Biomolecule Uniqueness Check Using HELM and Hashcode

David Deng
Director of Software Development

Tony Yuan
President, Head of R&D

InChI Symposium San Diego
Aug 23rd, 2019
Scilligence’s Uniqueness Check

- Scilligence’s cartridge supports both small molecules and biomolecules
- Uniqueness check uses a proprietary Hashcode approach
  - Similar to InChIKey
  - For both small and biomolecules
- Leading support in HELM for biomolecules
### HELM Monomer Library Management

<table>
<thead>
<tr>
<th>Monomer</th>
<th>Symbol</th>
<th>Alases:</th>
<th>Name:</th>
<th>Natural Analog:</th>
<th>Polymer Type:</th>
<th>Monomer Type:</th>
<th>Backbone:</th>
<th>Substructure:</th>
<th>Display:</th>
<th>Count:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>Alaine</td>
<td>Alaine</td>
<td>2-aminoacidic acid</td>
<td>PEPTIDE</td>
<td>Backbone</td>
<td>R1, H</td>
<td>R2, OH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Author:</td>
<td>Created Date: 2018-Aug-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Versions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ac</td>
<td>Ac</td>
<td>Abu</td>
<td>Name: 2-amino-2-propanoic acid</td>
<td>N-Terminal Acetic</td>
<td>PEPTIDE</td>
<td>Backbone</td>
<td>R1, H</td>
<td>R2, OH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Author:</td>
<td>Created Date: 2018-Aug-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Versions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asu</td>
<td>Asu</td>
<td>App</td>
<td>Name: 2-amino-3-butoic acid</td>
<td>gamma-amino-butyric acid</td>
<td>PEPTIDE</td>
<td>Backbone</td>
<td>R1, H</td>
<td>R2, OH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Author:</td>
<td>Created Date: 2018-Aug-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Versions:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HELM Sketcher: Monomer Selection
HELM Sketcher: HELM Structure/String Conversion

PEPTIDE1{A.S.D.F.D.F.S.D.F.A.S.F.D}$$$$V2.0
HELM Sketcher: RNA Complementary Strand

RNA1{R(A)P.[12ddR][5meC][sP].[aR][cpmA][bP].[qR][5fU][sP].[mR][meA][bP].[5FBC6][dabA][bP]}

$\$$\$$\$$V2.0
Backend: Scilligence’s Uniqueness Check
However ....

Expanding is not always practical
HELM Sketcher: Hybrid of Entities (Mixed Mode)
PEGS Monomer

Symbol: APeg
Name: Alanine with long PEGS side chain
Polymer Type: PEPTIDE
Monomer Type: Backbone
Natural Analog: A
Status: Retired

Structure:

R1: H
R2: OH
R3: 
R4: 
R5: 

Chiral
When to Hash??

Certain Limits

HASHCODE
Pros vs. Cons

• **Pros**
  - Works for most common biopolymers
  - Balance between accuracy and performances
  - “Limit” configurable

• **Cons**
  - Monomer library dependent
  - Modifying monomer structures will change the HashCode
Thank You!

www.scilligence.com