3DEXPERIENCE®

Support for HELM In BIOVIA Products

Moises Hassan, Matthew Sage

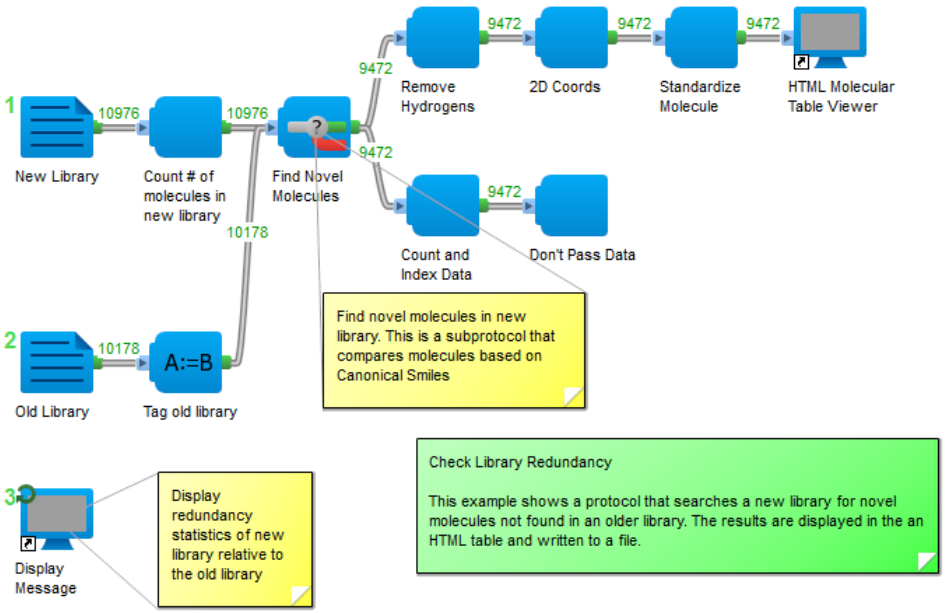
Feb 1st , 2016

DASSAULT SYSTEMES | The 3DEXPERIENCE® Company

Support for HELM in BIOVIA Products

- Pipeline Pilot
  - HELM components
  - HELM Editor integration
- Draw Sketcher
  - Import and Export HELM Strings and XHELM
  - HELM Editor integration
- Direct Chemistry Cartridge
  - Registration, import and export HELM and XHELM
- Biological Registration
  - HELM Editor integration
  - Register HELM strings as entities
- ScienceCloud
  - Centralized monomer libraries

Pipeline Pilot Chemistry Collection



- The Chemistry Collection offers a comprehensive suite of readers, writers, viewers, molecular property calculators, filters, and manipulators.
- This collection of modular components extends the standard capabilities of BIOVIA Pipeline Pilot to include compound processing and cheminformatics analysis

HELM Components in the Chemistry Collection

Readers

HELM Reader

XHELM Reader

Writers

HELM Writer

Converters

Molecule to HELM

Molecule to XHELM

Molecule from HELM

Molecule from XHELM

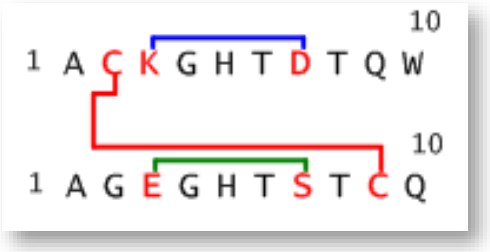
Utilities

HELM Monomer to SCSR Template

HELM Monomer from SCSR Template

A Quick Note on SCSR

- SCSR (Self-Contained Sequence Representation)
  - Chen et al, *J. Chem. Inf. Model.*, **2011**, *51* (9), pp 2186–2208
  - Extensions to V3000 MOL format to represent biological sequences using residue templates
  - Can represent chemical and biological features together
  - Modifications and small molecules attached to the sequence are represented using explicit chemistry
  - Supports Substructure Searching (SSS)



Example of mol file with SCSR

ACF amino acid sequence

Representation in mol file with SCSR

```
 ACCLDraw01271614482D

  0  0  0     0  0            999 V3000
M  V30 BEGIN CTAB
M  V30 COUNTS 3 2 1 0 1
M  V30 BEGIN ATOM
M  V30 1 Ala 6.0313 -3.75 0 0 CLASS=AA ATTCHORD=(2 3 Br) SEQID=1
M  V30 3 Cys 7.1887 -3.7472 0 0 CLASS=AA ATTCHORD=(4 1 Al 2 Br) SEQID=2
M  V30 2 Phe 8.3462 -3.7472 0 0 CLASS=AA ATTCHORD=(2 3 Al) SEQID=3
M  V30 END ATOM
M  V30 BEGIN BOND
M  V30 1 1 3 1
M  V30 2 1 3 2
M  V30 END BOND
M  V30 END CTAB
M  V30 BEGIN TEMPLATE
```

```
M  V30 TEMPLATE 1 AA/Ala/A/
M  V30 BEGIN CTAB
M  V30 COUNTS 7 6 3 0 1
.....
.....
M  V30 END CTAB
M  V30 TEMPLATE 2 AA/Cys/C/
M  V30 BEGIN CTAB
.....
.....
M  V30 END CTAB
M  V30 TEMPLATE 3 AA/Phe/F/
M  V30 BEGIN CTAB
.....
.....
M  V30 END CTAB
M  V30 END TEMPLATE
M  END
```

HELM Reader Component

Create SCSR sequences or fully explicit macromolecules

Files with HELM monomers, SCSR templates and mapping between them

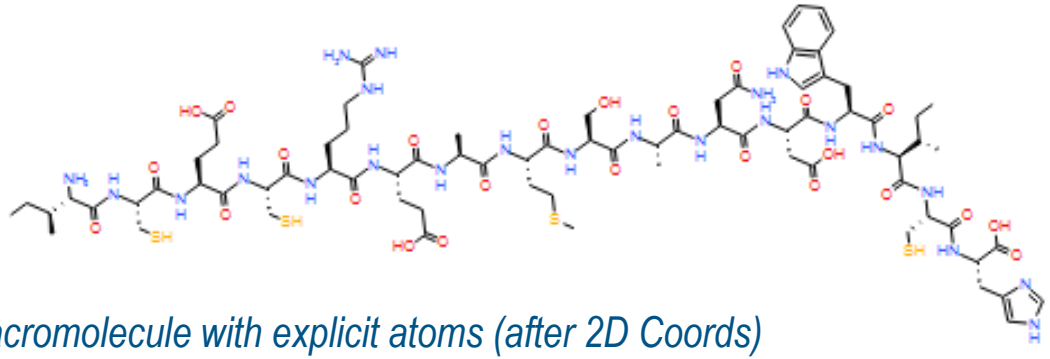
Parameters	
Source	data\Examples\HELM Strings\ICECREAMSANDWICH.helm
Monomers Source	data\HELM\HELMMonomers.sd
Structure to Create	Molecule with SCSR
Additional Options	
Global Templates	data\HELM\Draw41Templates.mol
Monomer Map	data\HELM\HELM_SCSR_MonomerMap.txt
Create SCSR Templates for Unknown Monomers	True
Orient Residues	True

HELM Reader Component

PEPTIDE1{I.C.E.C.R.E.A.M.S.A.N.D.W.I.C.H}$$$$


```
                  10
I C E C R E A M S A  N D W I C H
```

SCSR Sequence

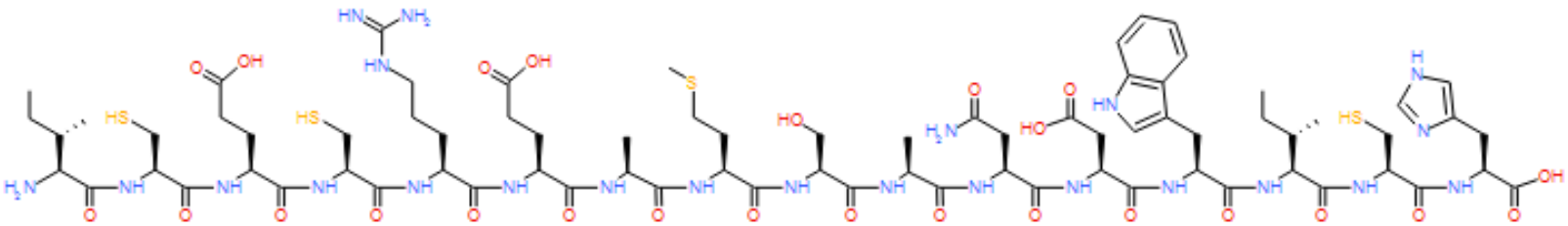


Macromolecule with explicit atoms (after 2D Coords)

HELM Reader Component

PEPTIDE1{I.C.E.C.R.E.A.M.S.A.N.D.W.I.C.H}$$$$

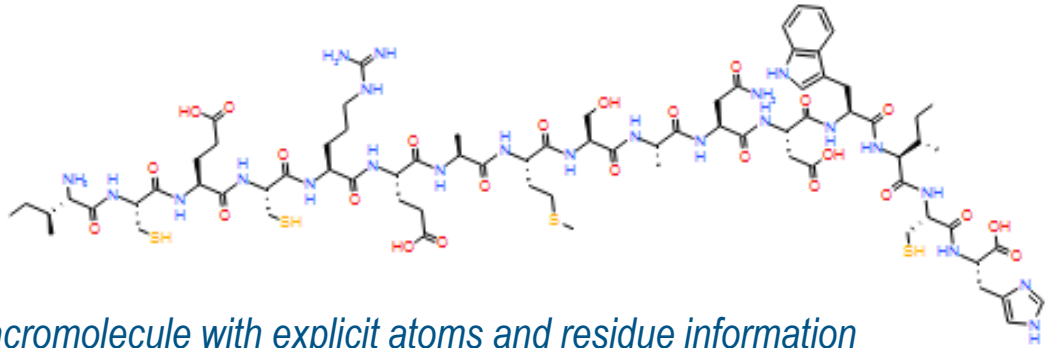
Improved 2D layout of macromolecules in latest release (2016)



HELM Writer Component

10
I C E C R E A M S A N D W I C H

SCSR Sequence

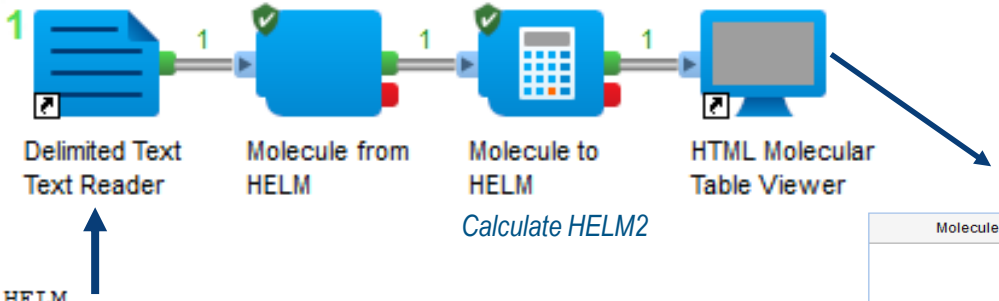


Macromolecule with explicit atoms and residue information

Writes HELM string from SCSR sequences or macromolecules

PEPTIDE1{I.C.E.C.R.E.A.M.S.A.N.D.W.I.C.H}$$$$

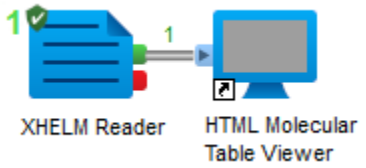
HELM Converters Components



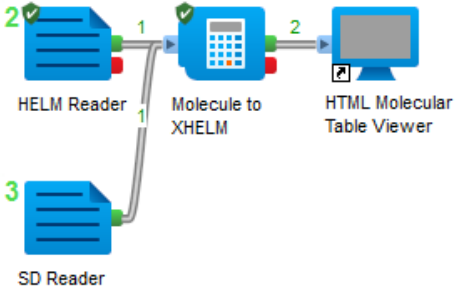
Molecule	HELM	HELM2
	PEPTIDE1{I.C.E.C.R.E.A.M.S.A.N.D.W.I.C.H}$$$$	PEPTIDE1{I.C.E.C.R.E.A.M.S.A.N.D.W.I.C.H}$$$$

- Conversion of molecules to and from HELM strings
- HELM string is a calculable property in PilotScript: calculate("HELM")

XHELM Support



Read XHELM files



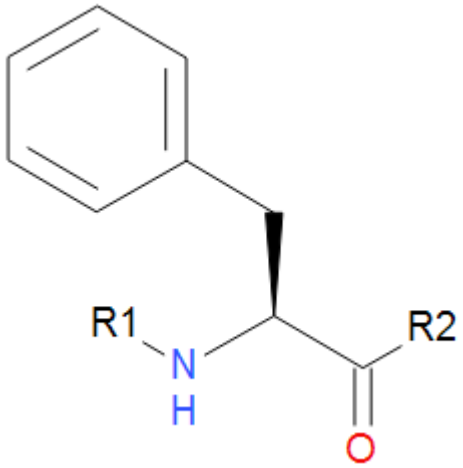
Create XHELM files from HELM Strings or SCSR mol files

XML-based format containing the HELM string and all required monomers

```
<?xml version="1.0" encoding="UTF-8"?>
<Xhelm>
  <HelmNotation>PEPTIDE1{A}$$$$</HelmNotation>
  <Monomers>
    <Monomer>
      <MonomerID>A</MonomerID>
      <MonomerSmiles>C[C@H](N[*])C([*])=O |$;;;_R1;;_R2;$|</MonomerSmiles>
      <MonomerMolFile>H4sIAAAAAAAAAKWSuw7CMAxF93yFJVhrOc57poipBXVgZ2RhYOD7SYI
      <MonomerType>Backbone</MonomerType>
      <PolymerType>PEPTIDE</PolymerType>
      <NaturalAnalog>A</NaturalAnalog>
      <MonomerName>Alanine</MonomerName>
      <Attachments>
        <Attachment>
          <AttachmentID>R2-OH</AttachmentID>
          <AttachmentLabel>R2</AttachmentLabel>
          <CapGroupName>OH</CapGroupName>
          <CapGroupSmiles>O[*] |$;_R2$|</CapGroupSmiles>
        </Attachment>
        <Attachment>
          <AttachmentID>R1-H</AttachmentID>
          <AttachmentLabel>R1</AttachmentLabel>
          <CapGroupName>H</CapGroupName>
          <CapGroupSmiles>[*][H] |$_R1;$|</CapGroupSmiles>
        </Attachment>
      </Attachments>
    </Monomer>
  </Monomers>
</Xhelm>
```

Interconversion between HELM Monomers and SCSR Templates

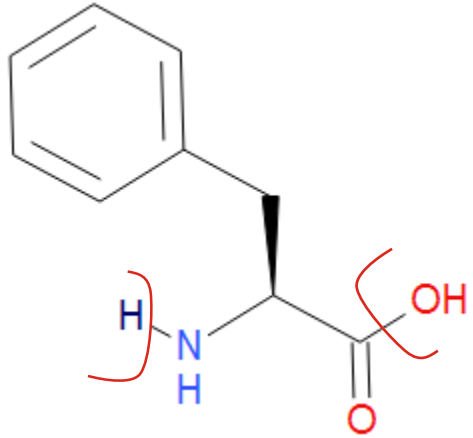
SCSR Templates for most monomers in the HELM Editor added in release 2016



HELM monomer
With attachment points

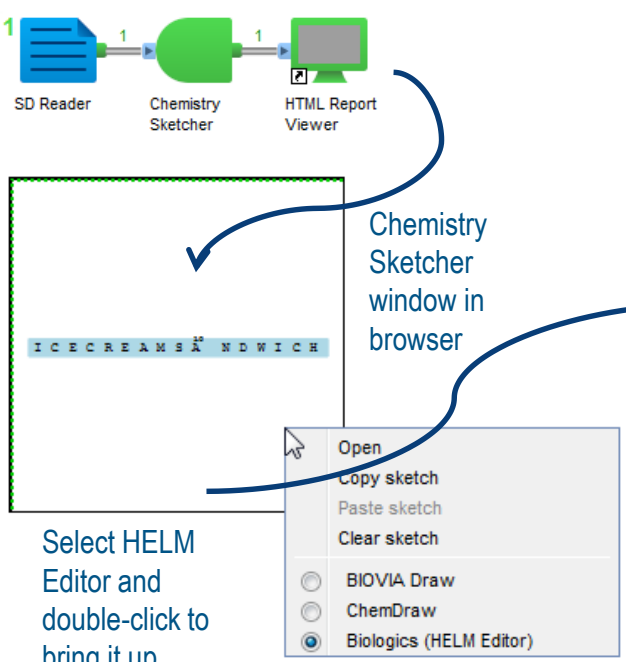
HELM Monomer from SCSR Template

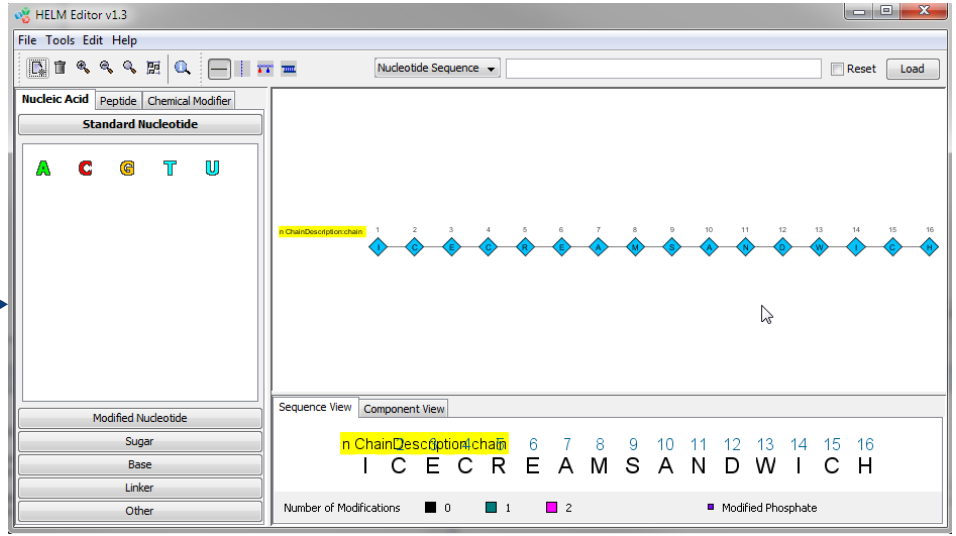
HELM Monomer to SCSR Template



SCSR Template
with leaving groups

Support for HELM Editor in Pipeline Pilot Chemistry Sketcher





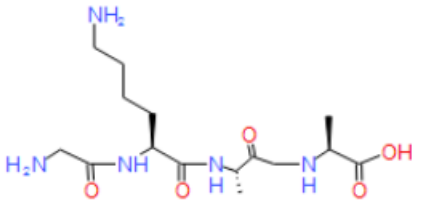
- Starts HELM Editor as JNLP app
- Sequence is loaded automatically (Thanks Sergio and Tianhong!)
- Use HELM Editor to edit the sequence
- Copy as HELM string and paste it back in the sketcher window (HELM string is converted to SCSR)

Advanced HELM Features Supported

- Support for in-line monomers represented by extended SMILES
  - PEPTIDE1{G.K.[C[C@H](N[*])C(=O)C[*] |$;;;_R1;;;;_R2$|].A}$$$$

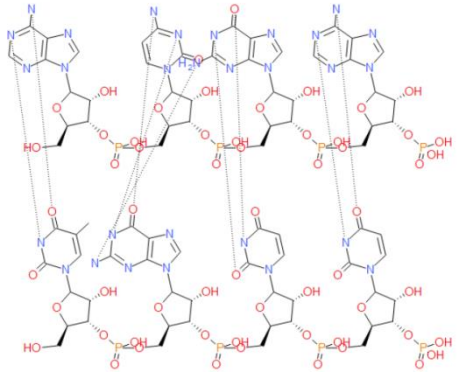
G K SMI1 A

sequence



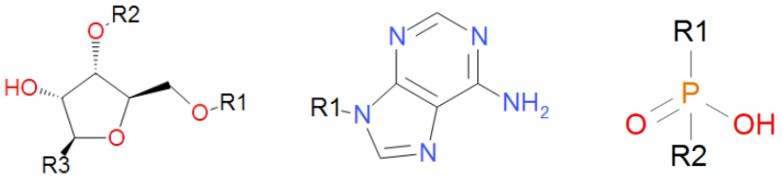
Explicit chemistry

- HBond information in HELM Strings converted to Hydrogen bond type in molecular data model

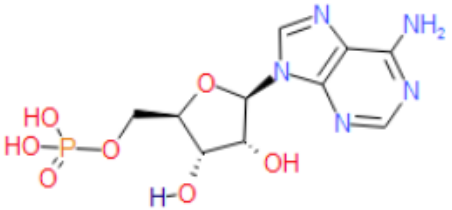


Limitations of Conversion of HELM Strings and SCSR Sequences

- Conversion between HELM and SCSR requires that all corresponding HELM monomers and SCSR templates are found in the specified configuration files (can be customized).
- SCSR with mutations and modifications can't be converted to HELM, as the HELM syntax doesn't support these features yet.
- Difficult to convert HELM strings with partial nucleotides (not the entire phosphate-sugar-base unit) to SCSR, as the SCSR templates for nucleic acids are based on complete nucleotides.



HELM: RNA1{R(A)P}



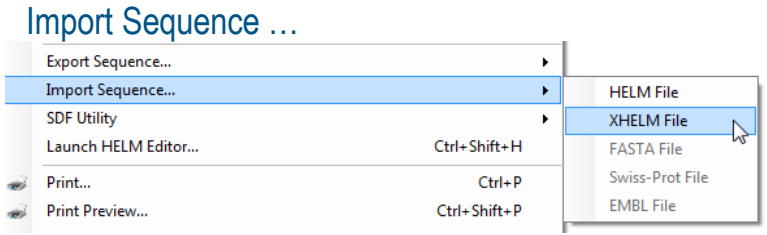
SCSR: RNA/A

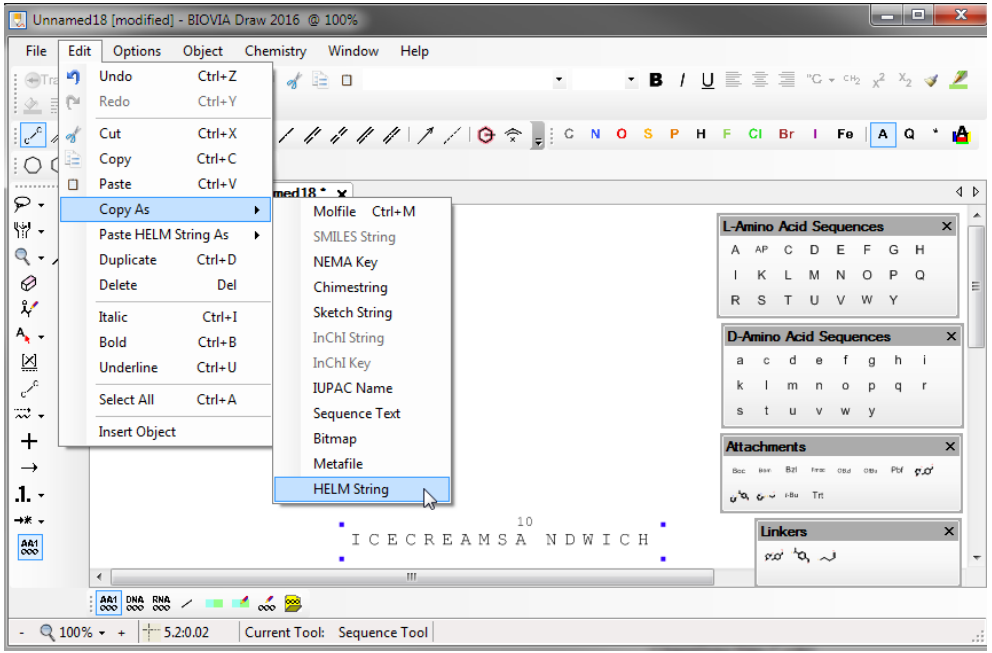
Support for HELM in BIOVIA Products

- Pipeline Pilot
  - HELM components
  - HELM Editor integration
  - Expanded SCSR template library and support for XHELM
- Draw Sketcher
  - Import and Export HELM Strings and XHELM
  - HELM Editor integration
- Direct Chemistry Cartridge
  - Operators to Import and Export HELM Strings
- Biological Registration
  - Register HELM strings as entities
  - Bring up HELM Editor

Support for HELM and XHELM in Draw Sketcher

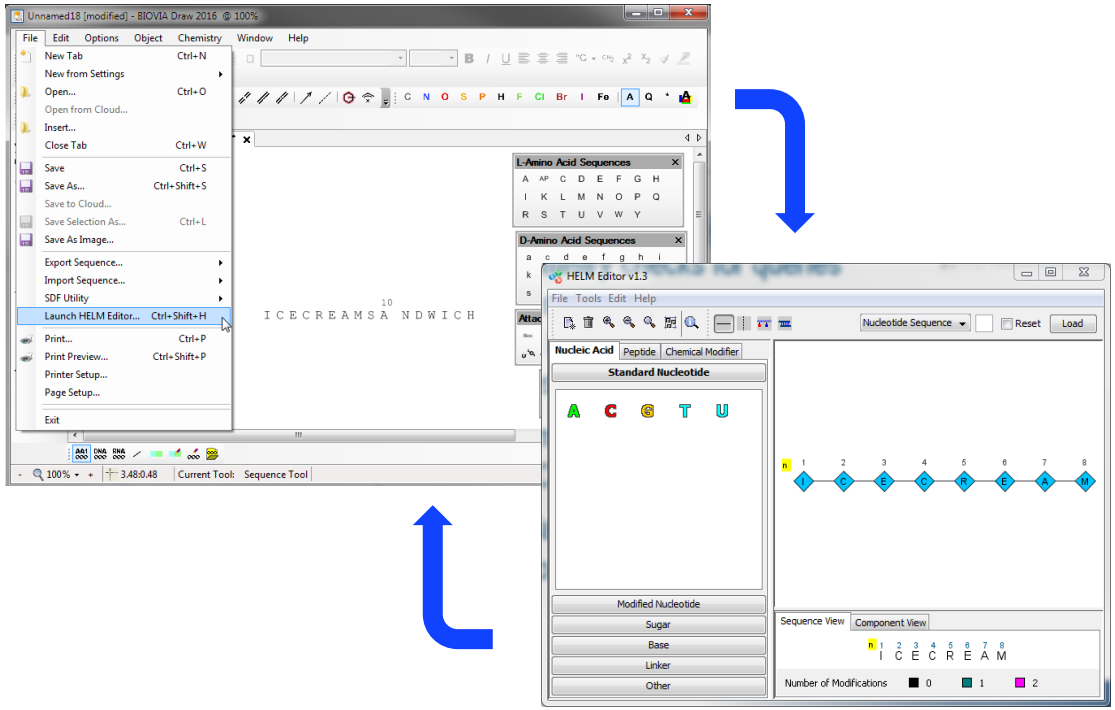
- Import and export sequences as HELM or XHELM
- PPChem .NET SDK used to read and process HELM and XHELM format





HELM Editor Integration In Draw

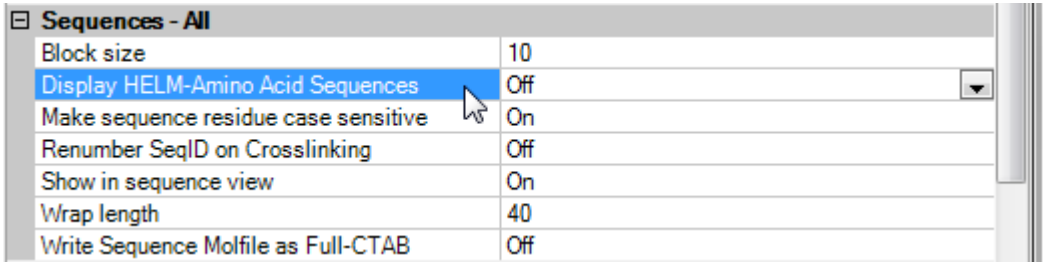
- Launch HELM Editor from new command in Draw File menu
- HELM Editor Java App is downloaded directly from Pistoia webpage
- Sequence in the canvas is loaded automatically into the HELM Editor
- Ability to edit sequences in the HELM Editor, copy as HELM notation and paste back into Draw



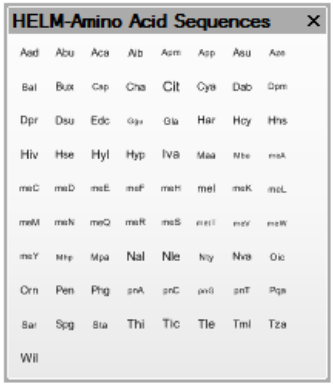
SCSR Templates for HELM Monomers in Draw

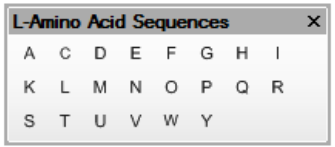
- Added SCSR Amino Acid templates corresponding to the HELM monomers found in the HELM Editor (optional toolbar)
- HELM AA monomers can be used in sequences with L-AA and D-AA

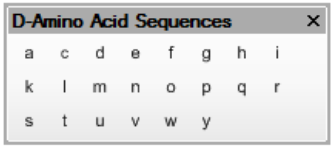
Settings



A D Aad Gla meD M K







Support for HELM in BIOVIA Products

- Pipeline Pilot
  - HELM components
  - HELM Editor integration
  - Expanded SCSR template library and support for XHELM
- Draw Sketcher
  - Import and Export HELM Strings and XHELM
  - HELM Editor integration
- Direct Chemistry Cartridge
  - Operators to Import and Export HELM Strings
- Biological Registration
  - Register HELM strings as entities
  - Bring up HELM Editor

Support for HELM in Direct Cartridge

- Register molecules from HELM and XHELM
  - HELM strings can be used anywhere you would normally expect a molecule - the code tries to interpret the input string as molfile, Chime, SMILES, **HELM** and **XHELM**

```
insert into molecules (name, ctab) values
('RNA1',  mol('RNA1{R(G)P.R(C)P.R(A)P.R(T)P.R(U)P.R(C)P.R(G)P.R(A)P}$$$$'));
```

- Function to convert either HELM or XHELM strings directly into a molfile (SCSR)

```
select
mdlaux.helmtomolfile('PEPTIDE1{A.C.D.E.F.G.H.I.K.L.M.N.P.Q.R.S.T.V.W.Y}$$$$')
from dual;
```

Support for HELM in Direct Cartridge

- Calculate HELM and XHELM for registered molecules
  - Given a database of sequence molecules you can get the HELM or XHELM strings with the following commands (‘ctab’ is the column name of the chemistry column in the table, ‘molecules’ is the name of the table):

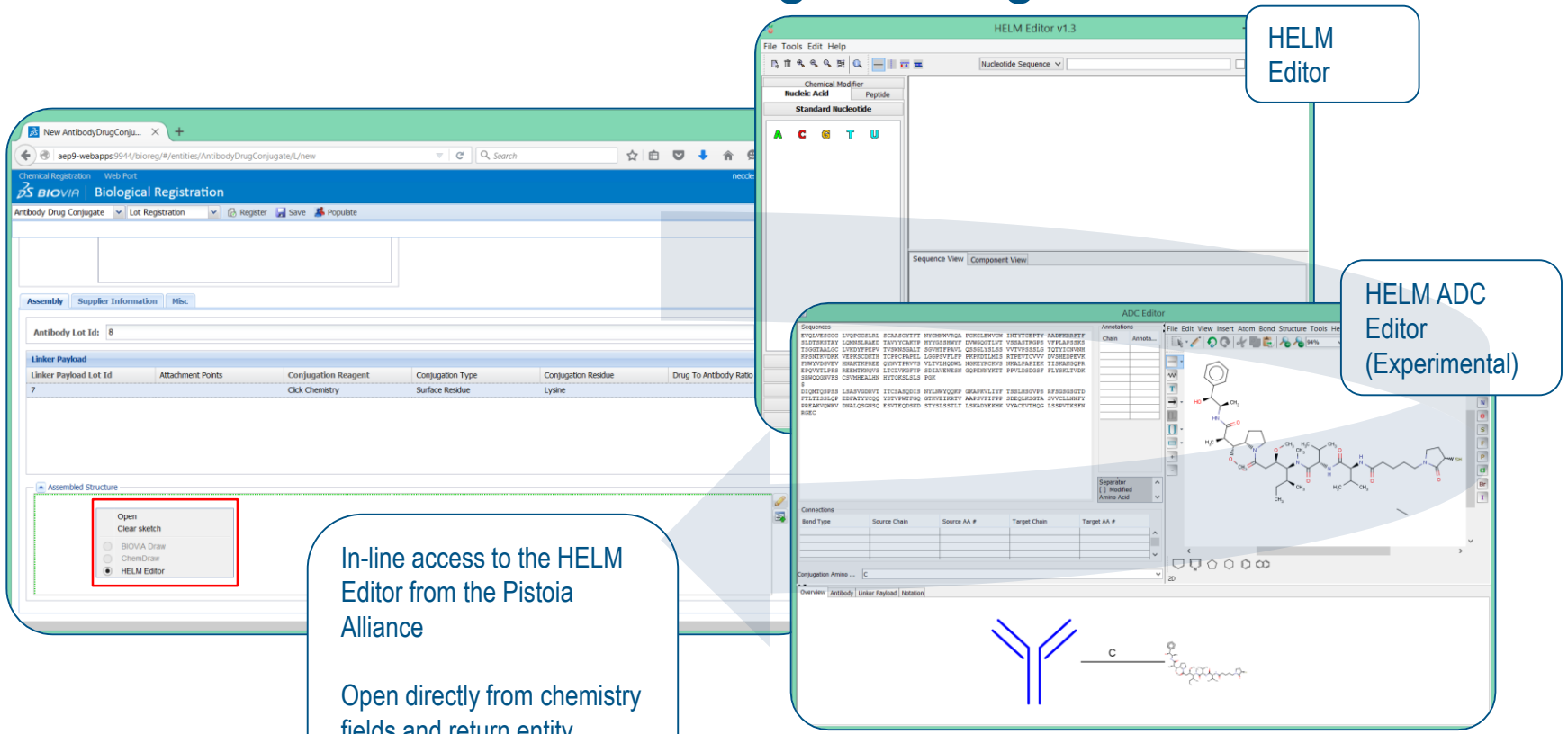
```
select name, helm(ctab) from molecules;

select name, xhelm(ctab) from molecules;
```

Support for HELM in BIOVIA Products

- Pipeline Pilot
  - HELM components
  - HELM Editor integration
  - Expanded SCSR template library and support for XHELM
- Draw Sketcher
  - Import and Export HELM Strings and XHELM
  - HELM Editor integration
- Direct Chemistry Cartridge
  - Operators to Import and Export HELM Strings
- Biological Registration
  - Register HELM strings as entities
  - Bring up HELM Editor

Use of HELM Editor in Biological Registration



HELM Editor

HELM ADC Editor (Experimental)

In-line access to the HELM Editor from the Pistoia Alliance

Open directly from chemistry fields and return entity representations via the clipboard

Future Directions

- Centralized Monomer Libraries
  - Tools to manage and share monomer libraries among several products (Pipeline Pilot, Draw, Direct, …)
  - Add monomers, check for duplicates, search by name or substructure, …
- Enable use of Draw sketcher in HELM Editor
  - Optional use of Draw to display and edit monomer structures (instead of Marvin)