HELM for the masses

A notation standard for representing Biologics

Roland Knispel, ChemAxon

Domain Lead – Sergio Rotstein, Pfizer

Pistoia Alliance

The Pistoia Mission…

Lower the barriers to innovation

by improving inter-operability of R&D business processes

through pre competitive collaboration

Pistoia Alliance

Q2 2013

Pistoia Alliance Membership

Unmet need

„At present most pharma companies have **no way of storing complete structural machine-readable information** about the macromolecules they prepare or use in their research projects.”

HELM project charter / Dec. 2012

Pistoia Alliance

HELM

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
- Pistoia Alliance to provide *ad interim* governance for the controlled evolution of the standard.

Business Case

- ***Projected Shared Project Costs ~$125,000 [Project Coordinator, Governance, Promotion, Meeting]***
- ***Estimated value for the industry of implementing this project is ~$437m***
  - Driven by productivity savings (e.g. fewer notation queries/resolutions from non-standard language)
  - Standard reference notation for macromolecules in pharma compound registration systems
  - The PA is the best-positioned, formally-established, x-company organization equipped to coordinate this kind of pre-competitive collaboration

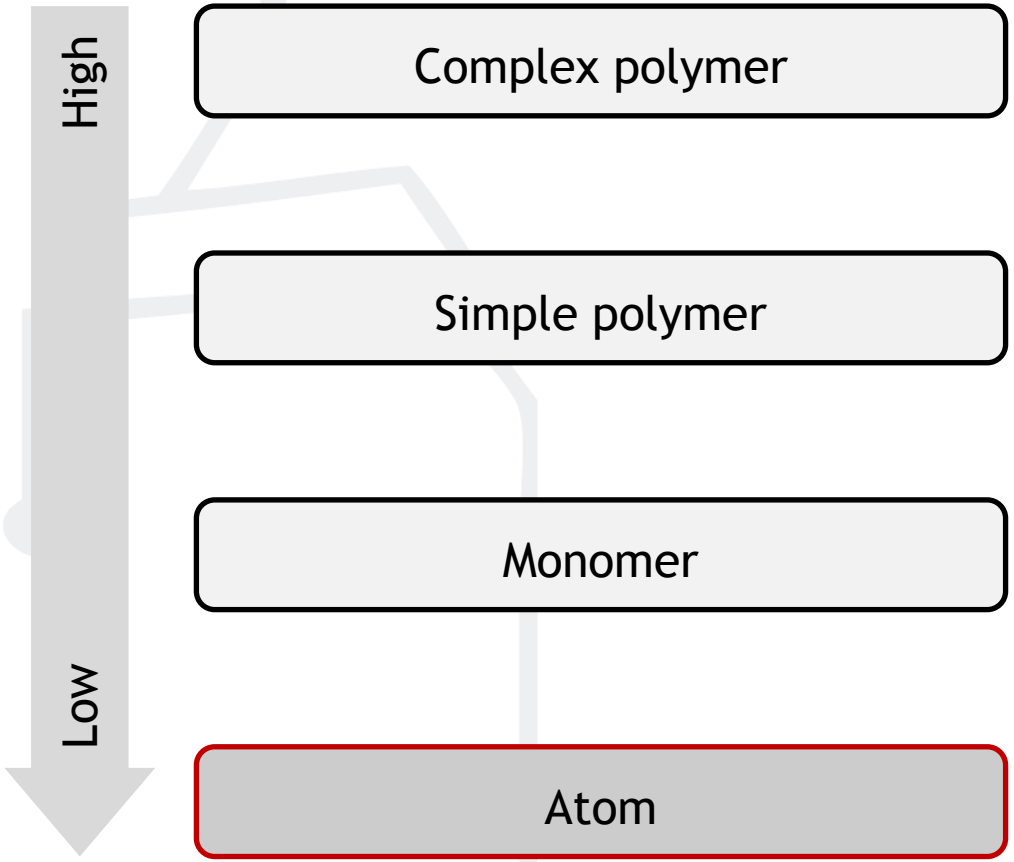
Pistoia Alliance

Hierarchical Editing Language for Macromolecules

- Macromolecules
  - Non-small molecules, Biologics
- Editing Language
  - Notation system
  - Grammar and vocabulary
- Hierarchy
  - Atom
  - Monomer
  - Simple Polymer
  - Complex Polymer

Structure hierarchy

- Higher level components are a combination of lower level components



- Molecules described by a HELM notation consist of atoms and bonds

Structure hierarchy

- Higher level components are a combination of lower level components

High

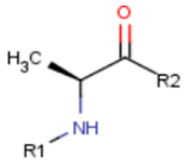
Low

Complex polymer

Simple polymer

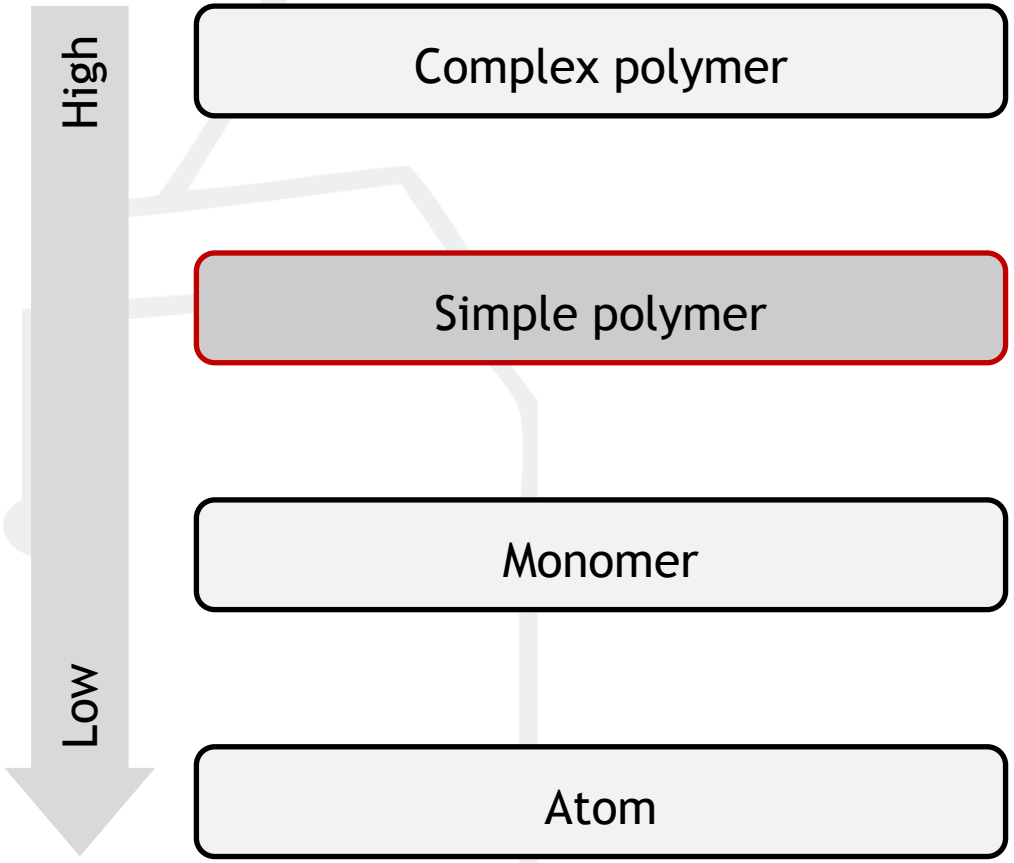
Monomer

Atom

Structure	
SMILES	C[C@H](N[*])C([*])=O |r,$;;;_R1;;_R2;$|
ID	A
Attachment Points	R1-H R2-OH
Natural Analog	A
Polymer Type	PEPTIDE
Monomer Type	Backbone
Name	L-Alanine

Structure hierarchy

- Higher level components are a combination of lower level components

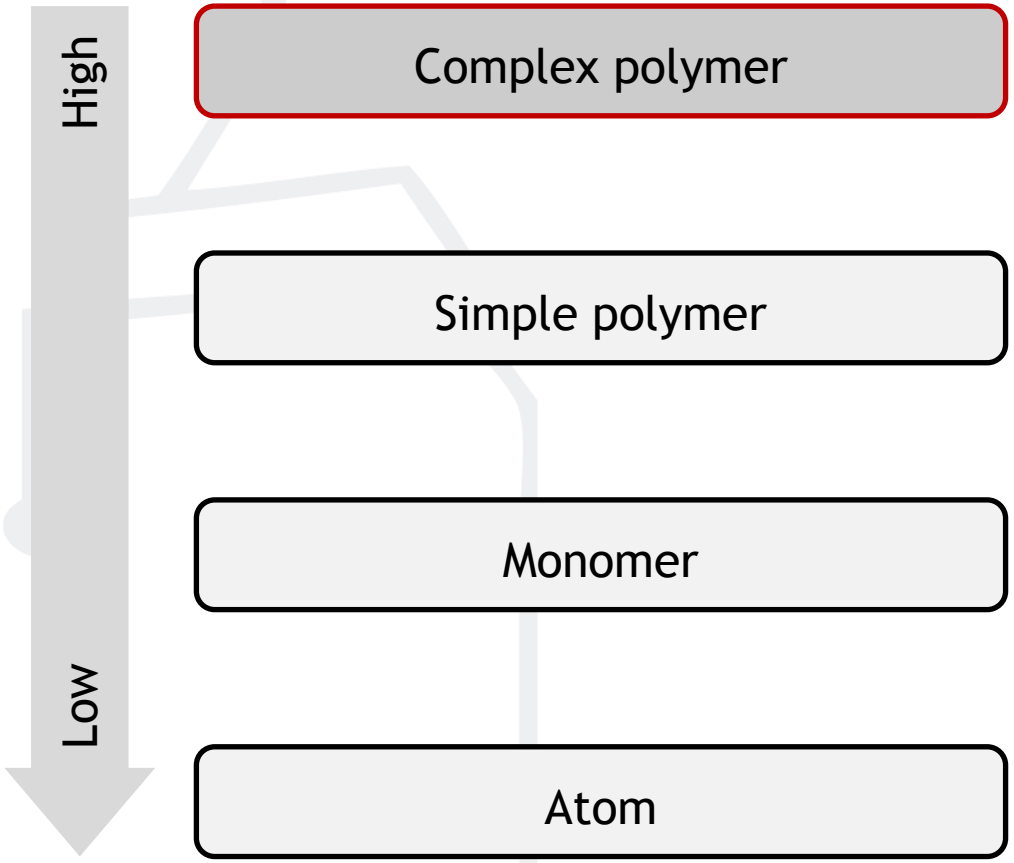


- Linear chains of Monomers or monomer units of a single biological molecule type (e.g. peptide chain, singular nucleic acid strand

Type	Monomer (unit)
Peptide	A - Alanine
Nucleic acid	R(A)P R – Ribose A – Adenine P – Phosphate
Chem	[PEG3] – Pegylation

Structure hierarchy

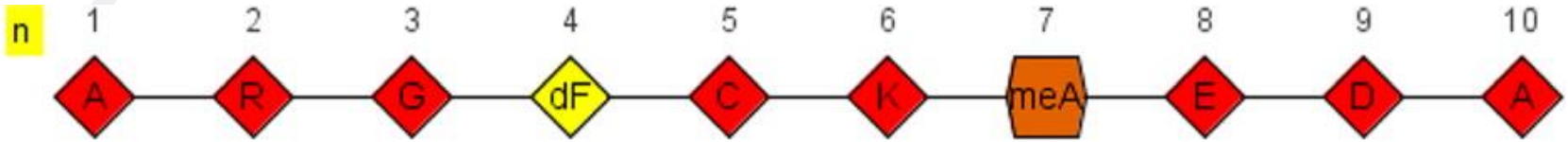
- Higher level components are a combination of lower level components



- Entire chemical structure information of the macromolecule
- May include annotations per Simple polymer

Example: Linear Peptide

A



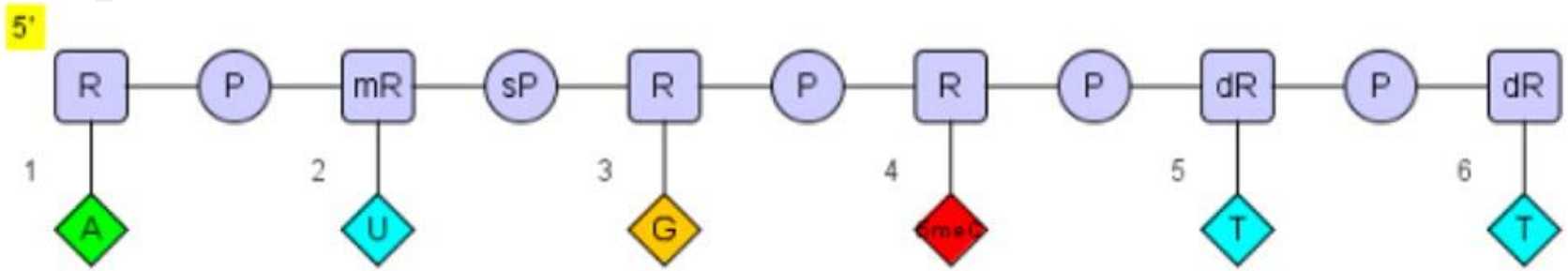
B PEPTIDE1{A.R.G.[dF].C.K.[meA].E.D.A}$$$$

A. Monomer Graph View

B. HELM Notation

Example: Linear Oligonucleotide

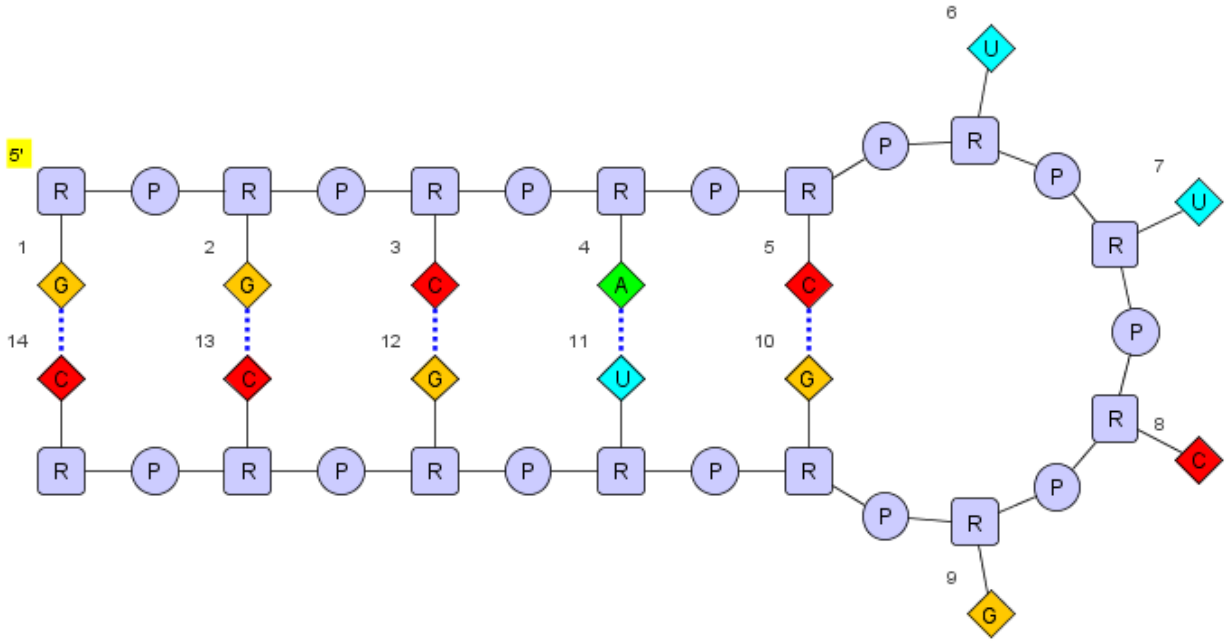
A



B RNA1{R(A)P.[mR](U)[sP].R(G)P.R([5meC])P.[dR](T)P.[dR](T)}$$$$

A. Monomer Graph View B. HELM Notation

Example: Stem-loops (e.g. Tetraloops)

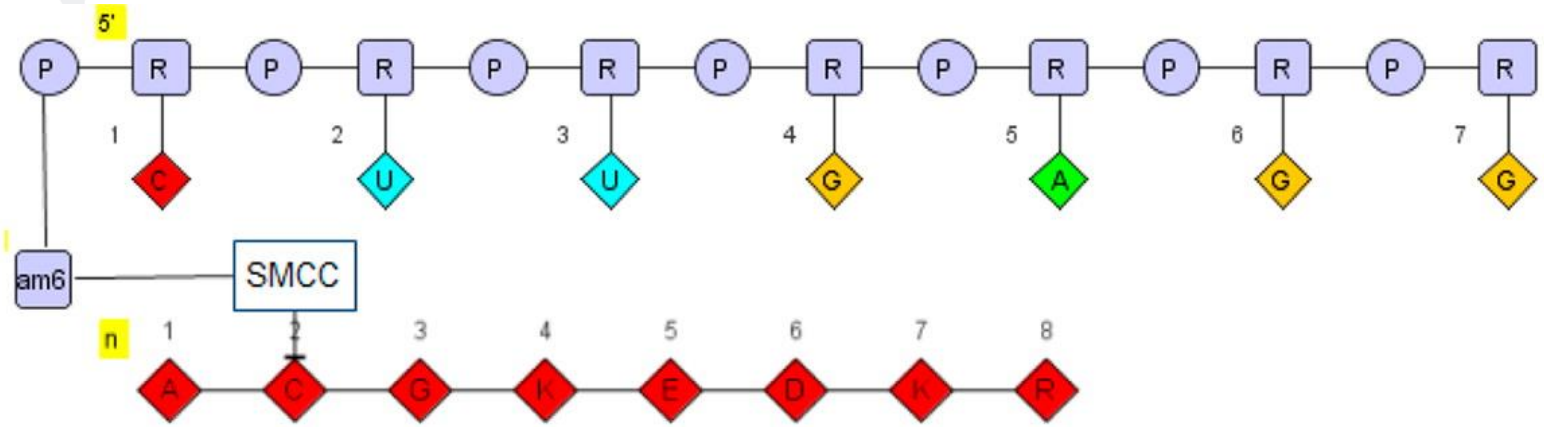


HELM notation:

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)}$$$RNA1,RNA1,11:pair-32:pair|RNA1,RNA1,5:pair-38:pair|RNA1,RNA1,14:pair-29:pair|RNA1,RNA1,8:pair-35:pair|RNA1,RNA1,2:pair-41:pair$$

Example: Oligonucleotide Peptide Conjugate

A

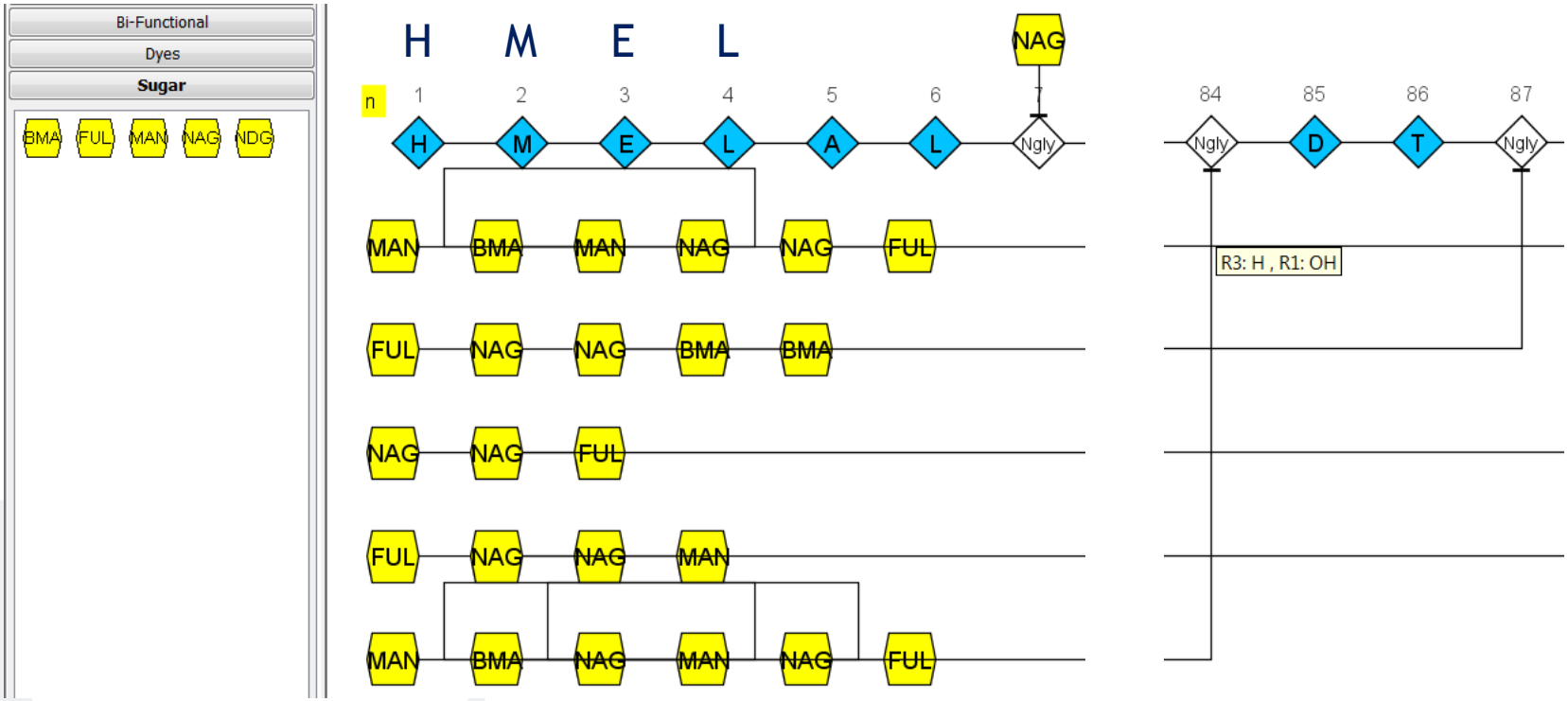


B

RNA1{[am6]P.R(C)P.R(U)P.R(U)P.R(G)P.R(A)P.R(G)P.R(G)}|PEPTIDE1{A.C.G.K.E.D.K.R}|CHEM1{SMCC}$PEPTIDE1,CHEM1,2:R3-1:R2|RNA1,CHEM1,1:R1-1:R1$$$

A. Monomer Graph View B. HELM Notation

Example: Glycoprotein (PDB:3FUS)

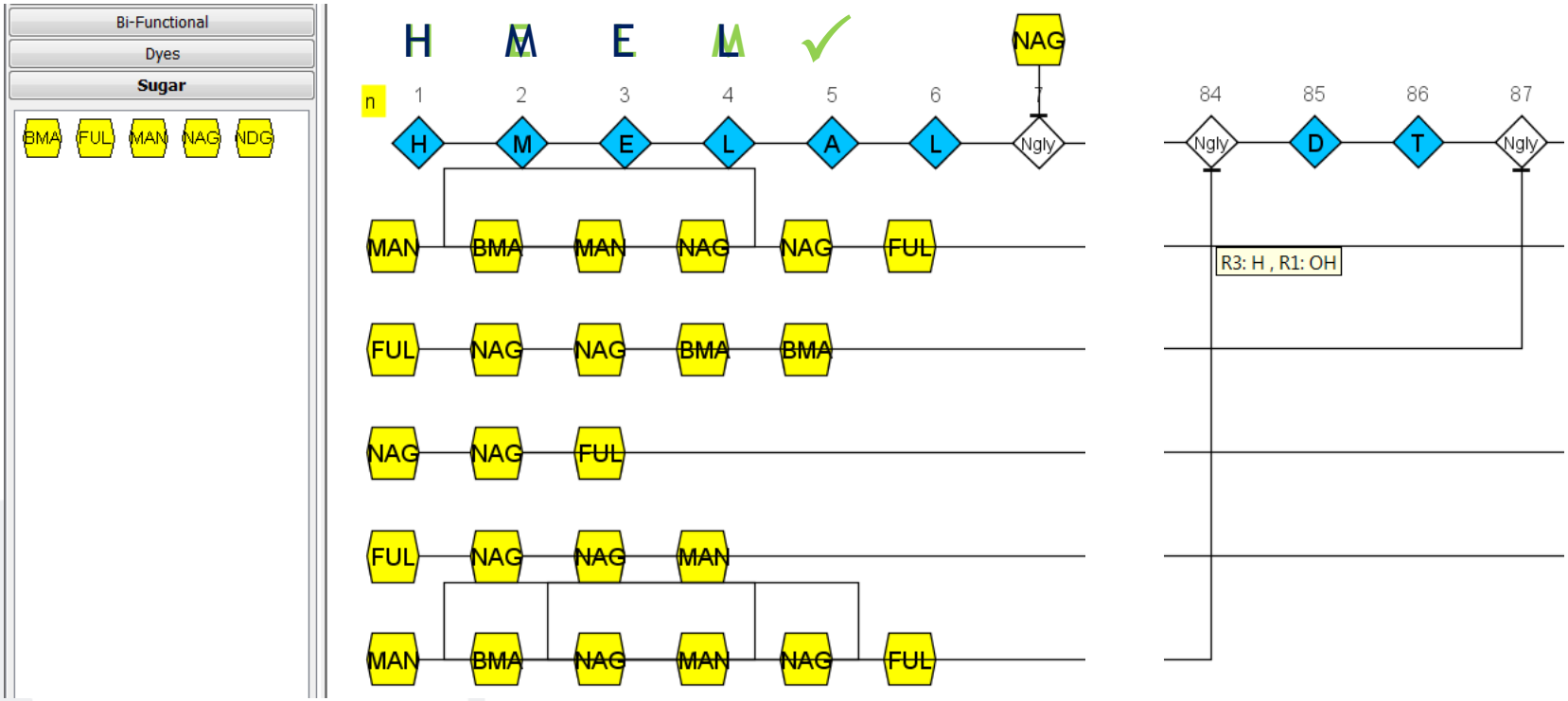


HELM notation:

CHEM1{BMA}|CHEM2{MAN}|CHEM3{MAN}|CHEM4{NAG}|CHEM5{FUL}|CHEM6{NAG}|CHEM7{NAG}|CHEM8{NAG}|CHEM9{FUL}|CHEM10{BMA}|CHEM11{BMA}|CHEM12{NAG}|CHEM13{NAG}|CHEM14{FUL}|CHEM15{NAG}|CHEM16{FUL}|CHEM17{NAG}|CHEM18{MAN}|CHEM19{BMA}|CHEM20{MAN}|CHEM21{MAN}|CHEM22{NAG}|CHEM23{NAG}|CHEM24{FUL}|CHEM25{MAN}|CHEM26{BMA}|CHEM27{NAG}|CHEM28{NAG}|PEPTIDE1{H.M.E.L.A.L.[Ngly].V.T.E.S.F.D.A.W.E.N.T.V.T.E.Q.A.I.E.D.V.W.Q.L.F.E.T.S.I.K.P.C.V.K.L.S.P.L.C.I.G.A.G.H.C.[Ngly].T.S.I.I.Q.E.S.C.D.K.H.Y.W.D.T.I.R.F.R.Y.C.A.P.P.G.Y.A.L.L.R.C.[Ngly].D.T.[Ngly].Y.S.G.F.M.P.K.C.S.K.V.V.V.S.S.C.T.R.M.M.E.T.Q.T.S.T.W.F.G.F.[Ngly].G.T.R.A.E.[Ngly].R.T.Y.I.Y.W.H.G.R.D.[Ngly].R.T.I.I.S.L.N.K.Y.Y.[Ngly].L.T.M.K.C.R.G.A.G.W.C.W.F.G.G.N.W.K.D.A.I.K.E.M.K.Q.T.I.V.K.H.P.R.Y.T.G.T.[Ngly].N.T.D.K.I.[Ngly].L.T.A.P.R.G.G.D.P.E.V.T.F.M.W.T.N.C.R.G.E.F.L.Y.C.K.M.N.W.F.L.N.W.V.E.D.R.D.V.T.N.Q.R.P.K.E.R.H.R.R.N.Y.V.P.C.H.I.R.Q.I.I.N.T.W.H.K.V.G.K.N.V.Y.L.P.P.R.E.G.D.L.T.C.[Ngly].S.T.V.T.S.L.I.A.N.I.D.W.T.D.G.[Ngly].Q.T.[Ngly].I.T.M.S.A.E.V.A.E.L.Y.R.L.E.L.G.D.Y.K.L.V.E.I.T}|CHEM29{NAG}|CHEM30{NAG}|CHEM31{BMA}|CHEM32{MAN}|CHEM33{MAN}|CHEM34{MAN}|CHEM35{NAG}|CHEM36{NAG}|CHEM37{BMA}|CHEM38{MAN}|CHEM39{BMA}|CHEM40{BMA}|CHEM41{MAN}|CHEM42{NAG}|CHEM43{NAG}|CHEM44{BMA}|CHEM45{NAG}|CHEM46{NAG}|CHEM47{FUL}|CHEM48{NDG}|CHEM49{NAG}|CHEM50{MAN}|CHEM51{BMA}|CHEM52{BMA}|CHEM53{NAG}|CHEM54{NDG}|CHEM55{FUL}$CHEM1,CHEM2,1:R2-1:R1|CHEM5,CHEM6,1:R1-1:R3|CHEM16,CHEM15,1:R1-1:R3|PEPTIDE1,CHEM30,52:R3-1:R1|PEPTIDE1,CHEM8,87:R3-1:R3|CHEM19,CHEM21,1:R3-1:R1|CHEM8,CHEM7,1:R1-1:R2|CHEM22,CHEM23,1:R1-1:R2|PEPTIDE1,CHEM28,292:R3-1:R1|PEPTIDE1,CHEM43,273:R3-1:R1|PEPTIDE1,CHEM46,146:R3-1:R1|PEPTIDE1,CHEM15,135:R3-1:R1|CHEM10,CHEM8,1:R1-1:R2|PEPTIDE1,CHEM36,289:R3-1:R1|CHEM51,CHEM50,1:R2-1:R1|CHEM39,CHEM41,1:R3-1:R1|PEPTIDE1,CHEM12,124:R3-1:R1|PEPTIDE1,CHEM29,7:R3-1:R1|CHEM39,CHEM38,1:R1-1:R2|CHEM20,CHEM19,1:R1-1:R2|CHEM17,CHEM15,1:R1-1:R2|CHEM12,CHEM13,1:R2-1:R1|CHEM46,CHEM47,1:R3-1:R1|CHEM32,CHEM31,1:R1-1:R3|CHEM52,CHEM51,1:R3-1:R1|PEPTIDE1,CHEM23,84:R3-1:R1|CHEM28,CHEM27,1:R2-1:R1|CHEM4,CHEM1,1:R2-1:R1|CHEM31,CHEM33,1:R2-1:R1|PEPTIDE1,CHEM54,190:R3-1:R1|CHEM34,CHEM33,1:R1-1:R2|CHEM18,CHEM17,1:R1-1:R2|CHEM31,CHEM35,1:R1-1:R2|CHEM9,CHEM7,1:R1-1:R3|CHEM26,CHEM27,1:R1-1:R2|CHEM54,CHEM53,1:R2-1:R1|CHEM6,CHEM4,1:R2-1:R1|CHEM1,CHEM3,1:R3-1:R1|CHEM25,CHEM26,1:R1-1:R2|CHEM54,CHEM55,1:R3-1:R1|CHEM43,CHEM42,1:R2-1:R1|CHEM40,CHEM37,1:R2-1:R1|CHEM46,CHEM45,1:R2-1:R1|CHEM12,CHEM14,1:R3-1:R1|CHEM42,CHEM40,1:R2-1:R1|PEPTIDE1,CHEM6,118:R3-1:R1|CHEM24,CHEM23,1:R1-1:R3|CHEM22,CHEM19,1:R2-1:R1|CHEM53,CHEM52,1:R2-1:R1|CHEM48,CHEM49,1:R2-1:R1|PEPTIDE1,CHEM48,184:R3-1:R1|CHEM45,CHEM44,1:R2-1:R1|CHEM37,CHEM38,1:R2-1:R1|CHEM11,CHEM10,1:R1-1:R2|CHEM36,CHEM35,1:R2-1:R1$$$

Pistoia Alliance

Example: Glycoprotein (PDB:3FUS)



HELM notation:

CHEM1{BMA}|CHEM2{MAN}|CHEM3{MAN}|CHEM4{NAG}|CHEM5{FUL}|CHEM6{NAG}|CHEM7{NAG}|CHEM8{NAG}|CHEM9{FUL}|CHEM10{BMA}|CHEM11{BMA}|CHEM12{NAG}|CHEM13{NAG}|CHEM14{FUL}|CHEM15{NAG}|CHEM16{FUL}|CHEM17{NAG}|CHEM18{MAN}|CHEM19{BMA}|CHEM20{MAN}|CHEM21{MAN}|CHEM22{NAG}|CHEM23{NAG}|CHEM24{FUL}|CHEM25{MAN}|CHEM26{BMA}|CHEM27{NAG}|CHEM28{NAG}|PEPTIDE1{H.M.E.L.A.L.[Ngly].V.T.E.S.F.D.A.W.E.N.T.V.T.E.Q.A.I.E.D.V.W.Q.L.F.E.T.S.I.K.P.C.V.K.L.S.P.L.C.I.G.A.G.H.C.[Ngly].T.S.I.I.Q.E.S.C.D.K.H.Y.W.D.T.I.R.F.R.Y.C.A.P.P.G.Y.A.L.L.R.C.[Ngly].D.T.[Ngly].Y.S.G.F.M.P.K.C.S.K.V.V.V.S.S.C.T.R.M.M.E.T.Q.T.S.T.W.F.G.F.[Ngly].G.T.R.A.E.[Ngly].R.T.Y.I.Y.W.H.G.R.D.[Ngly].R.T.I.I.S.L.N.K.Y.Y.[Ngly].L.T.M.K.C.R.G.A.G.W.C.W.F.G.G.N.W.K.D.A.I.K.E.M.K.Q.T.I.V.K.H.P.R.Y.T.G.T.[Ngly].N.T.D.K.I.[Ngly].L.T.A.P.R.G.G.D.P.E.V.T.F.M.W.T.N.C.R.G.E.F.L.Y.C.K.M.N.W.F.L.N.W.V.E.D.R.D.V.T.N.Q.R.P.K.E.R.H.R.R.N.Y.V.P.C.H.I.R.Q.I.I.N.T.W.H.K.V.G.K.N.V.Y.L.P.P.R.E.G.D.L.T.C.[Ngly].S.T.V.T.S.L.I.A.N.I.D.W.T.D.G.[Ngly].Q.T.[Ngly].I.T.M.S.A.E.V.A.E.L.Y.R.L.E.L.G.D.Y.K.L.V.E.I.T}|CHEM29{NAG}|CHEM30{NAG}|CHEM31{BMA}|CHEM32{MAN}|CHEM33{MAN}|CHEM34{MAN}|CHEM35{NAG}|CHEM36{NAG}|CHEM37{BMA}|CHEM38{MAN}|CHEM39{BMA}|CHEM40{BMA}|CHEM41{MAN}|CHEM42{NAG}|CHEM43{NAG}|CHEM44{BMA}|CHEM45{NAG}|CHEM46{NAG}|CHEM47{FUL}|CHEM48{NDG}|CHEM49{NAG}|CHEM50{MAN}|CHEM51{BMA}|CHEM52{BMA}|CHEM53{NAG}|CHEM54{NDG}|CHEM55{FUL}$CHEM1,CHEM2,1:R2-1:R1|CHEM5,CHEM6,1:R1-1:R3|CHEM16,CHEM15,1:R1-1:R3|PEPTIDE1,CHEM30,52:R3-1:R1|PEPTIDE1,CHEM8,87:R3-1:R3|CHEM19,CHEM21,1:R3-1:R1|CHEM8,CHEM7,1:R1-1:R2|CHEM22,CHEM23,1:R1-1:R2|PEPTIDE1,CHEM28,292:R3-1:R1|PEPTIDE1,CHEM43,273:R3-1:R1|PEPTIDE1,CHEM46,146:R3-1:R1|PEPTIDE1,CHEM15,135:R3-1:R1|CHEM10,CHEM8,1:R1-1:R2|PEPTIDE1,CHEM36,289:R3-1:R1|CHEM51,CHEM50,1:R2-1:R1|CHEM39,CHEM41,1:R3-1:R1|PEPTIDE1,CHEM12,124:R3-1:R1|PEPTIDE1,CHEM29,7:R3-1:R1|CHEM39,CHEM38,1:R1-1:R2|CHEM20,CHEM19,1:R1-1:R2|CHEM17,CHEM15,1:R1-1:R2|CHEM12,CHEM13,1:R2-1:R1|CHEM46,CHEM47,1:R3-1:R1|CHEM32,CHEM31,1:R1-1:R3|CHEM52,CHEM51,1:R3-1:R1|PEPTIDE1,CHEM23,84:R3-1:R1|CHEM28,CHEM27,1:R2-1:R1|CHEM4,CHEM1,1:R2-1:R1|CHEM31,CHEM33,1:R2-1:R1|PEPTIDE1,CHEM54,190:R3-1:R1|CHEM34,CHEM33,1:R1-1:R2|CHEM18,CHEM17,1:R1-1:R2|CHEM31,CHEM35,1:R1-1:R2|CHEM9,CHEM7,1:R1-1:R3|CHEM26,CHEM27,1:R1-1:R2|CHEM54,CHEM53,1:R2-1:R1|CHEM6,CHEM4,1:R2-1:R1|CHEM1,CHEM3,1:R3-1:R1|CHEM25,CHEM26,1:R1-1:R2|CHEM54,CHEM55,1:R3-1:R1|CHEM43,CHEM42,1:R2-1:R1|CHEM40,CHEM37,1:R2-1:R1|CHEM46,CHEM45,1:R2-1:R1|CHEM12,CHEM14,1:R3-1:R1|CHEM42,CHEM40,1:R2-1:R1|PEPTIDE1,CHEM6,118:R3-1:R1|CHEM24,CHEM23,1:R1-1:R3|CHEM22,CHEM19,1:R2-1:R1|CHEM53,CHEM52,1:R2-1:R1|CHEM48,CHEM49,1:R2-1:R1|PEPTIDE1,CHEM48,184:R3-1:R1|CHEM45,CHEM44,1:R2-1:R1|CHEM37,CHEM38,1:R2-1:R1|CHEM11,CHEM10,1:R1-1:R2|CHEM36,CHEM35,1:R2-1:R1$$$

Pistoia Alliance

Does HELM meet requirements?

"Standard reference notation for macromolecules in pharma compound registration systems"

Requirements:

- ❑ Use case coverage
- ❑ Specification and Documentation
- ❑ Interoperability
- ❑ Registrability
- ❑ Long-term stability
- ❑ Adoption

Pistoia Alliance

Use case coverage

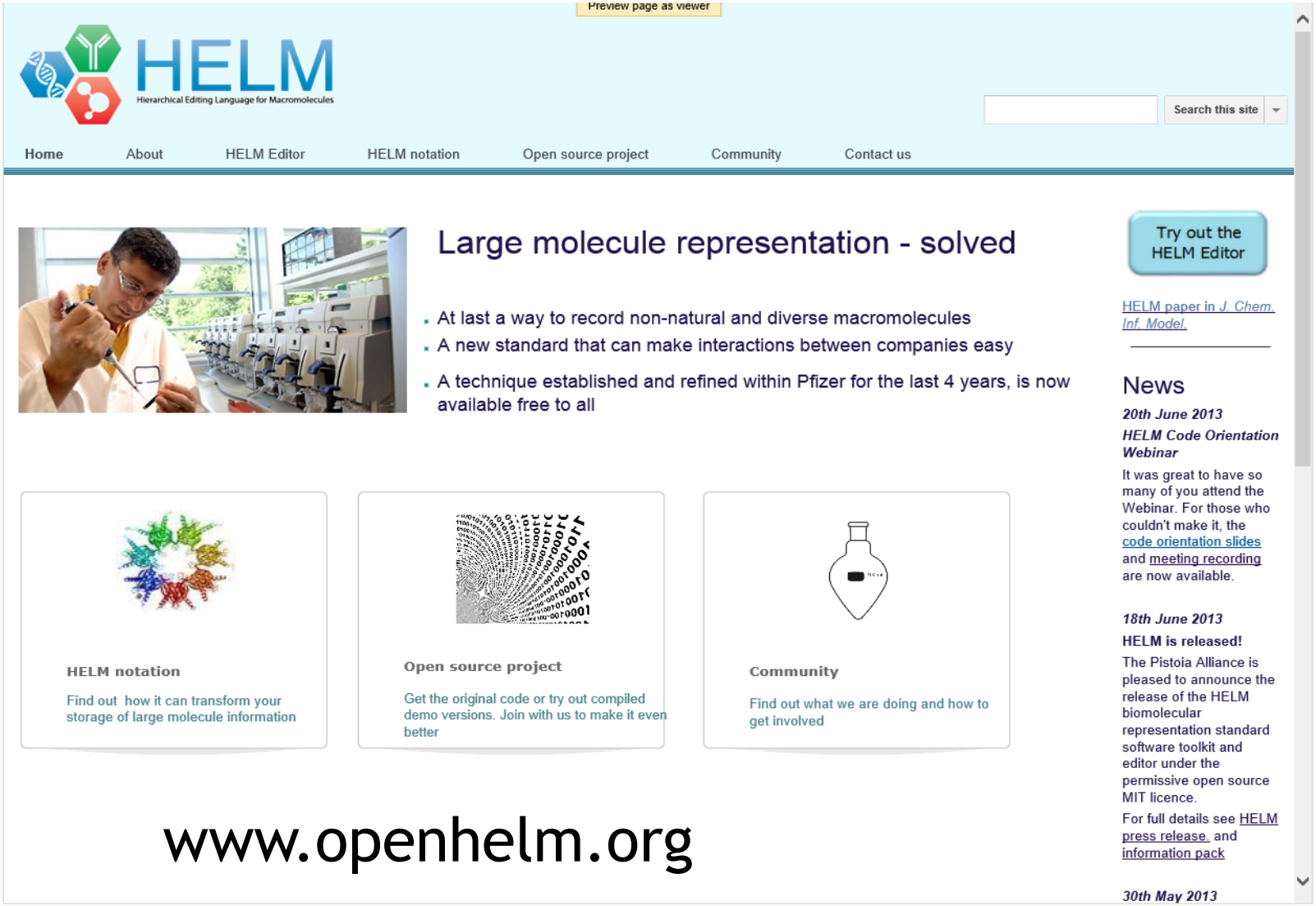
Unmodified and modified biological molecules

- ✓ Nucleic acids – dsDNA, siRNA, locked DNA/RNA...
- ✓ Proteins – Antibodies, proteins with PTMs
- ○ Glycoproteins

Conjugated molecules

- ✓ Covalently linked peptide-RNA hybrid molecules
- ○ Chemically crosslinked molecules – ADCs

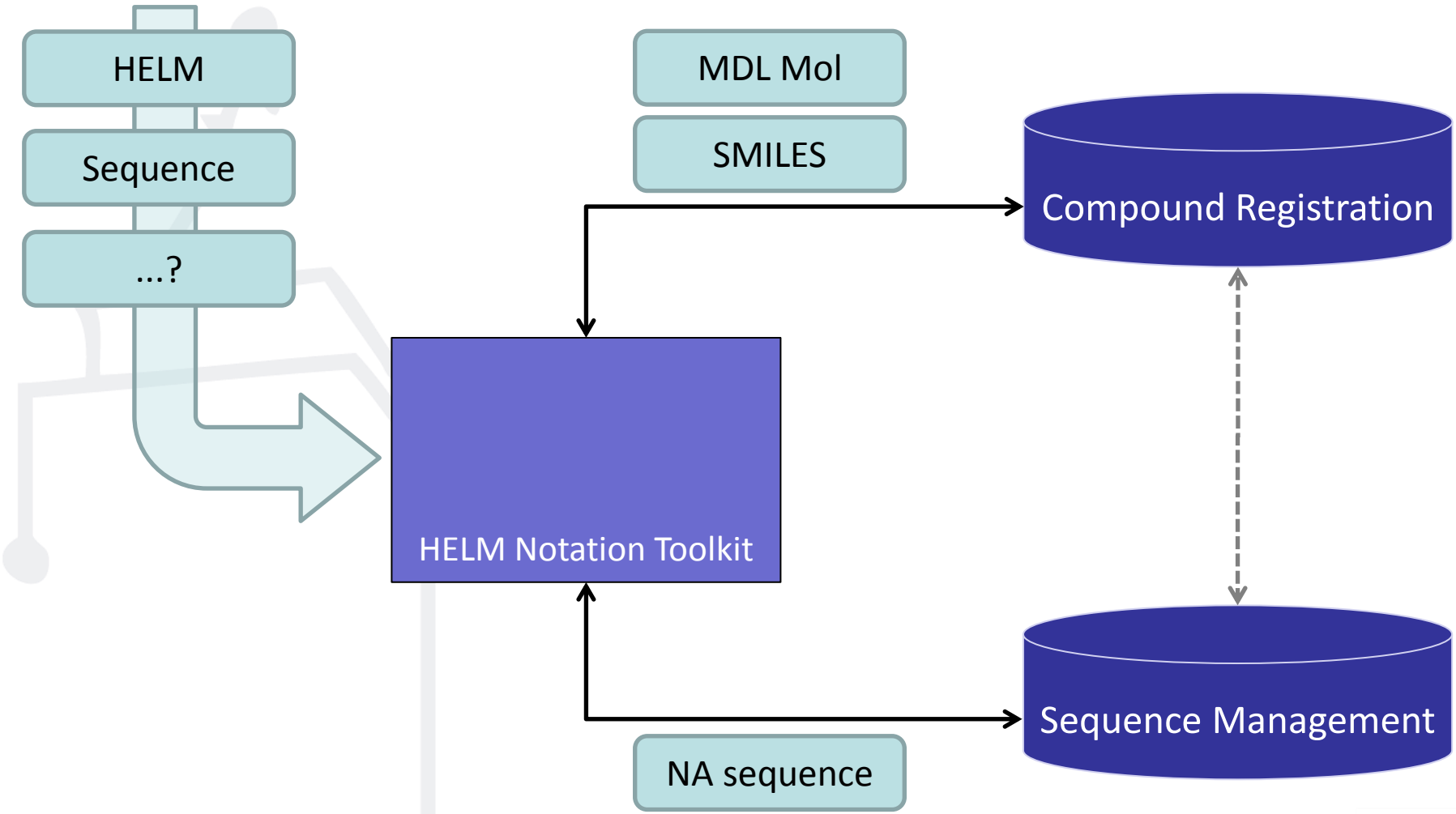
Specification and Documentation



www.openhelm.org

Pistoia Alliance

Interoperability



Central HELM monomer administration

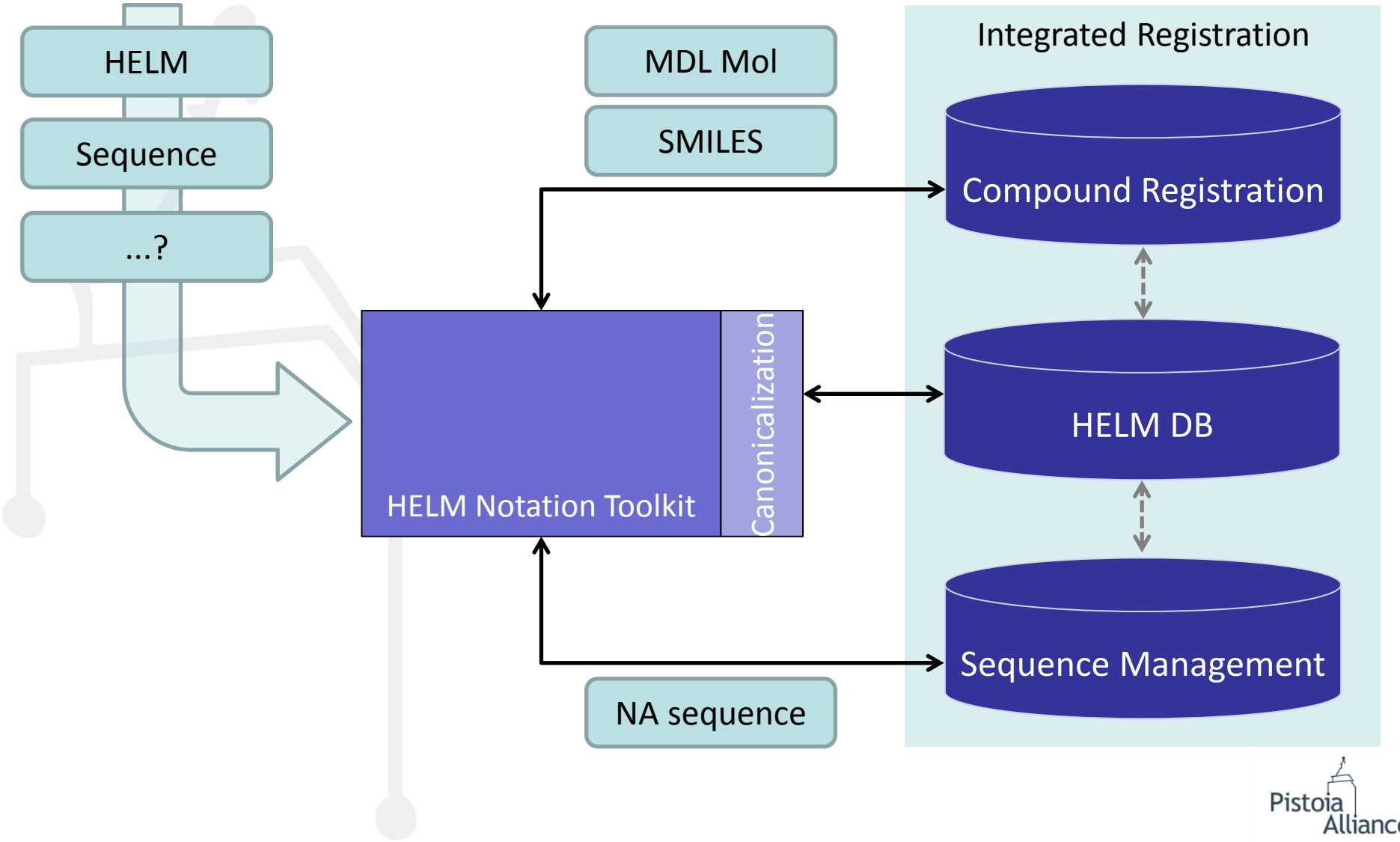
Pistoia Alliance

Interoperability

- Available tools currently export a HELM notation only
  - To regenerate the entire atomic structure the monomer information has to be passed on additionally (current PA project activity)
- Harmonization of differing monomer dictionaries between partnering organizations
  - To avoid ambiguity of macromolecule representation for registration

Pistoia Alliance

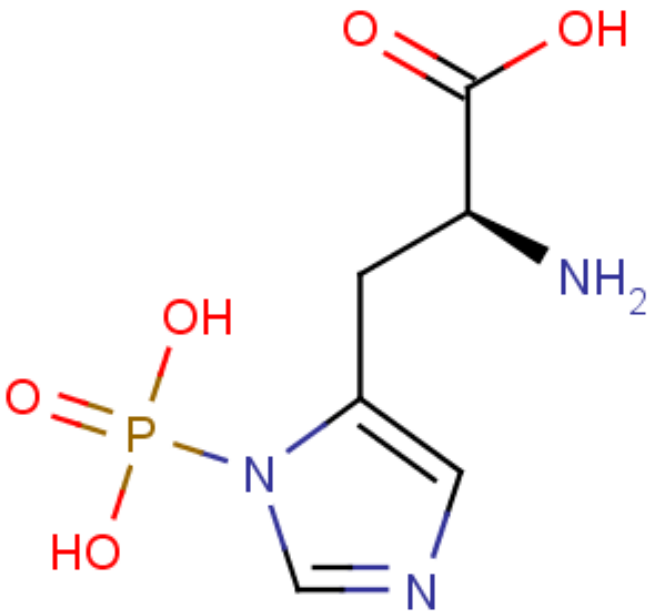
Registrability



Pistoia Alliance

Registrability

- Consistent automated registration of new HELM monomers is challenging



1-Phospho-histidine

Histidine + Phosphate

or

1-Phospho-histidine

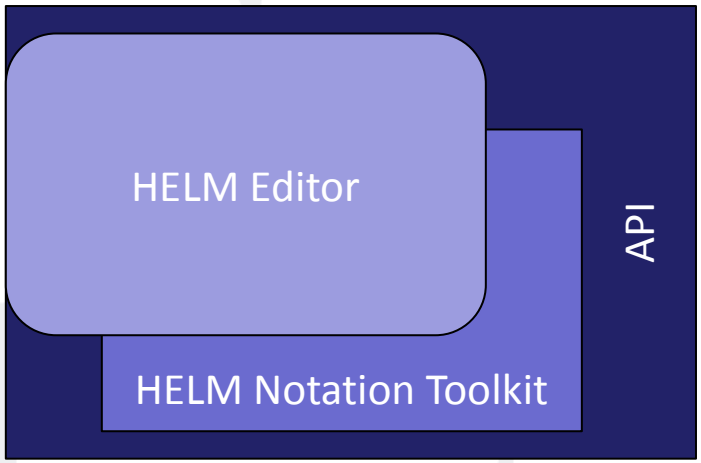
?

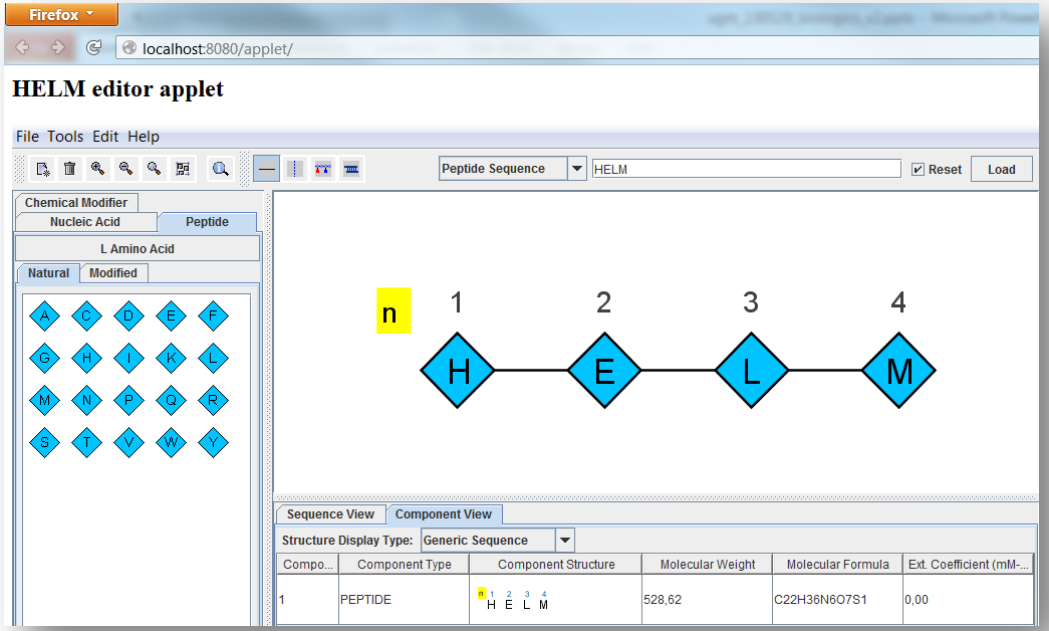
Long-term stability

- ✓ PA to „sign-post“ the HELM standard and foster adoption throughout PA community
- ✓ PA to provide *ad interim* governance for controlled evolution of the standard
- ✓ Release of HELM tools to the Open Source public domain facilitates public discourse
- ○ Widespread adoption

Pistoia Alliance

Which tools are available?





https://github.com/PistoiaHELM

Pistoia Alliance

Adoption

- ✓ Vendors
- ✓ Pharma-/Polymer chemistry companies
- ○ Content providers
- ○ Academia
- ○ State agencies

Open HELM Standard is young, requires thorough assessment

Pistoia Alliance

HELM is already meeting most of the requirements for broad adoption

With your help we can make HELM the “language of the macromolecules”

Requirements:

- ☑ Use case coverage
- ☑ Specification and Documentation
- ☑ Interoperability
- ☑ Registrability
- ☑ Long-term stability
- ☐ Adoption – over to you!

Pistoia Alliance

HELM

For more information please visit
http://openhelm.org

or contact us at info@openhelm.org

Pistoia Alliance