HELM:
Embracing Ambiguity

Key Information

Project Lead: Sergio Rotstein

Project Manager: Claire Bellamy

Goal of the HELM Project

Lower barriers to adoption of the HELM standard.

Team Members

Dennis Hansen
Sabrina Hecht
Jan Holst Jensen
Roland Knispel
Jeff Milton
Matthias Nolte
Yohann Potier
Joann Prescott-Roy
Sven Neumeyer
Stefan Klostermann
Eric Swayze
Markus Weisser
Tony (Yongwen) Yuan
Tianhong Zhang

Major contributors

Adopters/supporters

Pharma / Biotech / Institutes

BMS, GSK, Ionis, Merck, Novartis, Pfizer, Roche

Software vendors

Biovia, ChemAxon, NextMove Software, PerkinElmer, Scilligence

Regulatory

ISO 11238

Content / Service Providers

EBI (ChEMBL), NCBI (PubChem), quattro research

Info@openHELM.org

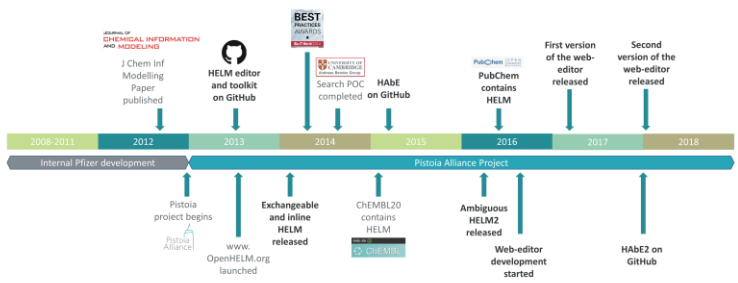
www.openHELM.org

Tianhong Zhang*, Hongli Li, Hualin Xi, Robert V. Stanton, and Sergio H. Rotstein *HELM: A Hierarchical Notation Language for Complex Biomolecule Structure Representation*, J. Chem. Inf. Model., 2012, 52 (10), pp 2796–2806

Introduction

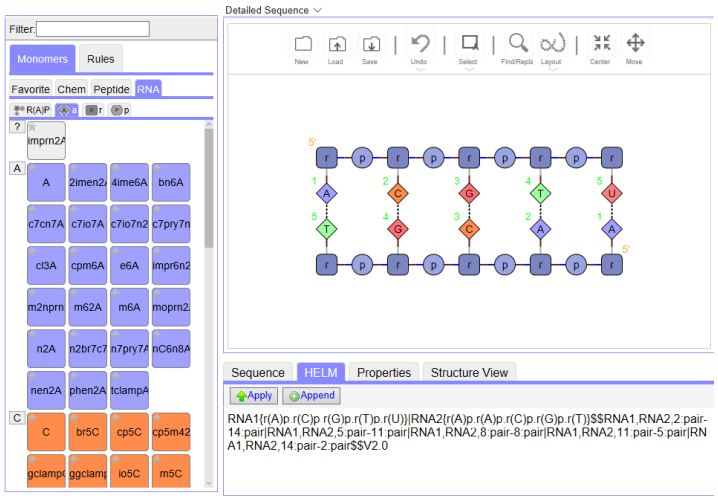
As the therapeutic utilization of complex and non-standard biomolecules has become commonplace in drug discovery R&D, scientists have struggled to represent these entities in their information systems, forcing them to use very non-standard "pick and mix" approaches that include multiple nomenclatures and textual descriptions. HELM, the open biomolecular representation standard, has solved this problem by providing a means to represent many 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. Since 2013 the project has extended the tools and notation and supported community initiatives such as monomer.org with its 'starter set' of well curated and systematically named monomers.



Web-editor

The original HELM Applet editor, supporting HELM Notation 1.0 and developed using Java, made it easier for scientists to draw and view macromolecules. The evolution of browser technology, however, resulted in a general reduction of support for Java applets. In response, the team released the HELM Web Editor in February 2017 which was developed from the ground up in JavaScript, which removed dependencies and maximized browser compatibility"



Solving the problem of 'known unknowns'

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 often exist in these structures. For example, the site of conjugation of a drug to an antibody may not be known or there may be multiple options (Fig 1)

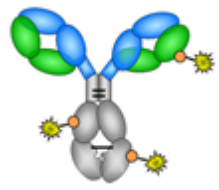


Figure 1: Antibody-drug conjugate

To solve this problem HELM includes three different types of uncertainty or ambiguity:

Component Ambiguity: The chemical structure of part or all of the molecule is not known.

Connection Ambiguity: The details of the connections between polymers are unknown.

Composition Ambiguity: The ratio of the components is not known.

HELM2's notation and model allows the scientist to capture the known structure while not over-claiming the extent of knowledge, but the visualization in the editor communicates the information to the scientist who does not want to learn to 'speak HELM'.

HELM: RNA1{R(A)(bP,sP)"Some notes".R(T)P.R(C)P.R(G)} $$$$V2.0

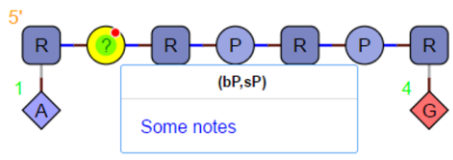


Fig 2: Alternative monomers and annotations.

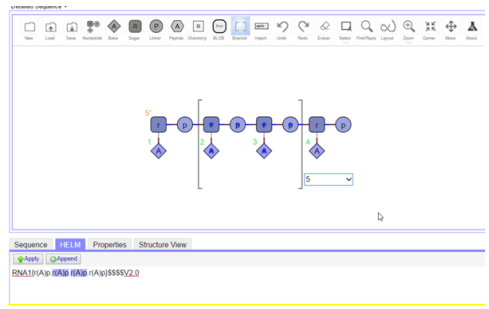


Fig 3: Repeating groups

Conclusion

HELM is now established as the notation of choice for complex biomolecules. With an expanding ecosystem of users and supporting tools, new adopters are well supported.

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 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.

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