Making Open Source Collaboration Work

Over the past few years the biopharmaceutical industry has deployed an arsenal of increasingly complex biomolecular entities in their search for effective therapeutic agents. Portfolios that were once comprised mainly of small molecules and antibodies have expanded to include modalities like antibody-drug conjugates (ADCs); proteins, peptides and antibodies with synthetic, unnatural amino acids; oligonucleotides with unnatural bases, etc. Investigators at Pfizer came to the realization that their scientific informatics infrastructure and methodology was not able to appropriately deal with these types of molecules and, furthermore, no commercial solutions existed. This led to their development of the Hierarchical Editing Language for Macromolecules (HELM). It soon became apparent that many other organizations were facing the same challenge, and that more would be gained from sharing the technology than from the proliferation of disparate, incompatible standards. When the Pistoia Alliance suggested the initiation of a project to release the work into the open source, it was clear that this was the right thing to do. The result: An open standard that is rapidly being adopted throughout the life sciences ecosystem of biopharmaceutical companies, scientific content and solution providers, research institutes, regulatory agencies, etc.

Complex Molecules

Many current biologic entities are significantly more complex than those found in nature, often including synthetic components (e.g. unnatural amino acids) and / or covalently attached modifiers (e.g. ADCs.) Their large size renders them too unwieldy for small molecule informatics tools, while their inclusion of detailed chemistry (as found in ADCs) rules out the use of sequence-based bioinformatics tools. It is this gap that HELM was designed to address.

"Before HELM, our systems had no way of establishing deterministically if one ADC was identical to another ADC, a situation that seemed quite outrageous for this day and age!"

Sergio Rotstein, Director of Business Technology, Pfizer

HELM consists of a notation standard for the in silico representation of complex biomolecule entities, along with the toolkit that leverages this standard to enable the depiction, storage, and retrieval of said entities. It was designed from the ground up to gracefully handle both the sequence and chemical aspects of complex biomolecules, and also to be extensible to new polymer types beyond those that are amino acid- or nucleotide- based.

As awareness of HELM was growing across the life sciences community due to the seminal paper published in the Journal of Chemical Information and Modeling, the Pistoia Alliance surveyed its member organizations on the most important common challenges they were facing. The problem of handling complex biomolecules computationally emerged as a major challenge, along with widespread agreement that HELM-like capabilities would constitute a very viable solution.

From the Alliance's point of view being able to effect the release of HELM into the open source became an obviously ideal path forward. From Pfizer's perspective there were no downsides to this as the company does not consider itself to be either a software house or a standards body. It had nothing to lose and plenty to gain by having the rest of the industry speak its existing biomolecular language, the most obvious upside of which would be the ability to easily share structures with collaborators, vendors, and regulatory agencies. Opening up HELM would also allow the community at large to continue to evolve the standard and tools so Pfizer would no longer have to go it alone.

Figure 1: HELM Editor depiction and notation of an RNA stem loop
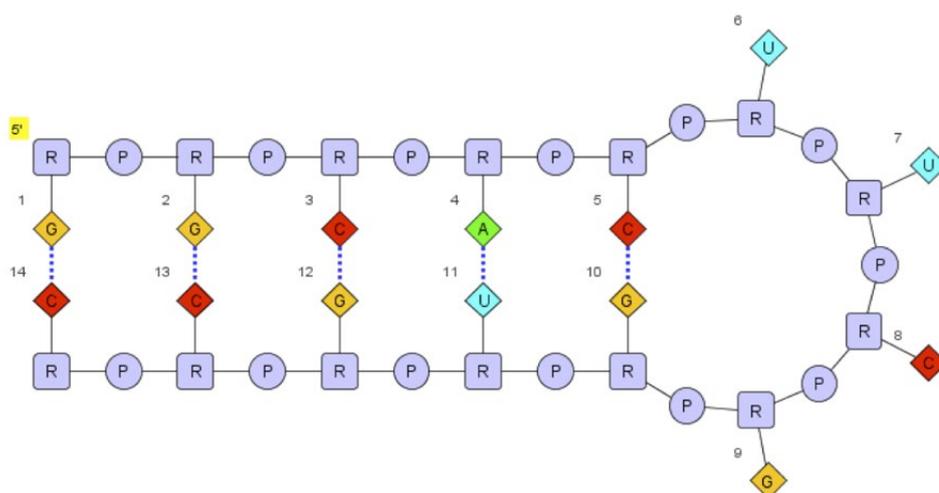
HELM notation
RNA1{R(G)P.R(G)P.R(C)P.R(A)P.R(C)P.R(U)P.R(U)P.R(C)P.R(G)P.R(G)P.R(U)P.R(G)P.R(C)P.R(C)}$$RNA1,RNA1,11:pair-32:pair|RNA1,RNA1,5:pair-38:pair|RNA1,RNA1,14:pair-29:pair|RNA1,RNA1,8:pair-35:pair|RNA1,RNA1,2:pair-41:pair$$

> *"At Pfizer, like everywhere else, we tend to be very busy and resource-limited. It is generally hard enough to find the time to write up a body of work for publication, let alone preparing it for release into the open source. The Pistoia Alliance provided a great framework, support system and set of resources to carry out all the work needed for public release, something that would have otherwise been cost-prohibitive for us to do alone. It was very symbiotic."*

Sergio Rotstein,
Director of Business
Technology, Pfizer

Facilitating Collaboration

Drug discovery R&D is more collaborative today than it has ever been, with pharmaceutical companies, contract research organizations, academia, non-profits and even government agencies working together on many aspects of any given project. The need for standards and methods for the exchange of data among scientists therefore transcends the walls of any individual organization. Further, a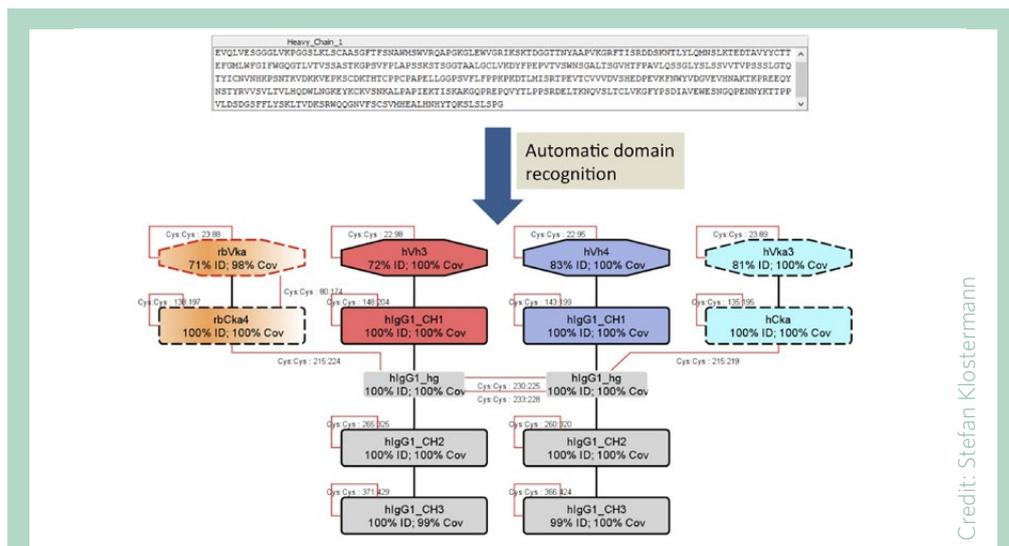s new types of molecular entities make their way downstream towards development, human trials and eventual commercialization, the adoption of these standards and methods by relevant organizations and agencies need to accompany them. To that end, the Pistoia team is collaborating very productively with the FDA to incorporate HELM into the relevant ISO documents that will make it an acceptable structural representation format for the submission of biomolecule data for many international regulatory agencies.

The benefits stack up

An additional benefit of opening HELM up for use by the whole community is the freedom it has given scientific software and content providers to incorporate the standard into their own offerings. Tools that can read and write biomolecule structure data in the format in which it is stored by Pfizer and now many others, vastly facilitate the work of scientists wishing to carry out computational manipulations and calculations with these types of molecules. Life is made easier for users and providers gain more customers, a win-win. Several software vendors now support HELM out of the box and public databases such as ChEMBL have begun to include some of their biomolecule structures in HELM format, making the integration of private and public data using standard off-the-shelf tools substantially easier. Such efforts, along with the anticipated regulatory agency adoption, will go a long way in solidifying HELM as a truly global standard.

Great Minds

Beyond the cost-related savings realized by sharing the burden of development with the open source community, the Pistoia HELM team quickly began to witness the great value of having a large, cross-industry set of sharp brains focusing on the same problem. The intellectual contributions made to the standard by this "team of competitors" has made it a more well-thought-out and well-designed standard, as having a dozen subject matter experts trying to poke holes in a design is far better than only one or two. One example contribution made by the community is Exchangeable HELM (xHELM), which allows users to exchange structures with collaborators that utilize a different monomer database or naming conventions. Roche used the HELM toolkit to develop a terrific antibody editor and registration tool (the HELM Antibody Editor, or HAbE), which they then contributed back into the open source community. These examples really serve to illustrate the power and value of open source approaches and have certainly validated Pfizer's decision to release the technology.

Figure 2: The HELM Antibody Editor (HAbE), developed by Roche, leverages the HELM technology to automate the process of going from sequence to a fully-annotated antibody. Conjugated chemical entities are also supported.

*Credit: Stefan Klostermann*

Conclusion

As we move forward, each organization will continue to have complete control over the technologies they wish to apply to their structural work. Having a common language – HELM – to exchange structural information across, platforms and organizations, will be the key factor in the interoperability of these technologies.

The wide array of companies and individuals that have contributed their time, resources, code and ideas to this endeavor clearly show that there is great value in pre-competitive alliances such as Pistoia, as they offer the best path to carrying out such projects.

More information
info@pistoiaalliance.org
http://www.pistoiaalliance.org

The HELM project website can be found at:
http://www.openhelm.org/
The HELM project team can be contacted
by emailing: **info@openhelm.org**
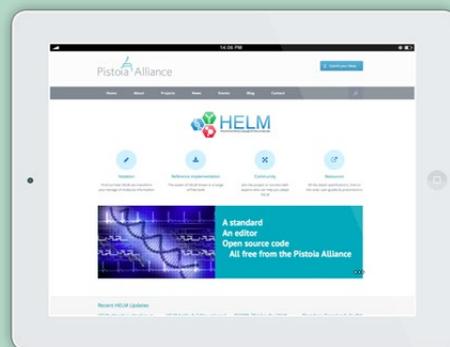
Pistoia Alliance: Lowering barriers to R&D innovation

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.