Pistoia Alliance New Project Portfolio

Hierarchical Editing Language for Macromolecules

Domain Lead - David Klatte
Business IT, Pfizer
Therapeutic Large Molecules

- ASOs, siRNAs
- Peptides
- Antibodies
- Therapeutic Proteins
- ADCs
- Vaccines

Structure Characteristics:
- Large
- Diverse
- Complex
- Chemical Modifications
- Multiple Components

HELM: A Hierarchical Editing Language for Macromolecules

1. Intuitive
   - Easy to understand by scientists
2. Hierarchical
   - Atom \( \rightarrow \) Monomer \( \rightarrow \) Simple Polymer \( \rightarrow \) Complex Polymer
3. Extensible
   - Supports different types of macromolecules
4. Computer Readable
   - Toolkit for notation parsing and validation
5. “Canonicalizable”
   - Canonical notation used for uniqueness checking in compound registration

HELM Notation Example

Oligonucleotide-ChemLinker-Peptide

Structure:
5'-AGmCUU~PEG3~n-KKEKG

Notation:
RNA1(R(A)R(P)(G)P)nR1(C)P.R(U)P.R(U)P.R(U)P.R(U)P.R1(PEPTIDE1(K,K,E,K,G,C)
|CHEM1{PEG3}
|$RNA1, CHEM1,18:R2-1:R1
|PEPTIDE1, CHEM1,3:R3-1:R2
|$

Specs:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>To delimit top level components</td>
</tr>
<tr>
<td></td>
<td>To delimit list item within a top level component</td>
</tr>
<tr>
<td>{}</td>
<td>To enclose simple polymer notation</td>
</tr>
<tr>
<td>.-</td>
<td>To delimit monomer groups within simple polymer notation</td>
</tr>
<tr>
<td>[ ]</td>
<td>To enclose a branch monomer within simple polymer notation</td>
</tr>
<tr>
<td>[ ]</td>
<td>To enclose a modified monomer (non-natural)</td>
</tr>
<tr>
<td>. . .</td>
<td>To describe connection and hydrogen bonding</td>
</tr>
</tbody>
</table>

Summary

1. HELM notation language was conceived in the summer of 2008 to support Pfizer’s Oligonucleotide Therapeutic Unit.
2. The first version of PME and PMR was released into production in December 2008.
3. HELM notation has been extended to support in-line chemical structure and ambiguous connections.
4. The current version of PME is 3.0, which supports structure drawing and visualization for oligonucleotides, peptides, antibodies and proteins, and antibody drug conjugates.
5. The current version of PMR is 3.1, which supports registration for oligonucleotides, peptides, antibodies and proteins.
## Canonical Models for Large Molecules

### Unmet Need

- There is no standard approach for a computer-based way of managing large molecules such as:
  - Peptides, Antibodies, Therapeutic Proteins and Vaccines
- Small molecules can be represented by atom bond representation, oligomers require monomer representations.
- Component-level representations are required for big proteins and antibodies.

### Value Proposition

- HELM has been developed, tested and used by Pfizer
- HELM is being published and its supporting software tools will be put into the open-source public domain
- Pistoia Alliance to “sign-post” this standard and foster its adoption throughout the Pistoia Alliance community

### Business Case

- **Aggregated Savings across the industry is estimated at $??? M** per year as a result of:
  - Standard reference notation for macromolecules in pharma compound registration systems
  - The Pistoia Alliance is the best-positioned, formally-established, x-company organisation equipped to co-ordinate this kind of pre-competitive collaboration

### High Level Implementation Timeline

<table>
<thead>
<tr>
<th>Event</th>
<th>Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding Agreed</td>
<td>Q4 - 2012</td>
</tr>
<tr>
<td>Funding Available</td>
<td>Q1 - 2013</td>
</tr>
<tr>
<td>Project Start</td>
<td>Q2 - 2013</td>
</tr>
<tr>
<td>Standard Available</td>
<td>Q3 - 2013</td>
</tr>
</tbody>
</table>

| Q4 - 2013 |
PA App Strategy

Vision

The next wave of computation in science will be centered on mobile devices and there needs to be services to cater for this need.

Function

We will foster the Life Sciences Mobile App eco-system, encourage a strong science App and services creation culture and be seen as the goto organization for mobile science

Value

- **Aggregated Savings across the industry is estimated at $209 M per year (estimate) as a result of:**
  - Mobile-device enabled life science R&D – “Science-on-the-Go”
  - Laboratory standard apps on mobile devices decrease the cost of hardware and software
  - PA is expertly equipped to co-ordinate this app ecosystem and services creation culture

- **Projected Shared Project Costs ~$150,000**

Timing

<table>
<thead>
<tr>
<th>Funding Agreed</th>
<th>Funding Available</th>
<th>Project Start</th>
<th>Robust App Catalogue in use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 - 2012</td>
<td>Q1 - 2013</td>
<td>Q2 - 2013</td>
<td>Q4 - 2013</td>
</tr>
</tbody>
</table>
Interested parties so far include:

<table>
<thead>
<tr>
<th>Pharma</th>
<th>Technology Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Klatte, Pfizer</td>
<td>Alex Drijver, ChemAxon</td>
</tr>
<tr>
<td>Ian Stott &amp; Jerry Winter, Unilever</td>
<td>John McCarthy, Accelrys</td>
</tr>
<tr>
<td>Simon Thornber, GSK</td>
<td>David Lowis, Certara</td>
</tr>
<tr>
<td>Dirk Bornemeier, Abbott</td>
<td>Hans DeBie, ACDLabs</td>
</tr>
<tr>
<td>Ramesh Durvasula, BMS</td>
<td>Jinbo Lee, Scilligence</td>
</tr>
<tr>
<td>Stefan Klostermann, Roche</td>
<td></td>
</tr>
<tr>
<td>Claus Stie Kallesøe</td>
<td></td>
</tr>
<tr>
<td>Steve Heller, InChI Trust</td>
<td></td>
</tr>
</tbody>
</table>