NY 10036, USA

\textsuperscript{b}Medicinal Chemistry, Respiratory and Inflammation, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Mölndal, Pepparedsleden 1, SE-431 83, Mölndal, Sweden

\textsuperscript{c}Medicinal Chemistry, Cardiovascular and Metabolic Diseases, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Mölndal, Pepparedsleden 1, SE-431 83, Mölndal, Sweden

Supplementary Information
1. Materials and Methods

Molecular topologies with highest frequency of IMHB occurrence (>70%), as described by Kuhn et al.\textsuperscript{1} (Table S1) and selected monoatomic transformations (Table S2) served to define the corresponding SMIRKS\textsuperscript{2} for the sought matched molecular pairs (MMP). These definitions were employed to query a subset of ChEMBL\textsuperscript{3} (i.e., compounds with molecular weight < 750 Da) using KNIME\textsuperscript{4} and the matches and associated biological data returned for further analysis. MMPs were removed if any of the following conditions applied: 1) the transformed compound was still capable to form an IMHB due to structural symmetry or additional IMHB features in its topology, 2) the transformation resulted in a net charge change for the compound, 3) the associated biological data was not of the “binding” type as from ChEMBL definitions\textsuperscript{3}, was missing, labelled as inconclusive or undefined, originated from single concentration response experiments or high throughput screening assays. Numerical analysis and visualization of the resulting data set was performed using Dotmatics’ Vortex\textsuperscript{5}.

2. Supplementary Tables

<table>
<thead>
<tr>
<th>IMHB pair (pseudo ring)</th>
<th>Topology</th>
<th>% H-bond occurrence in CSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>C=O···H-N (6-membered)</td>
<td>cC3aC3a</td>
<td>93</td>
</tr>
<tr>
<td>C=O···H-N (6-membered)</td>
<td>aC3cC3a</td>
<td>89.5</td>
</tr>
<tr>
<td>C=O···H-N (6-membered)</td>
<td>cNaC3a</td>
<td>96</td>
</tr>
<tr>
<td>C=O···H-N (6-membered)</td>
<td>aNaC3c</td>
<td>93.5</td>
</tr>
<tr>
<td>C=O···H-N (6-membered)</td>
<td>aNaC3a</td>
<td>85.3</td>
</tr>
<tr>
<td>C=O···H-O (6-membered)</td>
<td>aC3cC3a</td>
<td>84.7</td>
</tr>
<tr>
<td>N···H-N (6-membered)</td>
<td>aNaC3a</td>
<td>80.9</td>
</tr>
<tr>
<td>C-O···H-N (5-membered)</td>
<td>aC4aC3a</td>
<td>74.5</td>
</tr>
<tr>
<td>C-O···H-N (5-membered)</td>
<td>aC3cC3a</td>
<td>71.2</td>
</tr>
<tr>
<td>C-O···H-N (6-membered)</td>
<td>aC3cC3aC3a</td>
<td>87.9</td>
</tr>
<tr>
<td>C-O···H-N (6-membered)</td>
<td>aC3cC3aC3c</td>
<td>70</td>
</tr>
</tbody>
</table>

Table S1. Topologies with the highest occurrence of IMHB, based on Kuhn et al.\textsuperscript{1}
<table>
<thead>
<tr>
<th>Structure</th>
<th>Fragment</th>
<th>HBD</th>
<th>Alkox-NH</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Structures" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Details:**
- **HBD:** Heteroatom Bonding Domain
- **Alkox-NH:** Alkoxy-Nitrogen Heterodomain
- **Charge:** Charge status of the structure
### Table S2. SMIRKS definition of chemical transformations used to identify IMHB MMPs.

<table>
<thead>
<tr>
<th>Topology</th>
<th>IMHB Pair</th>
<th>Pseudo Ring</th>
<th>Size</th>
<th>Transform</th>
<th>Bioactivity Type</th>
<th>IMHB Compound Bioactivity Value</th>
<th>“No IMHB” Compound Bioactivity Value</th>
<th>∆log (Bioactivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aC3cC3a</td>
<td>Carbonyl-NH 6</td>
<td>HBD-CH</td>
<td>6</td>
<td>aC3cC3a</td>
<td>ALS IC_{50} (µM)</td>
<td>100</td>
<td>&gt;100</td>
<td>2.74</td>
</tr>
<tr>
<td>aC3cC3a</td>
<td>Carbonyl-NH 6</td>
<td>HBD-Me</td>
<td>6</td>
<td>aC3cC3a</td>
<td>JNK1 IC_{50} (µM)</td>
<td>0.008</td>
<td>&gt;20</td>
<td>-3.39</td>
</tr>
<tr>
<td>Compound</td>
<td>Assay</td>
<td>IC₅₀ (µM)</td>
<td>Kᵢ (µM)</td>
<td>pIC₅₀</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
<td>----------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aC₃C₃aC₃a Alkoxy-NH₆ HBA-CH</td>
<td>HCN1</td>
<td>0.032</td>
<td>&gt;199</td>
<td>-3.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aC₃C₃aC₃a Alkoxy-NH₆ HBD-O</td>
<td>5-HT4R</td>
<td>0.079</td>
<td>0.0001</td>
<td>2.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aC₃C₃cC₃c Alkoxy-NH₆ HBA-CH</td>
<td>CASP1</td>
<td>0.090</td>
<td>&gt;100</td>
<td>-3.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aNaC₃a Nhet-NH₆ HBD-CH</td>
<td>A2A AR</td>
<td>0.0048</td>
<td>47.2</td>
<td>-3.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aNaC₃c Carbonyl-NH₆ HBD-Me</td>
<td>TIE-2</td>
<td>0.0069</td>
<td>5.6</td>
<td>-2.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cC₃aC₃a Carbonyl-NH₆ HBD-Me</td>
<td>c-Met</td>
<td>0.001</td>
<td>&gt;20</td>
<td>-4.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cC3aC3a Carbonyl-NH</td>
<td>NS5B IC50 (nM)14</td>
<td>16</td>
<td>0.02</td>
<td>2.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 HBD-O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table S3.** Representative IMHB MMPs outliers defined as Δlog(Bioactivity) > 1.5 * interquartile range. IMHB topologies and atomic pairs are highlighted in bold.
1. References


