User Guide for the HELM Editor II

Last updated: January 18th 2018

Contents

Intended Audience

This document explains the essential functions of the HELM editor.

It assumes familiarity with the basics of the HELM notation and common types of macro-molecules, such as, nucleotides and peptides.

For more information on the HELM notation, please see the HELM website:

www.openHELM.org

Quick reference guide to the main screen
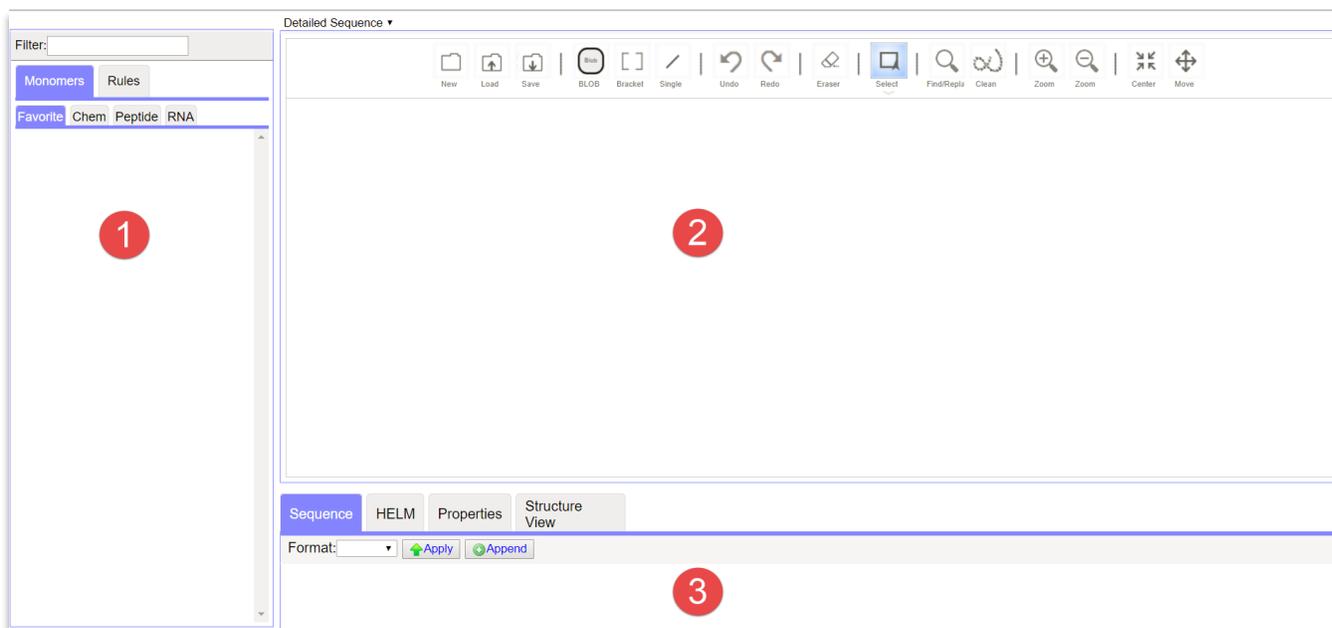
1. Monomer browser
   a. Provides access to monomers required to build macromolecules.
2. Canvas
   a. Drawing area where you can place monomers and build macromolecules.
3. Viewing area
   a. Provides alternative views of the molecule such as natural analog sequence, HELM, properties such as mol wt and the atom-bond structure.

Monomer Browser
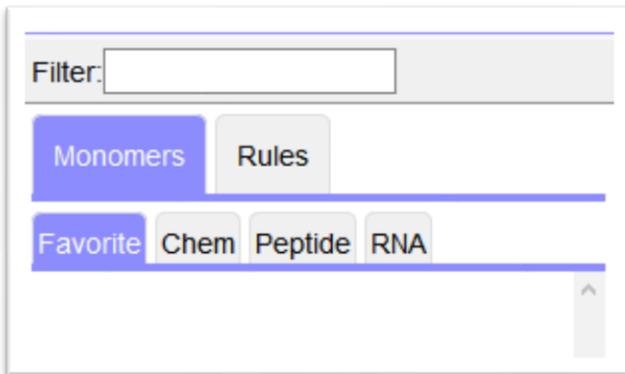
The monomer browser provides access to the monomer library, which is divided into chemical monomers, peptide and nucleotide monomers. The tabs are used to browse the library.

Frequently used monomers can be displayed in the favorites tab and frequently used functions can be saved under the Rules tab.

Filtering the monomer set
The Filter box can be used to search for monomers.
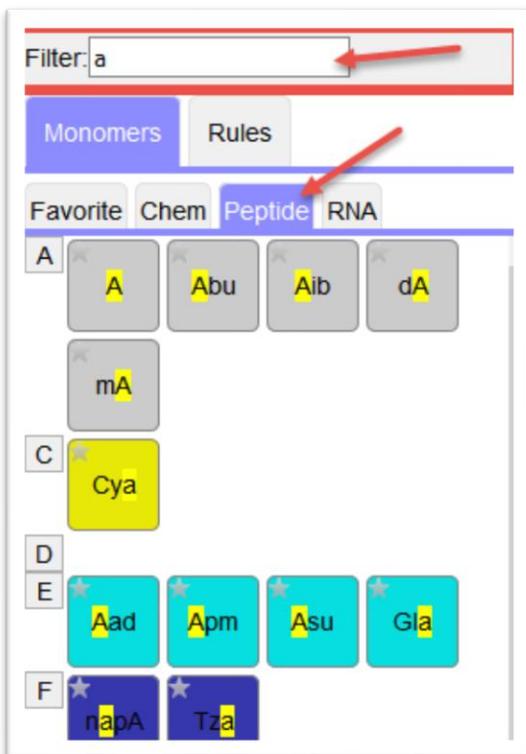
To search for a monomer, choose the tab for the type of monomer, then, in the Filter box, type letters or numbers from the monomer symbol.

Note: Entering lower case letters will display items with either upper or lower-case letters anywhere in the code.

## Using Monomer Favorites

Frequently used monomers can be saved under the Favorite tab.
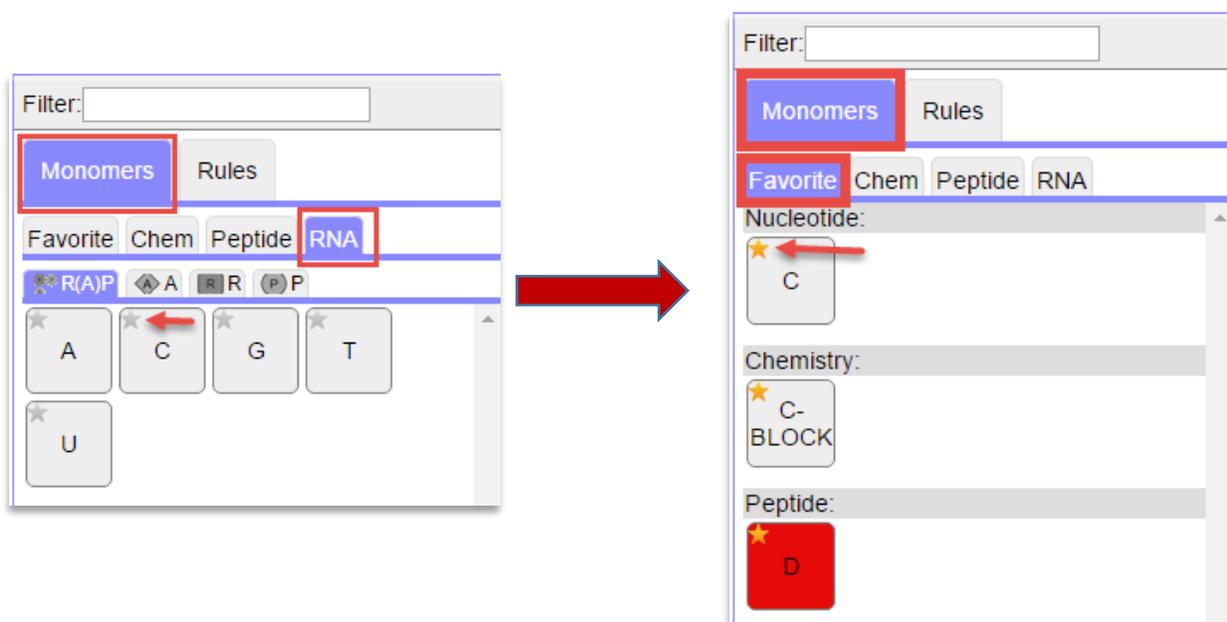
To create a favorite:

- Select the desired monomer under the Chem, Peptide or RNA tabs.
- Click on the star symbol in the upper left corner of the monomer icon.
- The star will turn from grey to yellow and the monomer icon will be saved in the Favorite tab.

Click the yellow star to deselect a Favorite.

Browsing the different monomer categories

Chem and peptides

Choose the Chem or peptide tab under the Monomer tab to display the different lists. The monomers will be displayed in the area below. Peptide structures are categorized according to their natural analog.

A tooltip of the chemical structure will appear by mouse-hovering over the chemical icon.

RNA

Nucleotides have subcategories as well as the main tab.
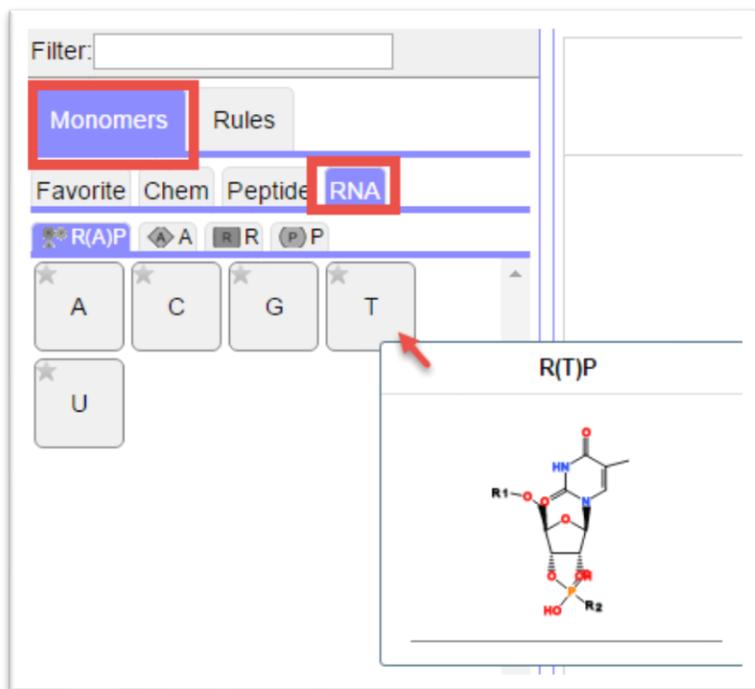
Choose the RNA tab under the Monomer tab to display the nucleotide list.
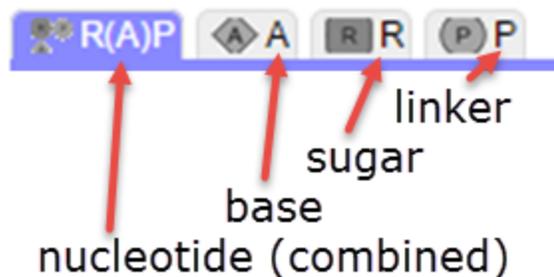
There are four tabs under the RNA tab:

1) R(A)P: Choose this tab to display the nucleotide (combined) options.
2) A: Choose this tab to display all the bases.
3) R: Choose this tab to display all the sugars.
4) P: Choose this tab to display all the linkers.

Monomer Ambiguity

You can add ambiguous monomers where the structure is not known. There are 4 different types:

- The character "*" represents 1 or more unknown monomers
- The character "X" represents one single unknown amino acid in a PEPTIDE polymer.

- The character "N" represents one single unknown base in a RNA polymer.
- The character "_" represents a deleted or missing single monomer. This is typically used for list elements

These monomers can be found in favorites section of the monomer palette (only the appropriate monomers for the polymer type). E.g.
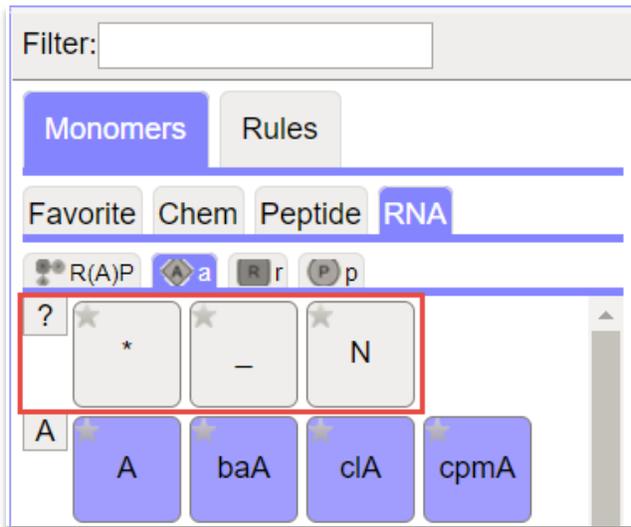
Rules
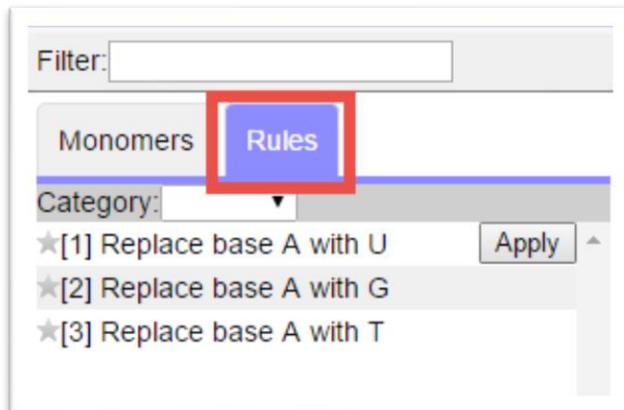
Choose the Rules tab to display preset rules.

Select a Category of preset rules from the Category dropdown menu.

Click the Apply button to apply the rule.

Canvas

The canvas is the main drawing area of the editor and contains a number of button to help the user construct a macromolecule.
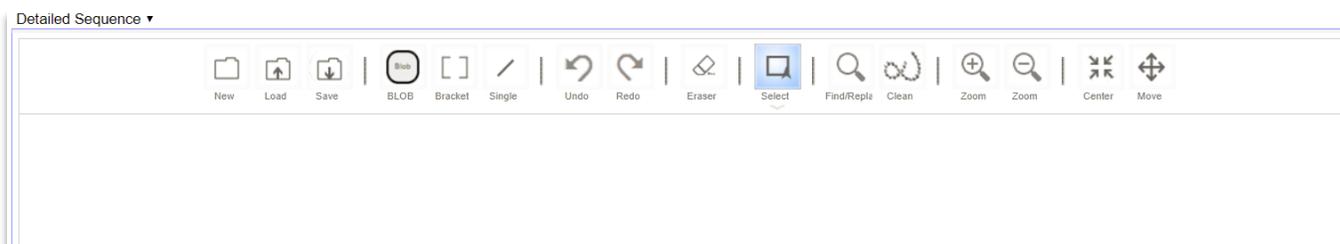
Buttons:

- New: Clears the contents in the canvas.
- Load: Loads a HELM or xHELM file.
- Save: Save contents as a HELM or xHELM file.
- BLOB: Used as unknown simple polymer type.
- Bracket: To select a monomer or fragment and specify repeating units.
- Single: Single bond connect.
- Undo/Redo: To undo or redo a change on the canvas.
- Eraser: Erases a selected monomer.
- Select: Dropdown menu with Select, Lasso, Fragment or All tools to capture some or all of the monomers displayed on the canvas.
- Find/Replace: To find a monomer or to replace a monomer with another.
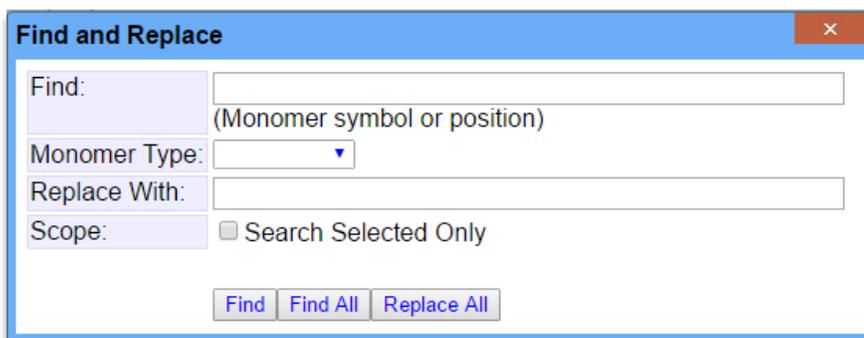
- Layout: Cleans up monomer display.
- Zoom +/-: Makes selected monomers larger or smaller.
- Center: Centers monomer display on canvas and displays zoomed monomers at default size.

- **Move**: Moves all monomers around the canvas. Also used to enlarge an area of a structure. Select move and use the mouse-wheel to zoom in and out on a structure. Mouse-click on area on which to focus.

Choose Detailed Sequence/Sequence from the dropdown menu to switch the display of the Canvas and Viewing Area.

Viewing area
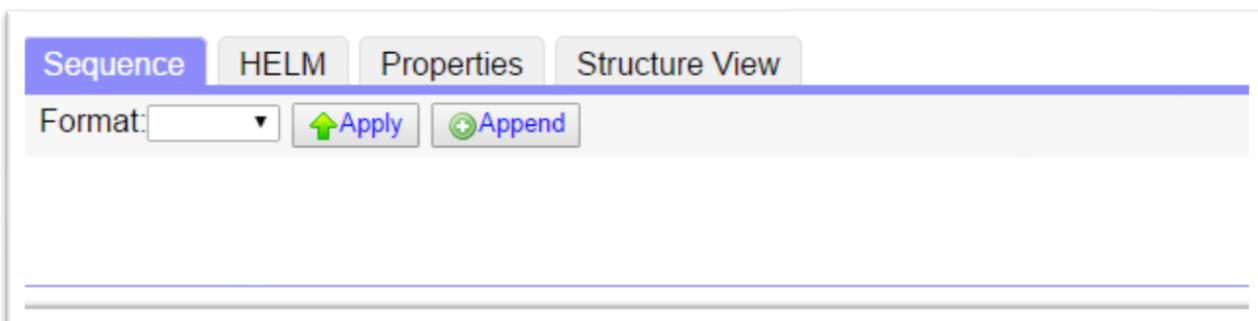
See a sequence view of a molecule or the component view that provides details such as Mol. Wt., Mol. Formula etc.

Sequence viewing
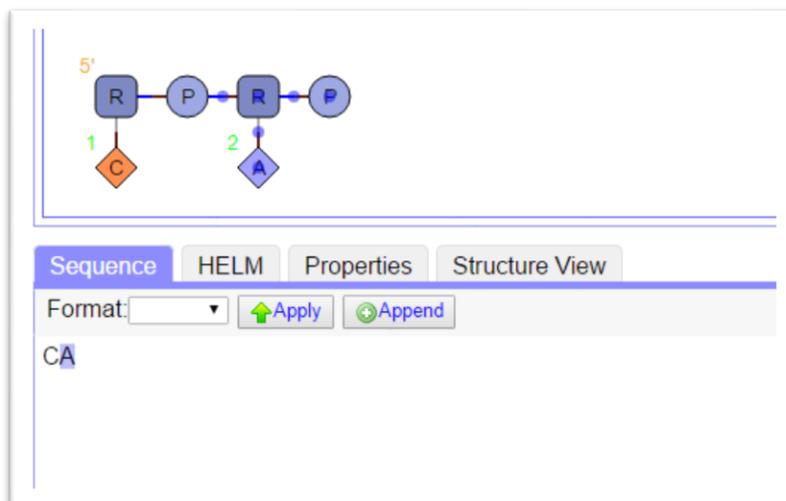
Choose the Sequence tab to display the sequence of monomers drawn on the canvas.

In the example sequence view the Nucleotide 'A' has been selected on the canvas and the corresponding nucleotide in the sequence display is highlighted.

Adding monomers via the sequence tab

A sequence can also be imputed directly by writing the sequence in the Sequence Viewing area or copying and pasting a sequence into the Viewing area and clicking Apply. Choose the format before applying!
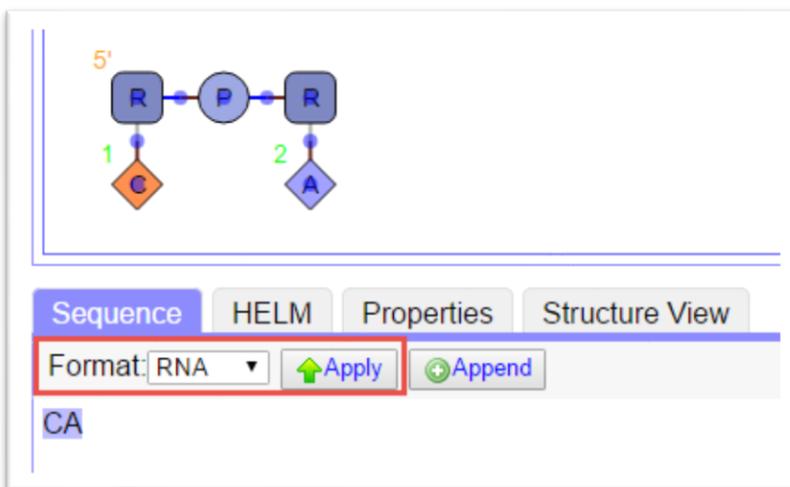
OR

You can add to an existing structure using the append function. This will add the new sequence to the canvas and you can then attach them together using normal drawing functionality.
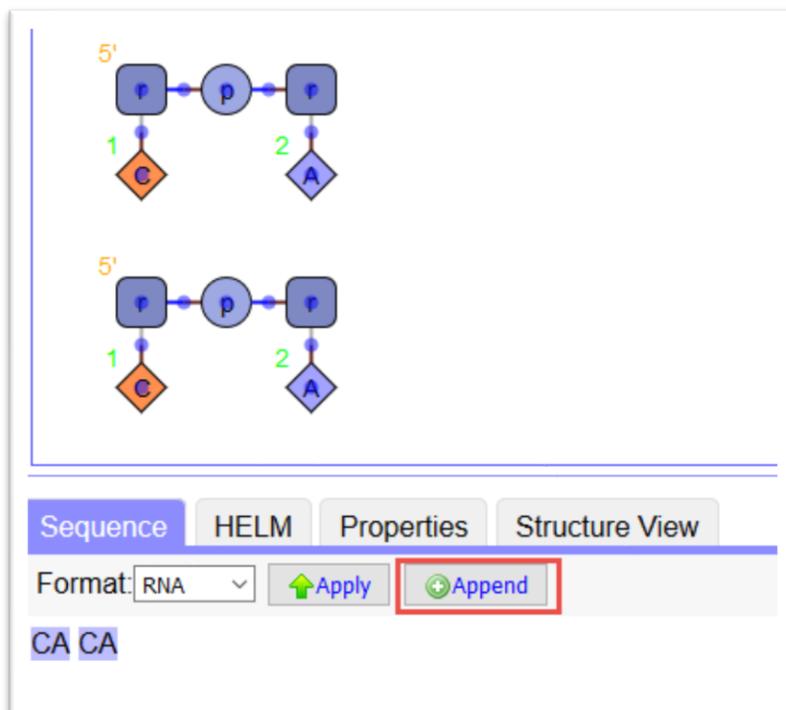
HELM notation

Choose the HELM tab to display the HELM notation. The notation for the sequence 'CA' is displayed with 'A' highlighted because this is highlighted on the canvas.
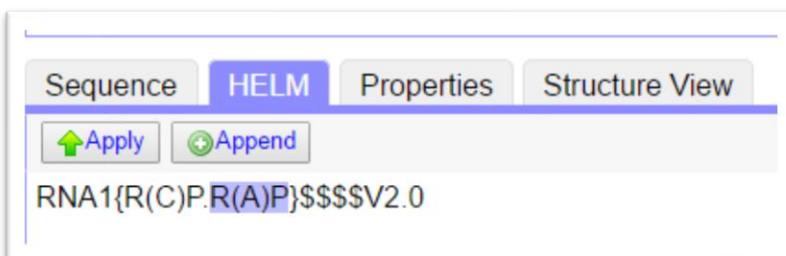
Properties

Choose the Properties tab to display the Molecular Weight, Molecular Formula and Extinction Coefficient for the entire sequence.

| Sequence | HELM | Properties | Structure View |

Molecular Weight: 652.4
Molecular Formula: $C_{19}H_{26}N_8O_{14}P_2$
Extinction Coefficient: 22.94

Structure View

Choose Structure View to see the structure of the selected monomer (in this case A).

| Sequence | HELM | Properties | Structure View |
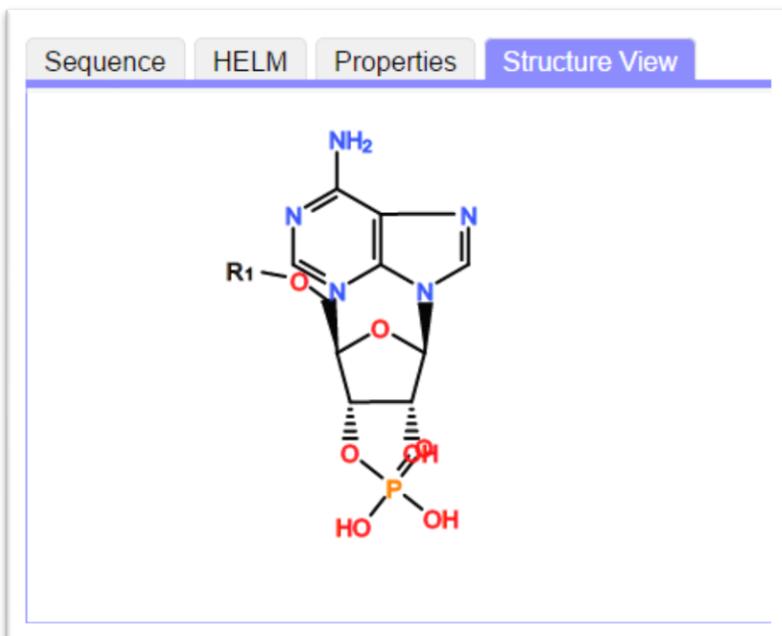
The Viewing area can be enlarged by dragging the line separating the Canvas and Viewing area up or down.

All or part of a structure can be viewed by using the Select button to choose the area to view.

Creating a new molecule

Using the drawing Canvas

Explore all the available monomers under the 'Chem', 'Peptide' and 'RNA' tabs.

There are three ways to add a monomer to the Canvas:

1) Click on the monomer in the Monomer Browser, and drag to the drawing Canvas.
2) Click on the monomer then click in the Canvas.
3) Write a sequence in the Viewing area.

After adding the first monomer, the next can be added in the same way.

A monomer on the Canvas can be changed to another monomer of the same type in three ways:

1. Highlight the replacement monomer in the Monomer Browser, hover over an existing monomer on the Canvas and click on the monomer,
2. Hover over an existing monomer on the Canvas and using the keyboard type the letter representing the monomer,
3. Highlight the replacement monomer in the Monomer Browser, hover over an existing monomer on the Canvas, left-mouse-click and drag left or right away from the monomer. In the example, an 'A' is added to the right, and then, a 'C' is added to the left. Note that the numbering changes when a monomer is added to the left.
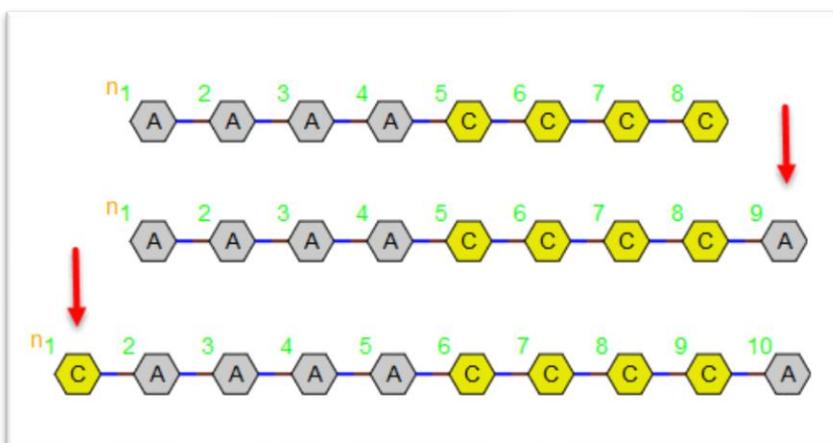
Monomers can also be added to any position in the sequence by adding them into the sequence in the Sequence Viewing area.
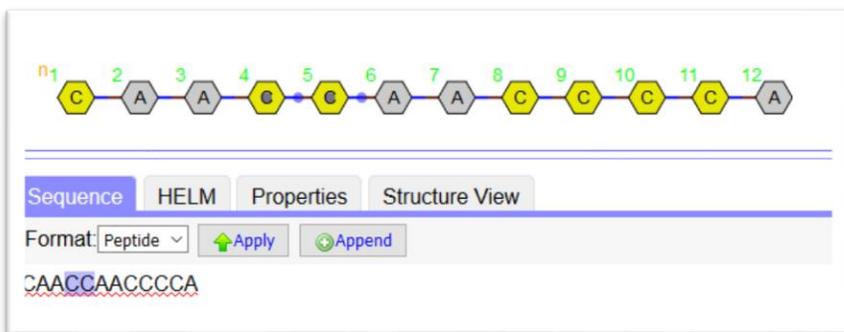
## Nucleotides

Nucleotides can be added to the Canvas in two ways:

1) Linkers, Bases, and Sugars from the RNA tab can be placed directly onto the canvas and individually linked. Note that the 5' end of the sequence is indicated.
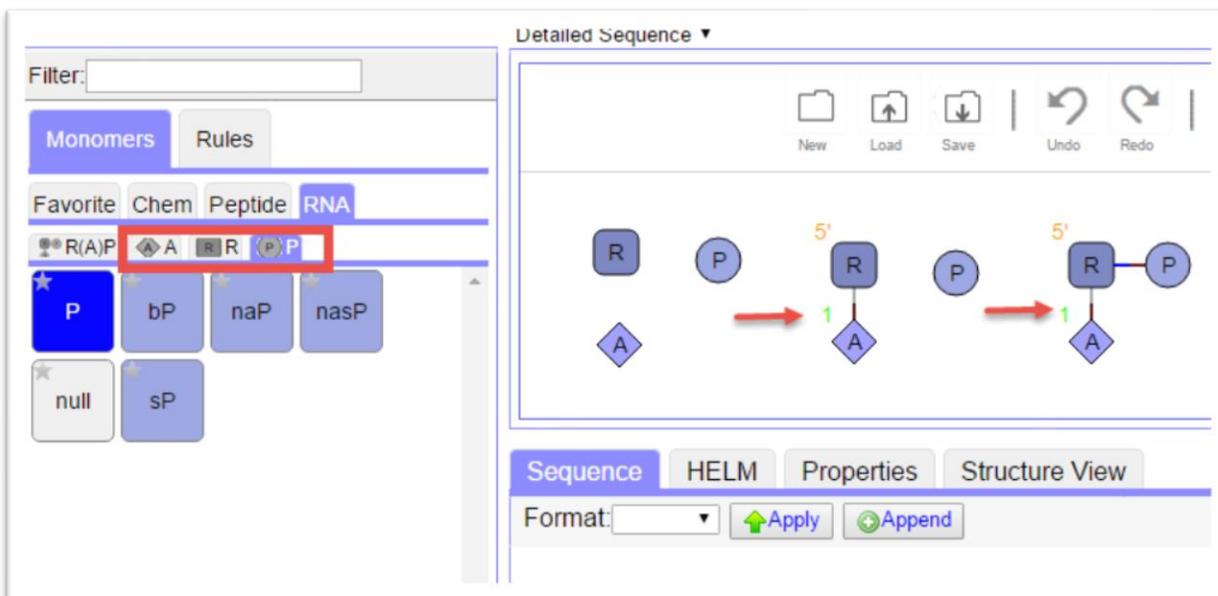
2) The Nucleotide can be added from under the Nucleotide (Combined) tab and then the individual linkers, bases and sugars can be changed by clicking on the desired linker, base or sugar in the monomer list then dragging to the desired position or clicking on the desired position.
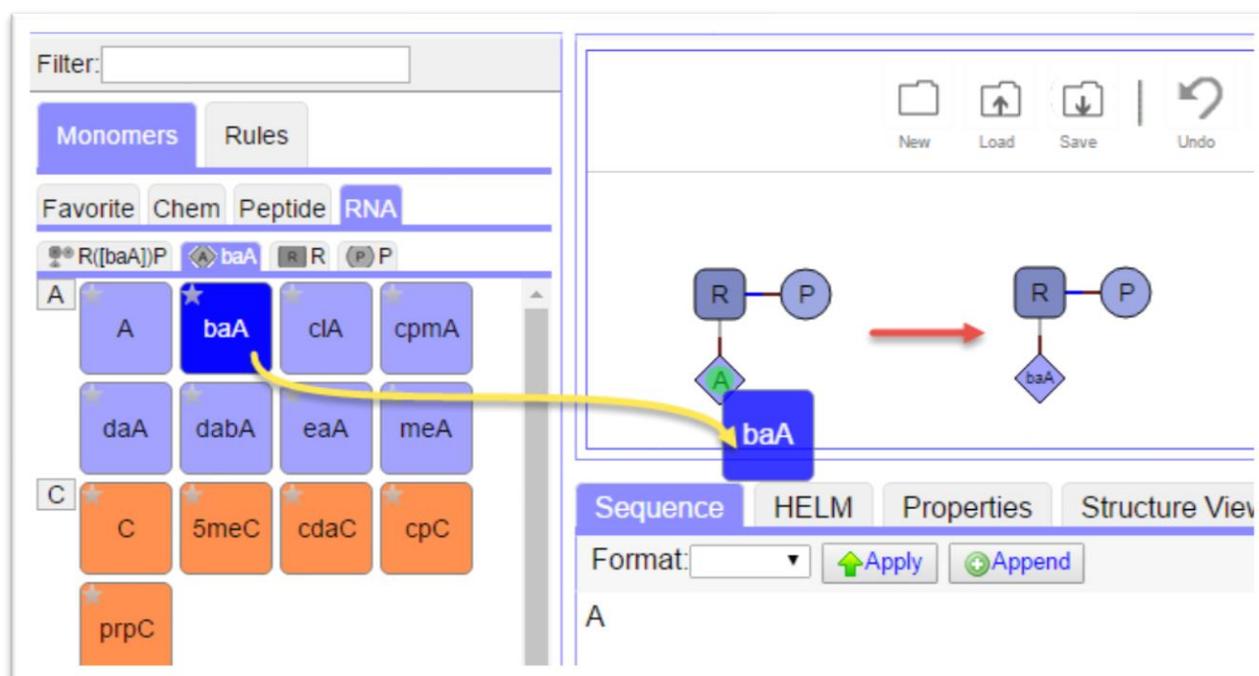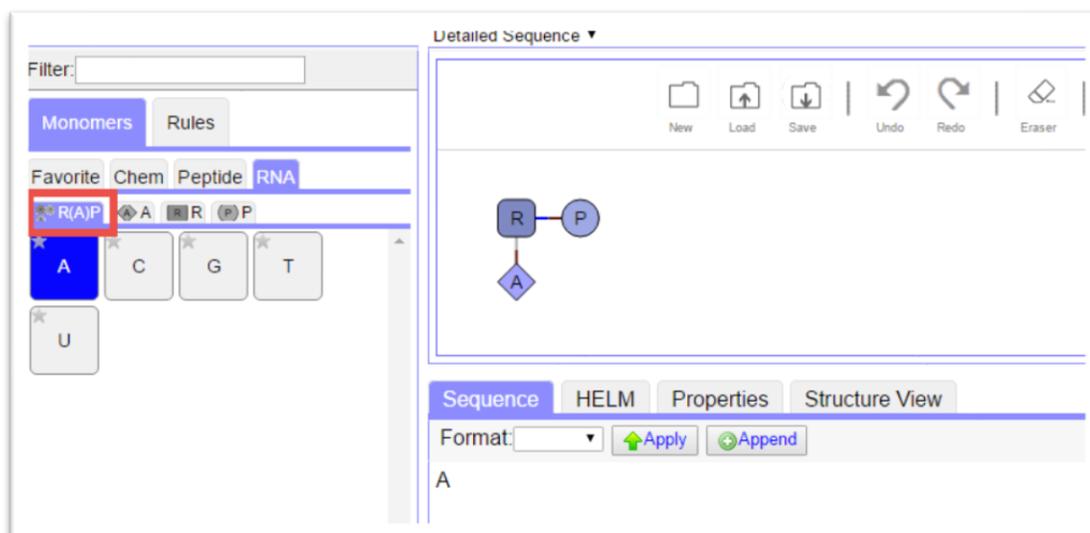
Once there is more than one monomer on the screen, a bond can be drawn between them by selecting one monomer, left-mouse-click, hold and drag to the second monomer.
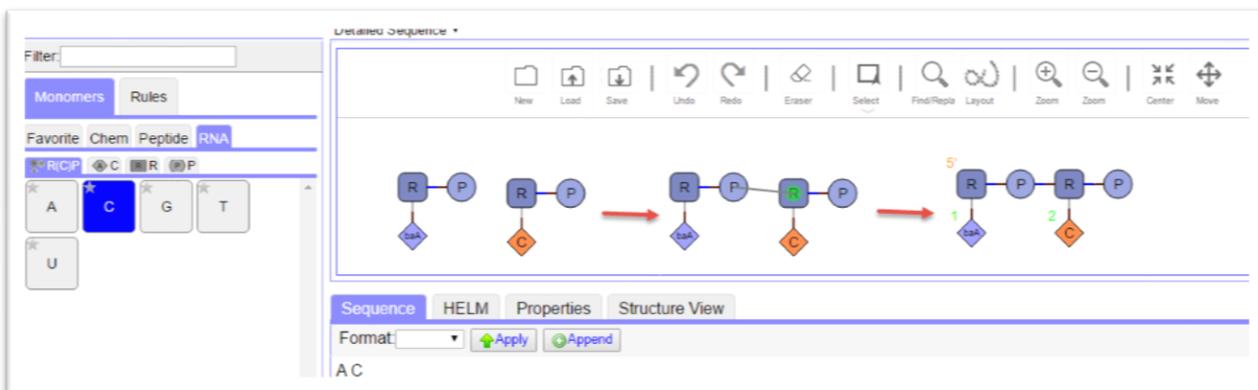
Please note the following principles:

- Connections are drawn from N to C terminus for peptides and from 5'->3' for the nucleotides.
- If monomers cannot be connected a dialogue box will appear with a message.
- If two sugars are connected, the phosphate linker is automatically inserted.

Using unknown monomers (Peptide and RNA)

Unknown monomers can be selected from the monomer palette in the same way as other monomers. The monomers are shown in yellow with a "?" to indicate that they are not specific.
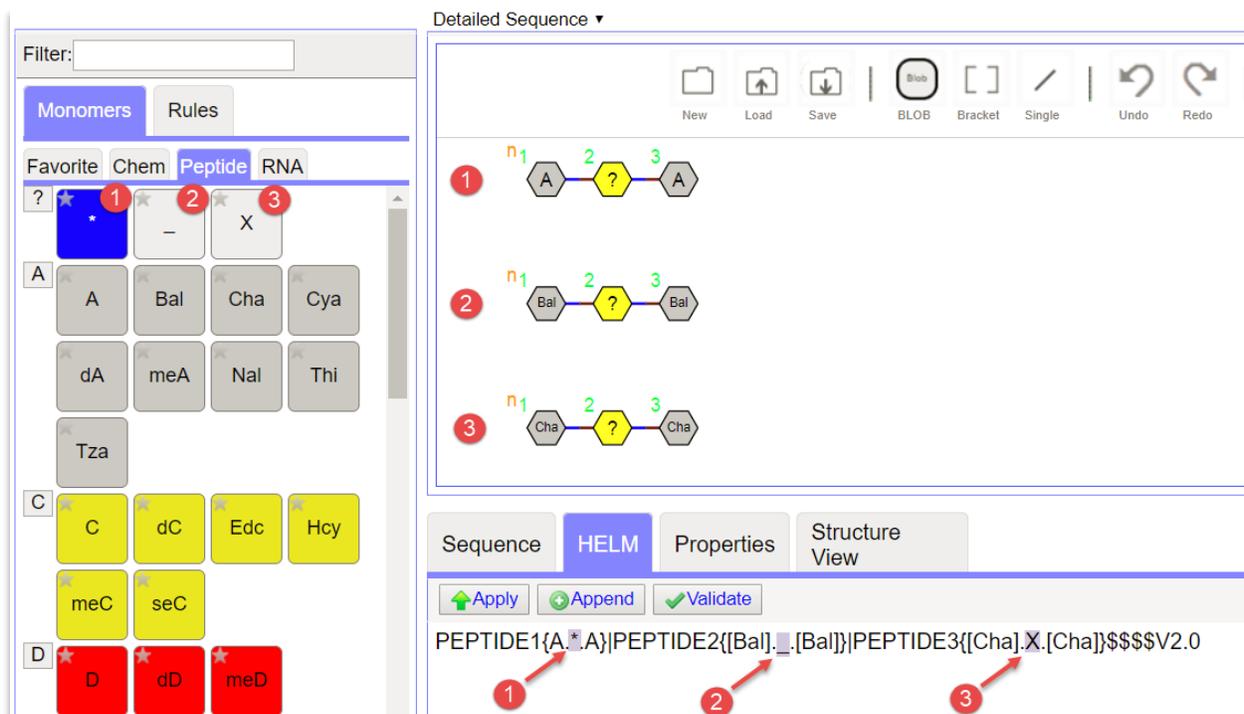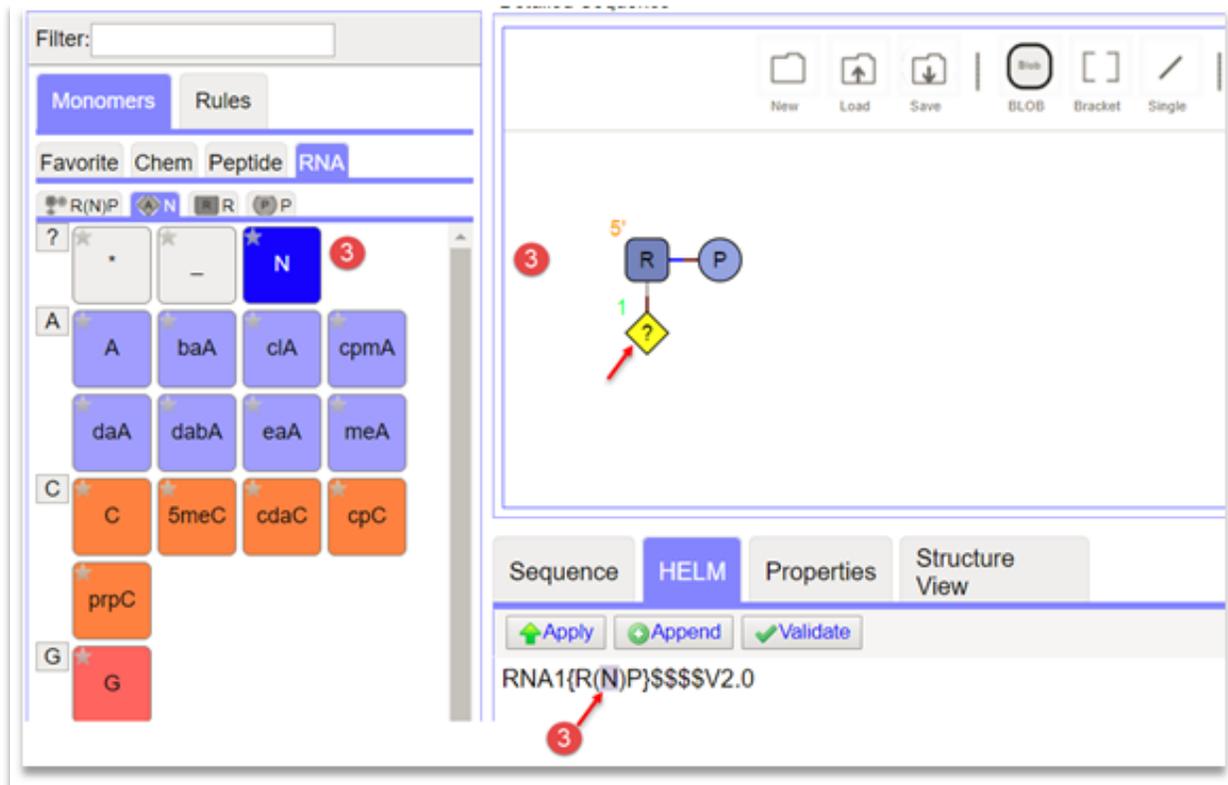
You can see the monomer details in the HELM string, or by hovering over the monomer on the canvas.

There are 4 types of unknown monomer:

- "*" represents 0..n unknown monomers.
- "X" represents one single unknown amino acid in a PEPTIDE polymer.
- "N" represents one single unknown base in a RNA polymer.
- "_" represents a deleted or missing single monomer. This is typically used in monomer lists (see Alternative monomers and mixtures)

Alternative monomers and mixtures

There may be situations where you have synthesized a molecule, but you either know that it is not pure, or are uncertain what the structure is. In these circumstances you may want to define a list of monomers at one or more positions in the molecule.

HELM provides the option to list monomers and specify whether you have

- monomer A OR monomer B.
- or monomer A AND monomer B

You can specify alternative monomers and mixtures by right-clicking on the monomer and choosing 'set monomer attributes'.
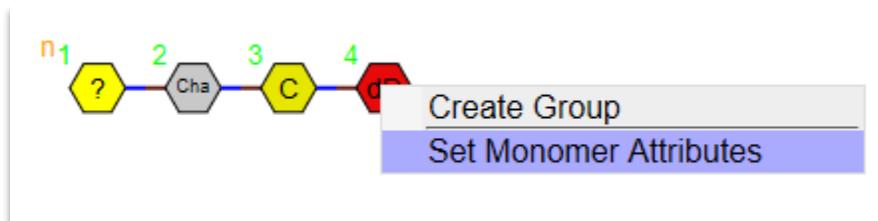
You can then type alternatives into the monomer symbol box


A comma between monomer symbols = alternatives

A plus sign = mixture.

If you know the relative proportions of each monomer, you also can add this to the monomer symbol box. For example:

This is reflected in the HELM for this molecule
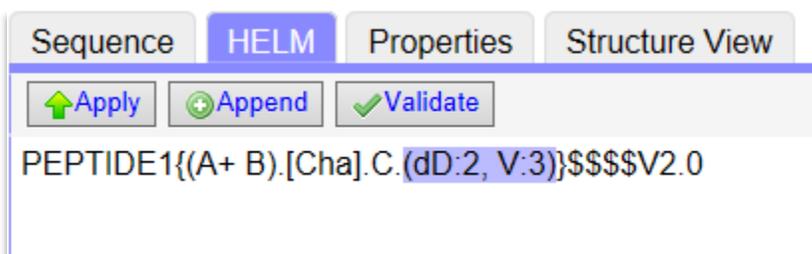
PEPTIDE1{(A+ B).[Cha].C.(dD:2, V:3)}$$$$V2.0

A more complex example is shown below:
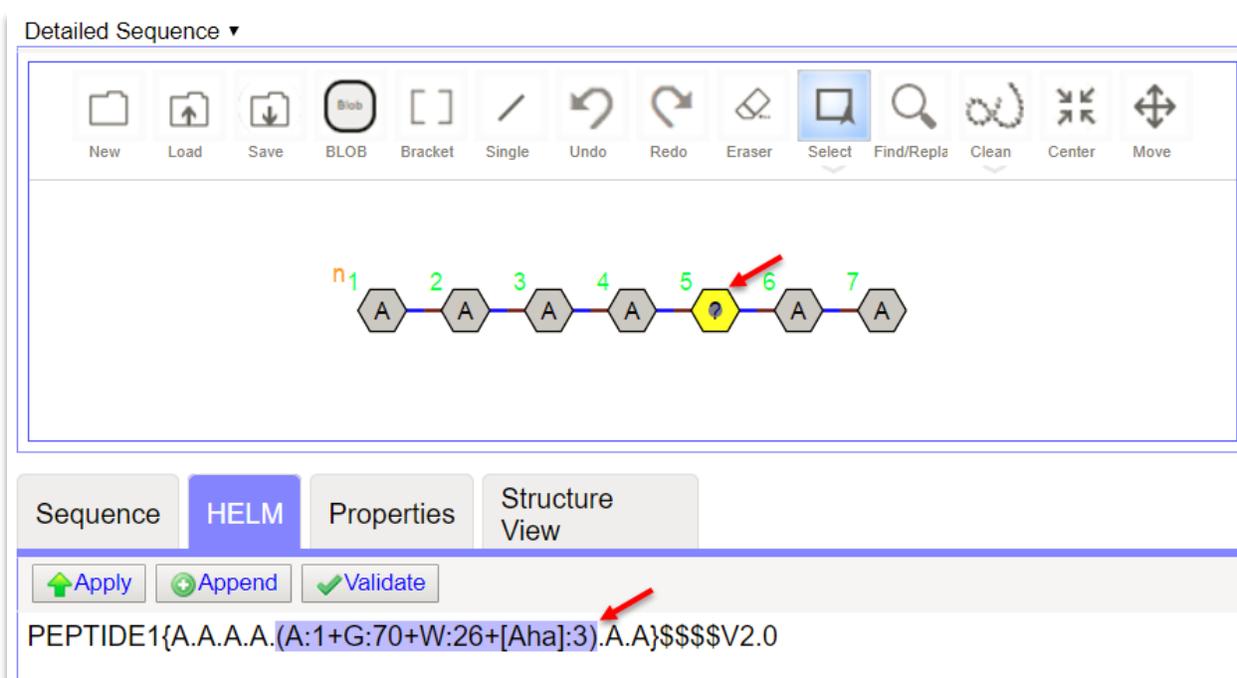
PEPTIDE1{A.A.A.A.(A:1+G:70+W:26+[Aha]:3).A.A}$$$$V2.0

Missing monomers

In some cases, a position can be occupied by a monomer or not. Example: A trailing Lysine might be present at the C-terminus of an antibody or not. An underscore is used to denote a potential blank position

New | Load | Save | BLOB | Bracket | Single | Undo | Redo | Eraser | Select | Find/Repla | Clean | Center | Move

n_1 1   2   3   4   5   6   7   8
A — A — C — D — D — E — E — ⚹

Sequence | HELM | Properties | Structure View

⬆Apply | ⊕Append | ✔Validate

PEPTIDE1{A.A.C.D.D.E.E.(_,K)}$$$$V2.0

Alternative polymers and mixtures

It is also possible to define a mixture or set of alternatives at a polymer level. Take this example of a mixture of 4 proteins with differing monomers at positions 2 and 5 and

10% PEPTIDE1{A.A.A.A.A}
40% PEPTIDE1{A.C.A.A.A}
10% PEPTIDE1{A.A.A.A.C}
40% PEPTIDE1{A.C.A.A.C}

To do this in the webeditor, you must:

1) Draw each peptide.
2) Make each peptide a group.
3) Right-click the border of each group, **Set Attributes**, enter ratio and click **Save**.
4) Select all four peptides and make a group.

Once you have done this for all groups you will see the values reflected in the HELM string.

Detailed Sequence ▾

New	Load	Save	BLOB	Bracket	Single	Undo	Redo	Eraser	Select	Find/Repla	Clean	Center	Move

ratio: 1

ratio: 4

ratio: 1

ratio: 4
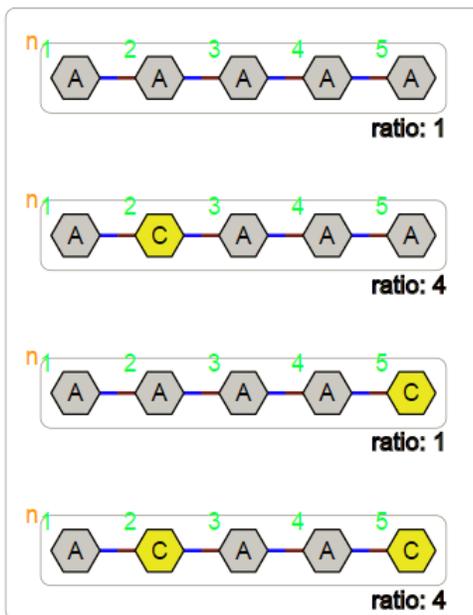
Sequence	**HELM**	Properties	Structure View

⬆Apply ⊕Append ✔Validate

PEPTIDE1{A.A.A.A.A}|PEPTIDE2{A.C.A.A.A}|PEPTIDE3{A.A.A.A.C}|PEPTIDE4{A.C.A.A.C}$$G5(PEPTIDE1:1+PEPTIDE2:4+PEPTIDE3:1+PEPTIDE4:4)$$V2.0

Monomer Annotations

You can add an annotation to a monomer by:

1) Placing a monomer on the canvas.
2) Right-click the monomer, right-click **Set Monomer Attributes**, add annotation, then click **Save.**
3) Mouse-hover over the monomer on the canvas to view the annotation
4) View HELM notation.
    - PEPTIDE1{*}"IL6"$$$$ represents a peptide chain with the name IL6.

Unknown simple polymers

Blobs represent simple polymers with an unknown structure/sequence.

Example

- BLOB1{Bead}"Aminated Polystyrene"$$$$ represents aminated polystyrene bead. Simple polymer type is "Bead", and simple polymer name is "Aminated Polystyrene".)

To add a blob:

1) Select **BLOB button.**
2) Left-click in the drawing canvas to place the BLOB.
3) Right-click the BLOB and select **Blob Type.**
4) Left-click the BLOB, right-click **Set Monomer Attributes**, add annotation/change monomer symbol, then click **Save.**
5) Click **Apply** button in the viewing area to apply HELM notation.

Connection ambiguity

Sometimes it is not possible to specify where or how many times components are connected to each other. You can represent this in the editor by using the bond attributes.

You can only change the bond attributes on bonds between groups or blobs.

Example: Connecting BLOB to Group and Setting Bond Attributes

1) Left-click the **Select** button and select the sequence on the canvas.
2) Right-click on the selected sequence and select **Create Group.**
   a. A border will now be around the sequence, indicating a group.
3) Place a BLOB on the canvas.
4) Left-click on the **Single** button and draw a bond connecting the Blob to the Group (Connect to any green circle on the Group border).
5) Left-click the **Select** button, right-click on the bond, select **Set Bond Attributes** (Position, R#, Ratio), and click **Save.**

Repeat sequences

One or more monomers may be repeated in a polymer. You can represent this in a compact way, by using brackets.

1) Select the **Bracket** button; click and drag to put brackets around desired sequence.
2) In the field, type a number, number range, or a "*" for ambiguity (e.g. 3, 3-6, *), then click on the canvas. To change the subscript, click the subscript and make changes in the field.

Conjugates
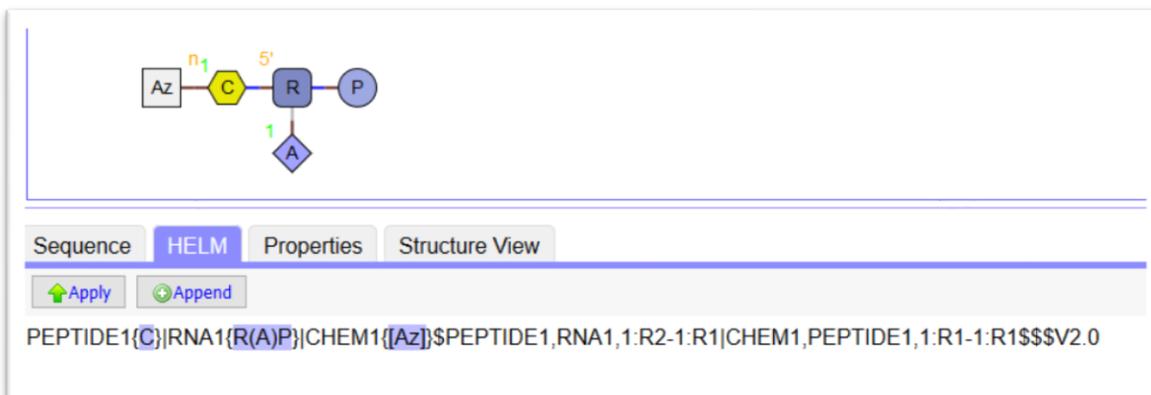
Conjugates can be drawn by selecting chemicals, peptides and nucleotides from the Monomer Browser and linking the monomers together.

If monomers cannot be connected a dialogue box will appear with a message.

R-Groups

It is important to understand the r-groups available in a monomer. The tooltip structure will show you where they are in the structure itself and the list at the bottom of the box shows the capping groups for each R group.

R groups are shown as capped in the Viewing area\Structure View tab. In this Alanine example R1 is H-capped.

A

Alanine

R2= OH
R1= H

Structure View
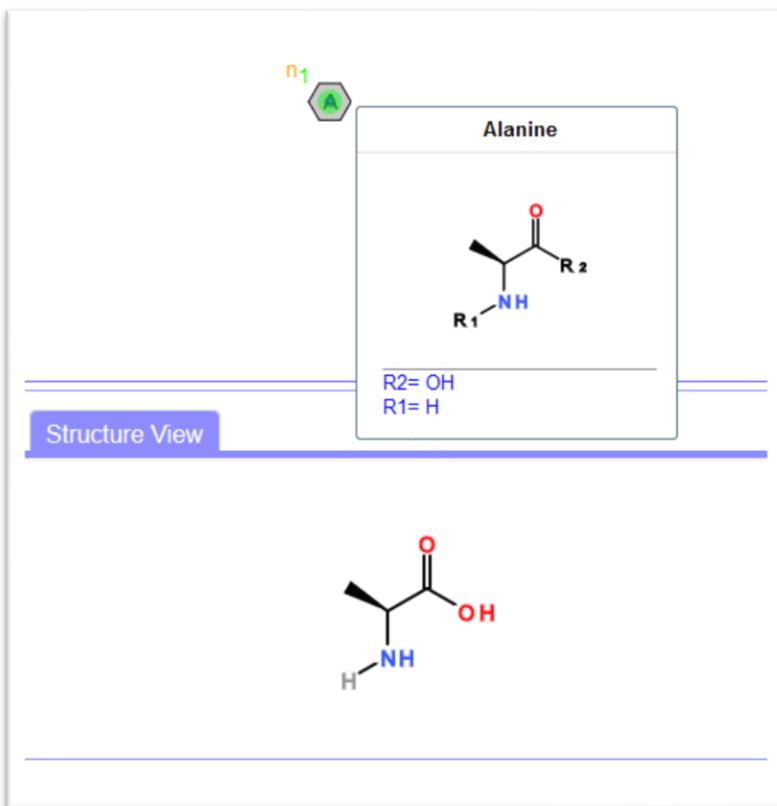
Monomers in the middle of a chain are typically linked to two other monomers, however certain monomers can form additional bonds. For example, Alanine (A) has two R groups (R1 and R2) with which to connect to other monomers, while Glutamic Acid (E) has three R groups (R1, R2 and R3).
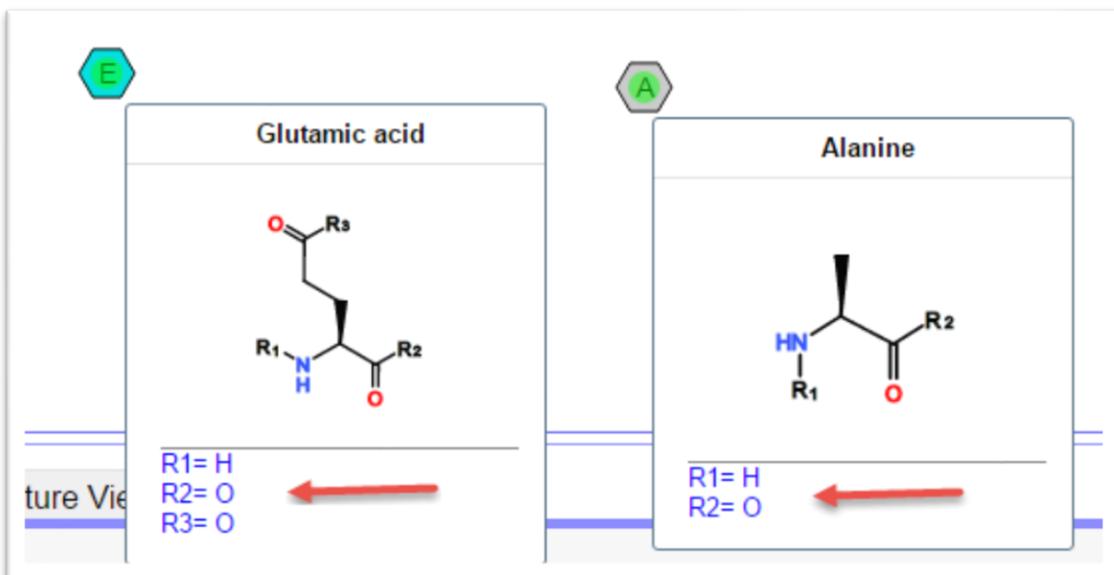
In this example, 'E' is attached to 'C'. When drawing the connecting bond between 'C' and 'E' a pop-up message will appear. The available connecting points for the two monomers will be displayed in the dropdown menu. There is only one available connecting point for C2 (R3). There are three available connecting point for E (R1, R2 and R3).

The number of attachment points of a monomer defines the number of covalent bonds in which the monomer can be involved.
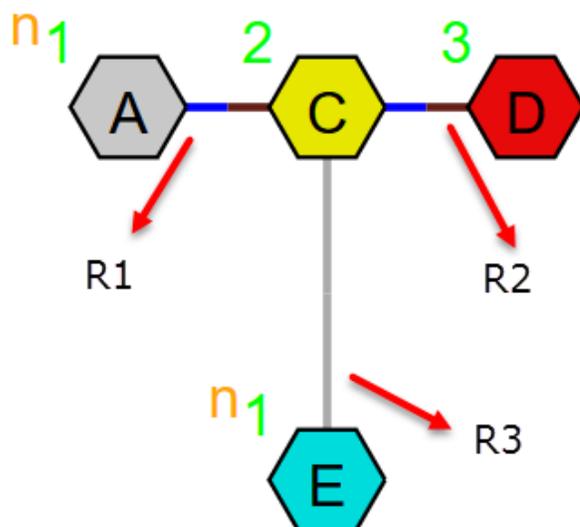
The Connection points are indicated by color: blue for R1, orange for R2, and gray for R3. The n-terminal is indicated by 'n'.

Working with structures

Monomers or bonds can be removed by mouse-hovering over them and hitting delete. Also, use the Select button to select the monomers to delete and hit delete.

To replace any of the monomers, drag a replacement monomer from the monomer browser and drop it onto the current monomer (as described previously), so long as their type and attachment points allow the replacement.

Replacing monomers

If a sequence contains a consistently modified monomer there are two options for updating it.

1. Use the Rules tab from the monomer browser (as described previously) to perform a search and/or replace the monomer.
2. Use the Find/Replace button (as described previously) in the Canvas menu to find a monomer or to replace a monomer with another.

Using the Find/replace box:

Find:

Enter the monomer in the Find box, choose the Monomer Type from the dropdown menu, and click Find to find the first monomer in a sequence or click the Find All button to find all monomers in a sequence. A pop-up message will display the number of monomers found. The monomers found will be highlighted in the Sequence or HELM Viewing area.

Replace:

Enter the monomer in the Find box, choose the Monomer Type from the dropdown menu, enter the replacement monomer in the Replace With box and click the Replace All button. A pop-up message will display the number of monomers replaced.

If only finding/replacing in a selected area, select that area using the Select button and check the Search Selected Only box.

Building a molecule using the sequence

Build new molecules by entering sequence information directly in the Viewing area, rather than dragging and dropping individual monomers, a process which is slow for larger polymers.

There are two sequence loading options: nucleotide sequence (RNA) and peptide sequence (Peptide).
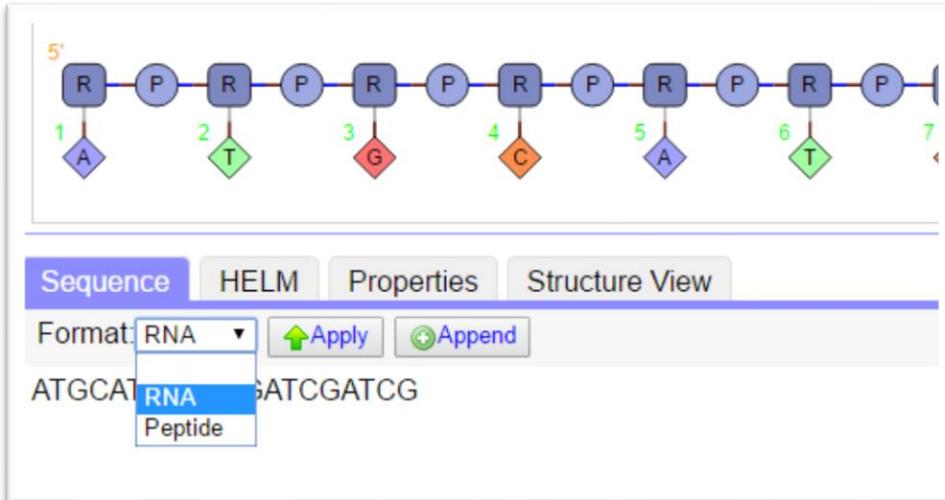
Cyclizing a sequence

To form a cyclic sequence, add a bond between two monomers that have suitable attachment points.

To create a backbone cycle, connect the first and the last monomers. In this example connect 'M', the first peptide, with 'N', the 19th peptide.
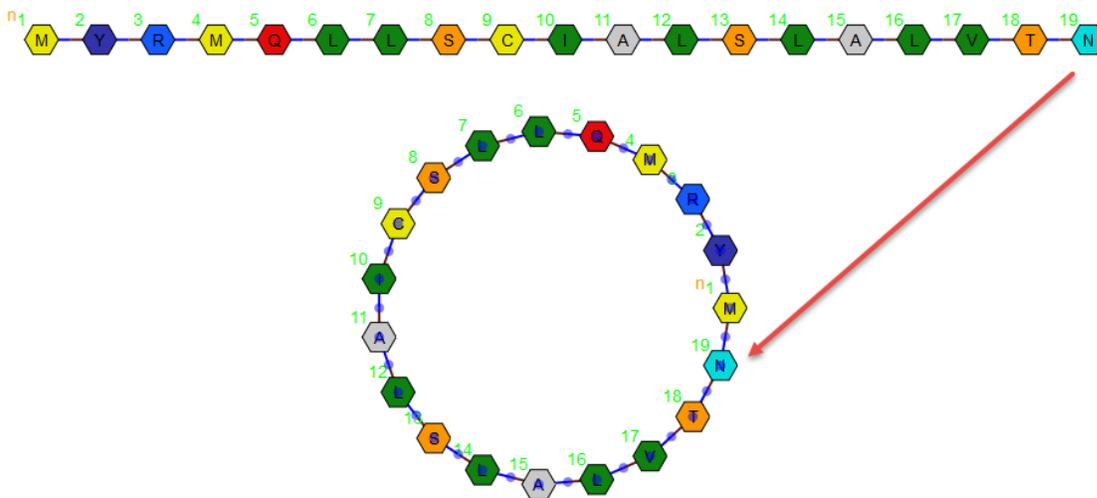
To create an intra-chain bridge, connect two monomers in the same sequence, that have more than two attachment points. Once the bond is added, the molecule will rearrange to form a cycle, and show the link as appropriately as possible.
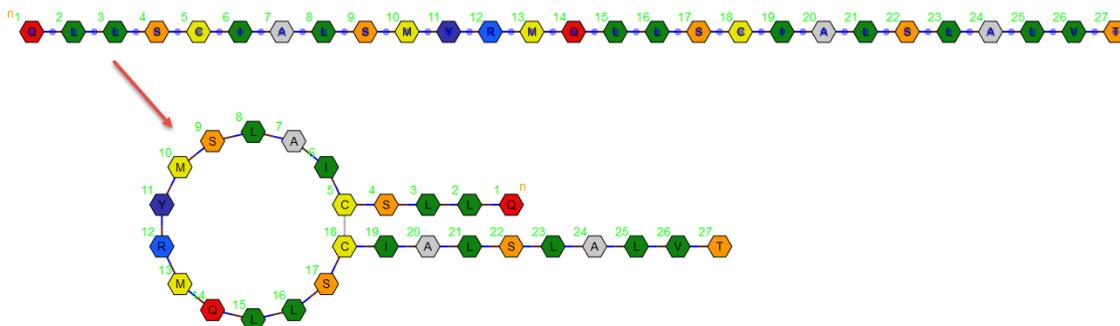
Saving and loading monomers

Once a new molecule is created or an existing monomer is modified, there are two formats to save and load the molecule data (using the Save and Load buttons in the Canvas), the HELM notation and the xHELM notation.

Helm notation

The HELM notation is the native data format of the HELM editor. The format is described extensively on the OpenHELM.org website. If you want to use your molecules mainly within your organization, HELM is the format of choice. It is a very compact and clear but it relies on a common set of monomers for the molecules.

Exchangeable HELM notation (xHELM)

The xHELM format was developed as the "portable" version of HELM. Like the HELM notation, xHELM is a unique notation. It contains the normal HELM notation in addition to the full set of monomers contained within the molecule. This is useful if you want to exchange your molecule data with other organizations, especially if you use non-standard monomers.

Managing monomers

HELM is a hierarchical notation where polymers are built from a predefined set of monomers. In the open source editor, these are stored in an XML file, but in an organization, these can be held in a central database.
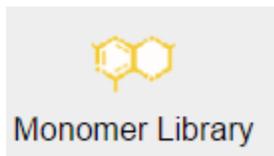
Add to the monomer set using the Monomer Library.

Using the Monomer Library

The Monomer Library is used to view the current monomer collection and register new monomers.

The Monomer library:

1. Monomer List Menu
2. Monomer List
3. Monomer information

Monomer List Menu

Hover over icons for identity key.

Refresh ⟳ to refresh the page.

New ✚ to add a new monomer to the list.

Edit ✎ to edit an existing monomer in the list.

Import Monomer XML Library 📂 to import an XML file.

Use the Symbol/Name box to search for an existing monomer via the dropdown list.

Use the Polymer Type dropdown list to filter monomers listed as RNA, CHEM or PEPTIDE.

Use the Monomer Type dropdown list to filter monomers listed as Backbone, Branch or Undefined.

Use the Count dropdown list to change the number of items displayed on a page.

Monomer List

Symbol	Name	Natural Analog	Polymer Type	Monomer Type	R1	R2	R3	Author	Created Date
MCC	4-(N-maleimidomethyl)cyclohexane-1-carboxylate		CHEM		OH				
A6OH	6-amino-hexanol		CHEM		H	H			
Az	Azide		CHEM	Backbone	OH			Fred	
Cys-BLOCK	Cys-BLOCK		CHEM		X				
SS3	Dipropanol-disulfide		CHEM		H	H			
EG	Ethylene Glycol		CHEM		H	OH			
Z	fvdasg		CHEM	Backbone	H	OH		Claire	
ds	gresye		CHEM	Branch	H				
hxy	Hexynyl alcohol		CHEM		H				
N-BLOCK	N-BLOCK		CHEM		OH				

To create a new monomer

1. Click on the 'New' button in the menu and a structure drawing editor will pop up.

2. Fill in the required information. Required information is indicated with a red asterisk.
3. Draw the molecule in the structure editor.
4. Make sure to choose the attachment points as R groups in the way recommended for that monomer type.
5. Click Save
6. The new monomer should now be available in the list.

Note: Both the monomer name and the symbol must be unique for the monomer to be registered. Please check carefully before registering.

Alternatively, an existing monomer can be used as a starting point.

1. Select the monomer to use as a template.
2. Click on the "Edit" button in the menu.
3. Change the structure to the new monomer.
4. Edit the monomer symbol and name.
5. Click Save.

Monomer information

Select a monomer in the list to display the available information in the "Monomer" window.

Managing Rulesets

Add to the Ruleset using the Ruleset icon in the HELM 2.0 top menu bar.
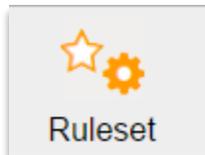
Using the Ruleset Manager

The Ruleset Manager is used to view the current ruleset and add new rules. These rules are then available to apply in the HELM canvas (see Rules under the Monomer Browser section).

Main page:

1. Rule Set: Window to add or edit rules and display existing rules.
2. Rule: Window to view selected rules from the Rule Set.

Rule Set

Hover over icons for identity key.

Refresh ⟳ to refresh the page.

New ⊕ to add a new rule to the list.

Edit ✎ to edit an existing rule in the list.

Use the Category box to choose category via the dropdown list or leave blank to display all items.

Use the Count dropdown list to change the number of items displayed on a page.

To add a rule to the Rule Set list:

1. Click on the 'New' button in the menu and a rule editor will pop up.

2. An ID will be automatically generated for the Rule.
3. Fill in the desired information.
   a. Category: Choose from the dropdown list or leave blank.
   b. Name: Enter the rule name.
   c. Description: Enter a rule description.
   d. Author: Enter the rule creators name.
4. JavaScript: Using the script provided, enter the JavaScript to implement the rule between the { }.
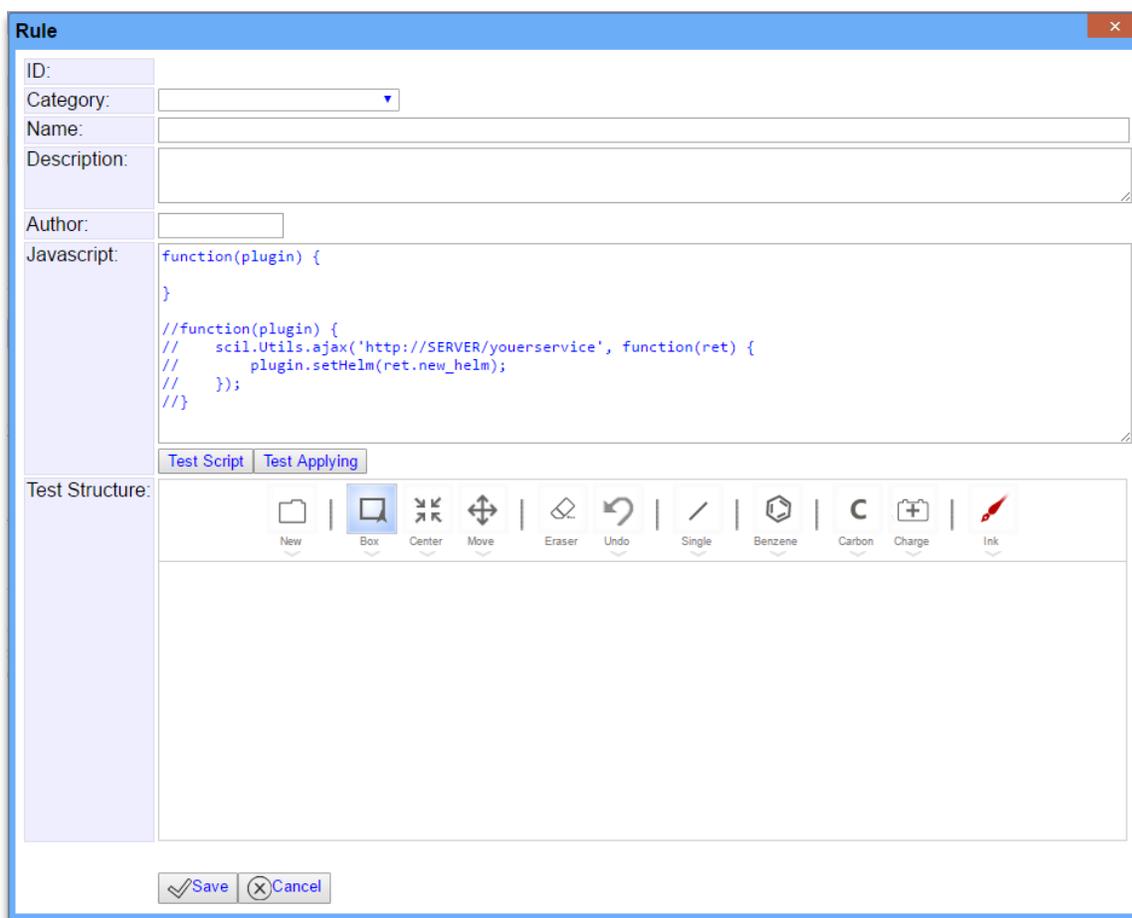
```javascript
Javascript:    function(plugin) {

               }
```

      a.   Test Script: Click Test Script for a check of the entered JavaScript.

      b.   Test Applying: Click Test Applying to test the rule application.

5. If desired draw a molecule in the Test Structure editor.
6. Click Save
7. The new rule should now be available in the Rule Set list.