Embracing ambiguity: Representation of macromolecules using the enhanced standard HELM 2.0

AUTHORS

Markus Weisser [1]

Cathrin Mayer [1]

Sabrina Hecht [1]

Bernhard Shirm [1]

Claire Bellamy[2]

Jan Holst Jensen [3]

Matthias Nolte [4]

Yohann Potier[5]

Sven Neumeyer [6]

Roland Knispel [7]

Stefan Klostermann [8]

Tianhong Zhang [9]

Sergio Rotstein [9]

Affiliations

- [1] quattro research GmbH, Fraunhoferstr. 18a, 82152 Planegg, Germany
- [2] Pistoia Alliance, Inc., 401 Edgewater Place, Suite 600, Wakefield, MA 01880, USA
- [3] Biochemfusion ApS, Lindegardsvej 44, 2920 Charlottenlund, Denmark
- [4] Bristol-Myers Squibb Inc., Research Pkwy, Wallingford, CT 06492, USA
- [5] Novartis Inc., 211 Massachusetts Ave, Cambridge, MA 02139, USA
- [6] Novartis Institutes for BioMedical Research, Fabrikstrasse 2, 4056 Basel, Switzerland
- [7] ChemAxon Ltd., Graphisoft Park, Záhony utca 7, Budapest, 1031 Hungary
- [8] Roche Innovation Center Penzberg, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany
- [9] Pfizer Inc., One Burtt Road, Andover, MA 01810, USA

References

Pistoia Alliance HELM Project on GitHub:

https://github.com/PistoiaHELM

Journal articles

T. Zhang at al. J.Chem. Inf. Model, 2012, 52(10), 2796

Introduction

As the therapeutic utilization of complex and non-standard biomolecules has become commonplace in drug discovery R&D, scientists have struggled to represent these entities in their information systems, forcing them to use very non-standard “pick and mix” approaches that include multiple nomenclatures and textual descriptions. HELM, the open biomolecular representation standard, has solved this problem by providing a means to represent multiple types of complex macromolecules (e.g. nucleotides, proteins, antibodies and antibody-drug conjugates) including those that contain non-natural elements such as chemically modified amino acids

The Pistoia Alliance formalized the HELM notation, originally created by Pfizer scientists, as an open standard in early 2013 and publicly released the related software toolkit and editor to the Open Source community.

While the original HELM solved the problem of representing unnatural complex biomolecules, it held the basic assumption that a scientist would know everything about the structures being represented. In biology, however, this is not always the case, as various types of ambiguity or uncertainty may exist in some of these structures.

Ambiguity in HELM 2.0

HELM 2.0 enables scientists to incorporate into their structural representation a systematic description of a number of forms of ambiguity, thus allowing for the rigorous representation of structures, even when uncertainty is present. A typical example is when the exact conjugation site of an antibody-drug conjugate cannot be determined.

Monomer Ambiguity

- Not fully characterized monomer sequences. Examples:

CHARACTER	DESCRIPTION	HELM 2.0 EXAMPLE
*	0..n unknown monomers	PEPTIDE1{A.*.G.C}$$$$V2.0
X	Single unknown amino acid in a PEPTIDE	PEPTIDE1{A.X.G.C}$$$$V2.0
N	Single unknown base in a RNA	RNA1{R(A)P.R(N)P.R(C)P.R(C)P.R(C)}$$$$V2.0
(,)	One of a list of monomer is possible	PEPTIDE1{A.(A:10,G:90).G.C}$$$$V2.0
(+)	Mixture of monomers	PEPTIDE1{A.(A+G+C).G.C}$$$$V2.0
–	Deleted or missing single monomer	PEPTIDE1{A.(A,_).G.C}$$$$V2.0
´ ´	Repeating monomers	PEPTIDE1{A.G.A.C.A'5-30'}$$$$V2.0

Table 1: Ambiguity examples

Polymer Ambiguity

- Complete sequence of a polymer is unknown.

Connection Ambiguity

- The details of the connections between polymers are unknown.

Grouping Ambiguity

- Unknown details of a grouping.

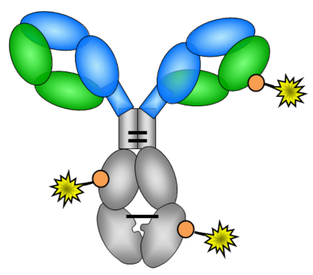

Figure 1: Antibody-drug conjugate

New HELM2 Notation Toolkit architecture

In addition to the representation of ambiguity, the HELM 2.0 toolkit provides a couple of additional enhancements that will make it easier for companies to integrate HELM into their infrastructure:

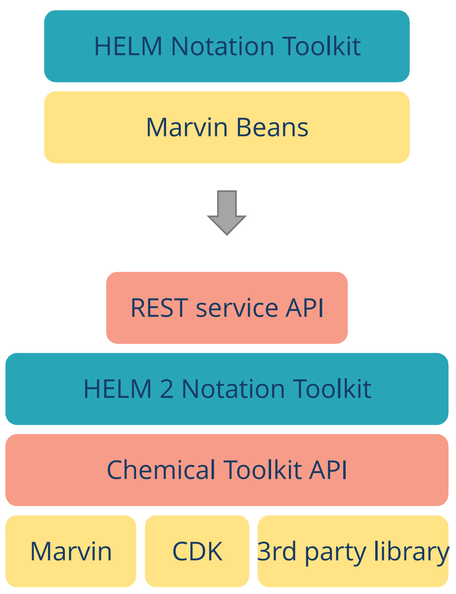

Figure 2: New HELM architecture

1. An API that allows the toolkit to leverage different cheminformatics software libraries

2. A set of web services that abstract the toolkit functionality from the toolkit code to facilitate calls to and from other software

Conclusion

With the addition of ambiguity support, HELM 2.0 now provides researchers the unique capability of representing complex biological entities that have not yet been fully characterized at the structural level, rendering it an even more practical technology for the electronic representation of a wide array of biomolecules.

By additionally enabling the use of different chemical libraries and providing web-services, HELM is now more open and practical technology than ever before.

The HELM code is available on GitHub and uses the permissive MIT open source license, which gives anyone the right to freely download and customize it. Please visit OpenHELM.org for additional information about the project.

quattro research GmbH

quattro research GmbH is a software and service company located in Munich, Germany. quattro research offers services and products for data management in the pharmaceutical, chemical, and biotech industry. Within several projects, quattro research developed extensive knowledge in the management and registration of biologics, e.g. antibodies and peptides.

www.quattro-research.com

Pistoia Alliance: Lowering barriers to R&D innovation

The Pistoia Alliance is a global, not-for-profit alliance of life science companies, vendors, publishers, and academic groups that work together to lower barriers to innovation in R&D. Our members collaborate as equals on open projects that generate significant value for the worldwide life sciences community.

www.pistoiaalliance.org

@pistoiaalliance