

Biomolecule Uniqueness Check Using HELM and Hashcode

David Deng

Director of Software Development

Tony Yuan

President, Head of R&D

InChI Symposium San Diego

Aug 23rd, 2019



Scilligence's Uniqueness Check

- **Scilligence's cartridge supports both small molecules and biomolecules**
- **Uniqueness check uses a proprietary Hashcode approach**
 - Similar to InChIKey
 - For both small and biomolecules
- **Leading support in HELM for biomolecules**

HELM Monomer Library Management

Monomer List						
Symbol: <input type="text"/> Polymer Type: PEPTIDE Monomer Type: Backbone Substructure: <input type="text"/> Display As: Card Count: <input type="text"/>						
Previous Page 1 2 3 4 5 6 7 8 9 10 11 12 Next Page						
Symbol: A Aliases: Name: Alanine Natural Analog: A Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: Author: Versions: Created Date: 2018-Aug-03	Symbol: Aad Aliases: Name: 2-aminoadipic acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: OH Author: Versions: Created Date: 2018-Aug-03	Symbol: Abu Aliases: Name: 2-aminobutanoic acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: Author: Versions: Created Date: 2018-Aug-03	Symbol: ac Aliases: Name: N-Terminal Acetic acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: R2: OH R3: Author: Versions: Created Date: 2018-Aug-03	Symbol: Aca Aliases: Name: 2-aminocaproic acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: Author: Versions: Created Date: 2018-Aug-03	Symbol: Aib Aliases: Name: alpha-aminoisobutyric acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: Author: Versions: Created Date: 2018-Aug-03	Symbol: am Aliases: Name: C-Terminal amine Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: R3: Author: Versions: Created Date: 2018-Aug-03
Symbol: Apm Aliases: Name: 2-aminopimelic acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: OH Author: Versions: Created Date: 2018-Aug-03	Symbol: App Aliases: Name: gamma-amino-butyric acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: Author: Versions: Created Date: 2018-Aug-03	Symbol: Asu Aliases: Name: 2-aminosuberic acid Natural Analog: X Polymer Type: PEPTIDE Monomer Type: Backbone R1: H R2: OH R3: OH Author: Versions: Created Date: 2018-Aug-03				

HELM Sketcher: Monomer Selection

The screenshot shows the 'Monomer Explorer' window with a 'Peptide' tab selected. The interface includes a filter box and several tabs: Favorite, Chem, Peptide, RNA, and Rules. The main area displays a grid of monomer buttons organized by category (A, C, D, E, F, G). Each button is color-coded and includes a star icon in the top-left corner. The 'D' category buttons (D, dD, meD) are highlighted in red, indicating they are selected.

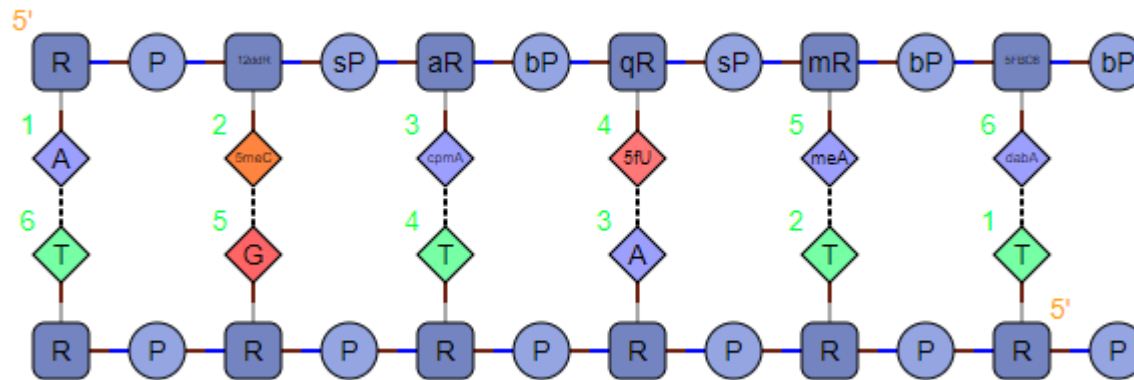
Category	Monomer	Color	Star
A	A	Grey	No
	Bal	Grey	No
	Cha	Grey	No
	Cya	Grey	No
	dA	Grey	No
	meA	Grey	No
	Nal	Grey	No
	Thi	Grey	No
Tza	Tza	Grey	No
C	C	Yellow	Yes
	dC	Yellow	Yes
	Edc	Yellow	Yes
	Hcy	Yellow	Yes
	meC	Yellow	Yes
	seC	Yellow	Yes
D	D	Red	Yes
	dD	Red	Yes
	meD	Red	Yes
E	E	Cyan	Yes
	dE	Cyan	Yes
	Ggu	Cyan	Yes
	Gla	Cyan	Yes
	meE	Cyan	Yes
F	F	Dark Blue	Yes
	dF	Dark Blue	Yes
	meF	Dark Blue	Yes
G	C	Grey	No
	Dha	Grey	No
	Sar	Grey	No

HELM Sketcher: HELM Structure/String Conversion



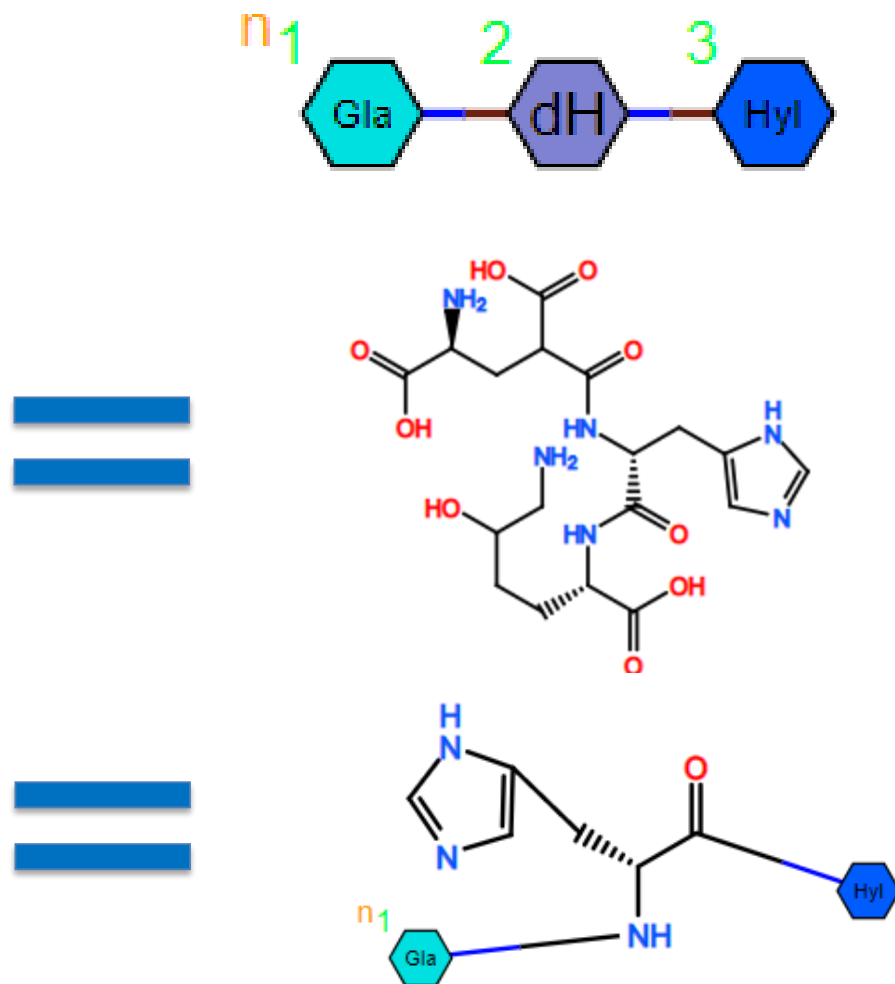
PEPTIDE1{A.S.D.F.D.F.S.D.F.A.S.F.D}\$\$\$\$V2.0

HELM Sketcher: RNA Complementary Strand



RNA1{R(A)P.[12ddR]([5meC])[sP].[aR]([cpmA])[bP].[qR]([5fU])[sP].[mR]([meA])[bP].[5FBC6]([dabA])[bP]}
 \$\$\$V2.0

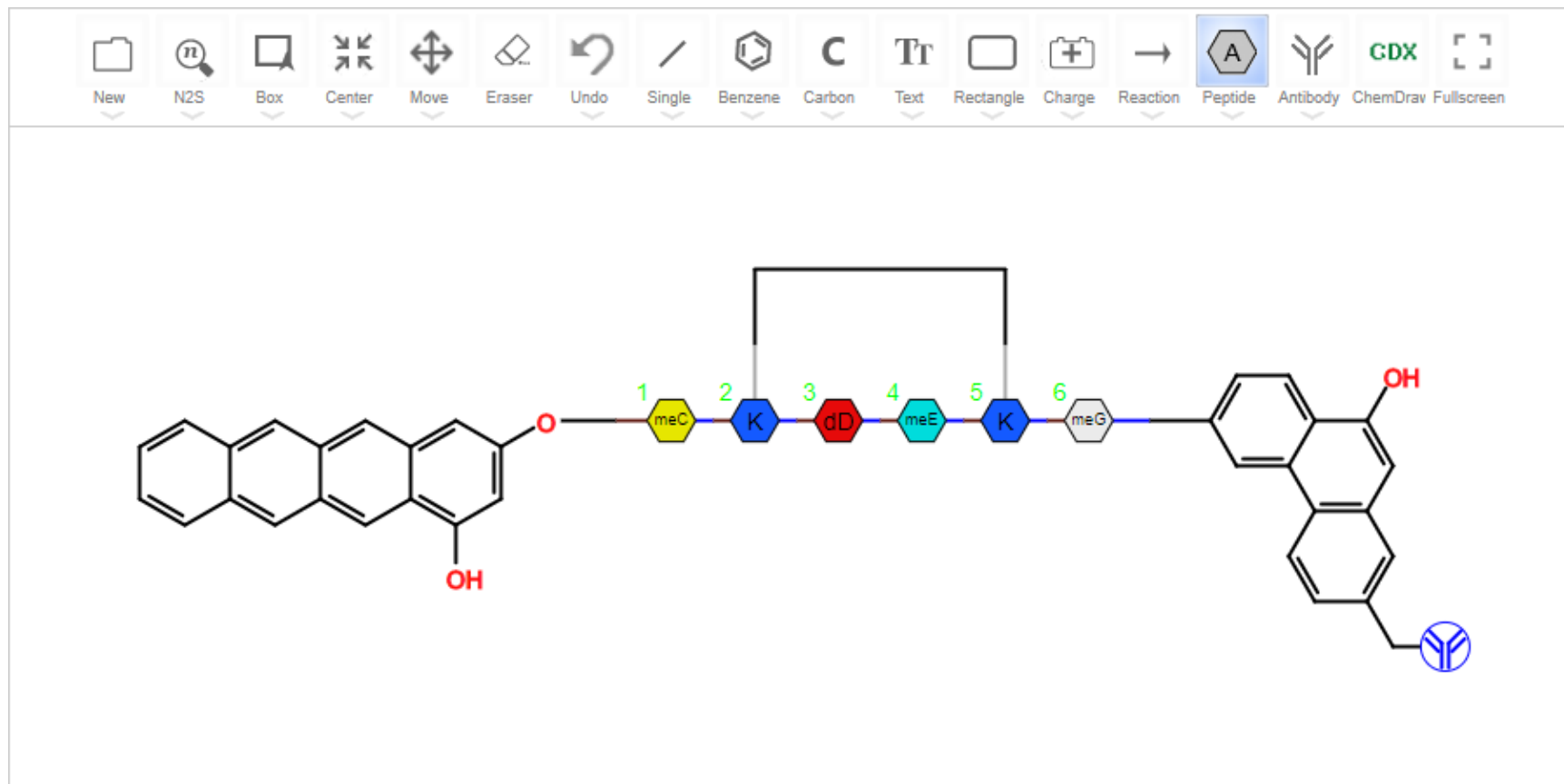
Backend: Scilligence's Uniqueness Check



However

Expanding is not always practical

HELM Sketcher: Hybrid of Entities (Mixed Mode)



PEGS Monomer

Monomer

Symbol:* APeg

Aliases:

Name:* Alanine with long PEGS side chain

Polymer Type:* PEPTIDE

Monomer Type:* Backbone

Natural Analog:* A

Status: Retired

Author:

SMILES:

Structure:

Chiral

R1: H

R2: OH

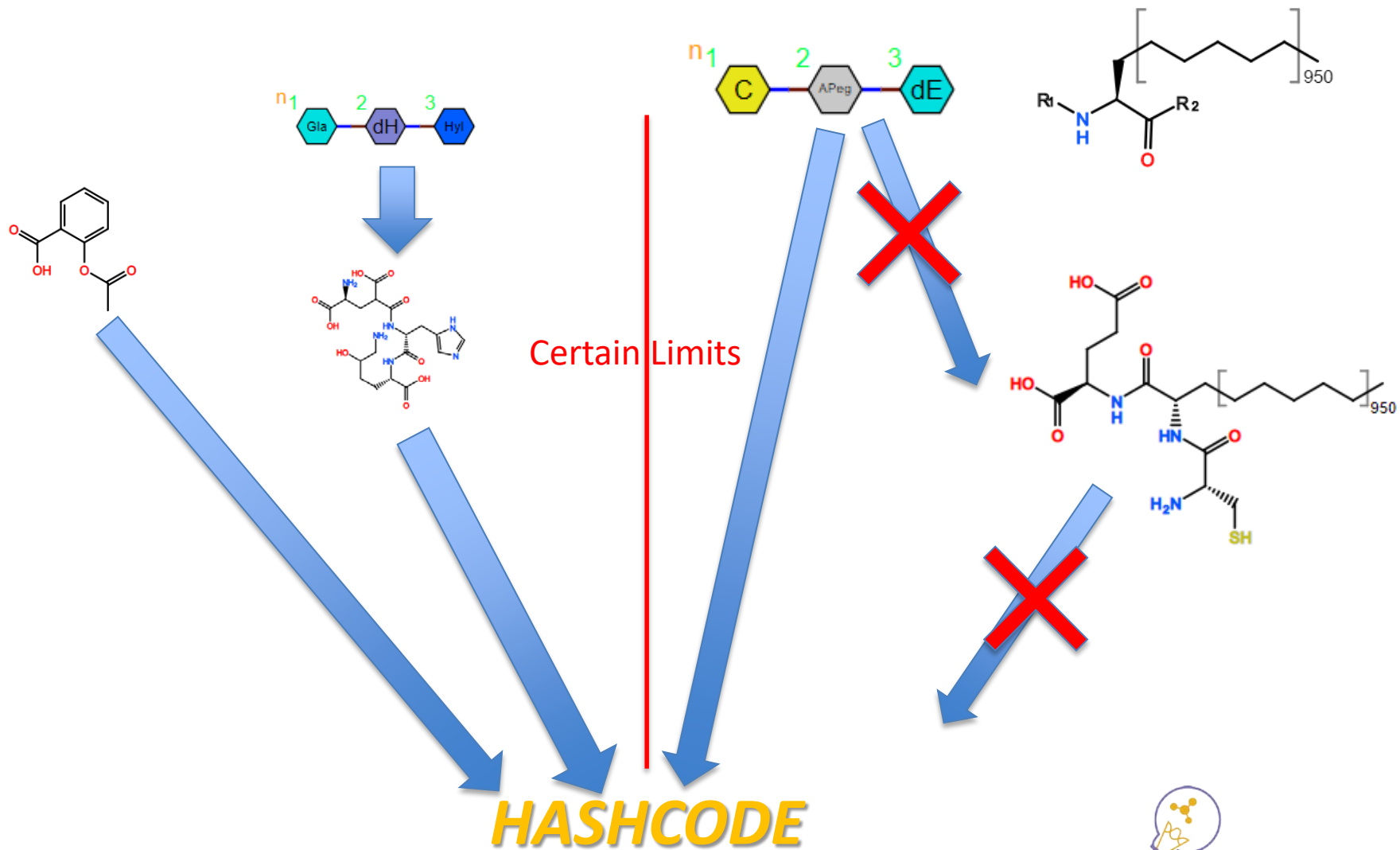
R3:

R4:

R5:

Save Delete Cancel

When to Hash??



Pros vs. Cons

- **Pros**

- Works for most common biopolymers
- Balance between accuracy and performances
- “Limit” configurable

- **Cons**

- Monomer library dependent
- Modifying monomer structures will change the HashCode

Thank You!
www.scilligence.com