# Proposal for a CARB polymer type in HELM

Table of Contents

[Proposal for a CARB polymer type in HELM 1](#_Toc62810891)

[Advisory panel 2](#_Toc62810892)

[Background 2](#_Toc62810893)

[HELM Principles 3](#_Toc62810894)

[CARB Polymer 6](#_Toc62810895)

[Monomer definition 6](#_Toc62810896)

[Polymer backbone construction 7](#_Toc62810897)

[Branching 8](#_Toc62810898)

[Branch prioritisation 9](#_Toc62810899)

[Monomer set 11](#_Toc62810900)

[Natural analogue 11](#_Toc62810901)

[Monomer naming 11](#_Toc62810902)

[Substituents 12](#_Toc62810903)

[Note about displaying substituents 12](#_Toc62810904)

[Display 13](#_Toc62810905)

[Complex Polymers 14](#_Toc62810906)

[CARB polymer to non-CARB polymer connections 14](#_Toc62810907)

[CARB polymer to CARB polymer connection 15](#_Toc62810908)

[Cyclisation 15](#_Toc62810909)

[Ambiguity 17](#_Toc62810910)

[Component Ambiguity 17](#_Toc62810911)

[Fully Unknown Monomer 17](#_Toc62810912)

[Partially specified monomers 17](#_Toc62810913)

[Unknown polymers 18](#_Toc62810914)

[Connection Ambiguity 18](#_Toc62810915)

[Unknown connection points 18](#_Toc62810916)

[Unknown stereochemistry at the anomeric carbon. 19](#_Toc62810917)

[Connections between a polymer and an unknown CARB polymer 19](#_Toc62810918)

[Composition Ambiguity 19](#_Toc62810919)

[Repeating groups – simple polymer (inc branches) 19](#_Toc62810920)

[Repeating groups – cyclisation 21](#_Toc62810921)

[Connection ambiguity 21](#_Toc62810922)

[Bibliography 23](#_Toc62810923)

**Version History:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Release Date** | **Author** | **History** |
| D 0.1 | 1st April 2019 | Claire Bellamy,  Tianhong Zhang | First draft version. |
| D0.2 | 14th April 2019 | Claire Bellamy,  Tianhong Zhang | Extended and includes review comments. |
| D0.3 | 30th April 2019 | Claire Bellamy,  Tianhong Zhang | Added repeating groups and monomer naming sections. Multiple edits for clarity. |
| D0.4 | 7th May 2019 | Claire Bellamy,  Tianhong Zhang | Minor tidying and circulated for review. |
| D0.6 | 1at January 2020 | Claire Bellamy,  Tianhong Zhang | Updates following wider glycan expert input and adjustments |

### Advisory panel

|  |
| --- |
| Martin Frank |
| Issaku Yamada |
| Kiyoko F. Aoki-Kinoshita |
| Frederique Lisacek |
| Philip Toukach |
| Thomas Lütteke |

## Background

The HELM notation was designed to be flexible and the ambitious aim was that it should be capable of handling all biomolecules: all polymer types and all variations on the natural monomers. The inventors recognised the need for a notation that included the ability to capture both monomer level information and the full atom/bond information.

So far, HELM has included three polymer types: PEPTIDE, RNA (which is really a general nucleotide type and includes DNA) and CHEM which allows any non-polymer chemical structure to be attached to the biomolecule.

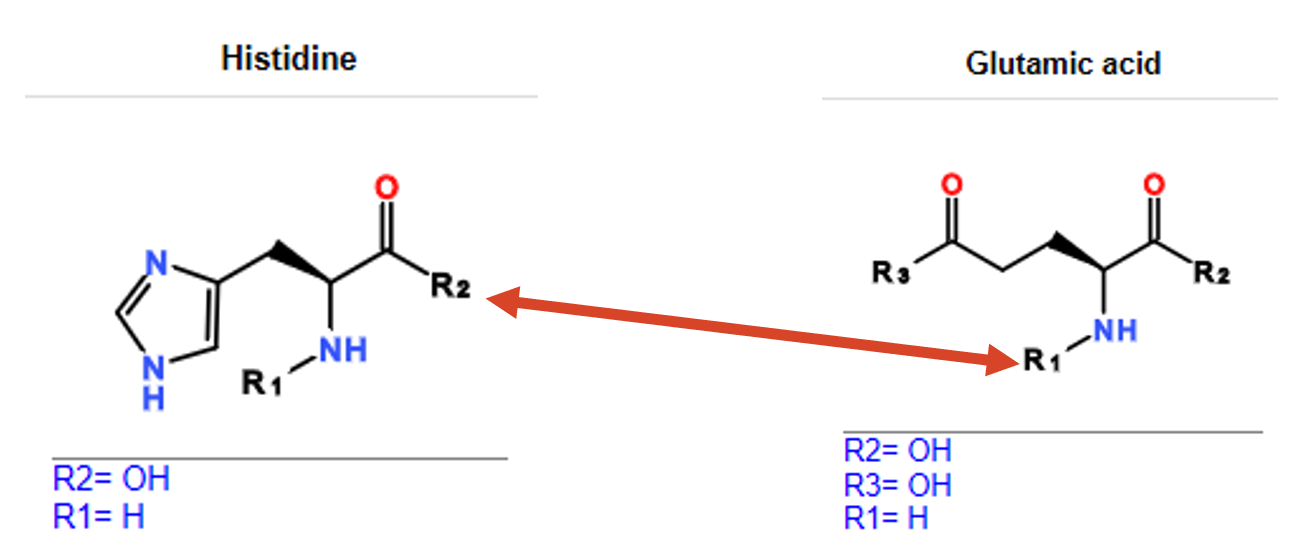
However, the design of HELM allows further polymer types to be added reasonably easily and the intention was always to do so. Polysaccharides and lipids are obvious candidates and neither have obvious solutions at present. For example, polysaccharides have rigorous and well-defined notations such as glycoCT and WURCS, but still have the issue of combining this with other polymer types. HELM would provide a mechanism for representing both.

Nevertheless, before a new polymer is added, specific rules for that polymer type need to be established and a suitable monomer set identified. This document sets out the issues and the proposed solution.

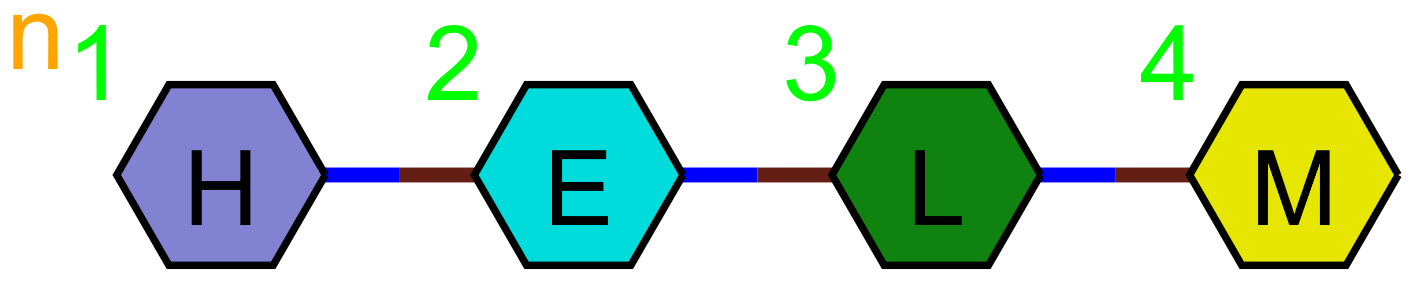
## HELM Principles

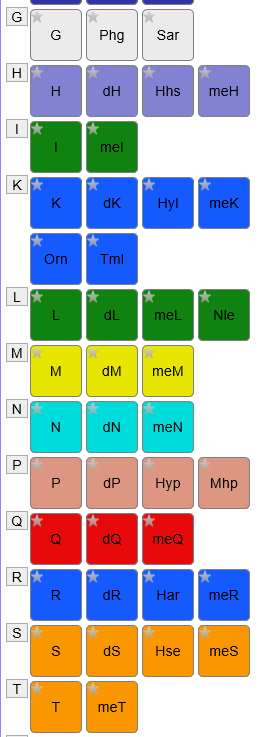
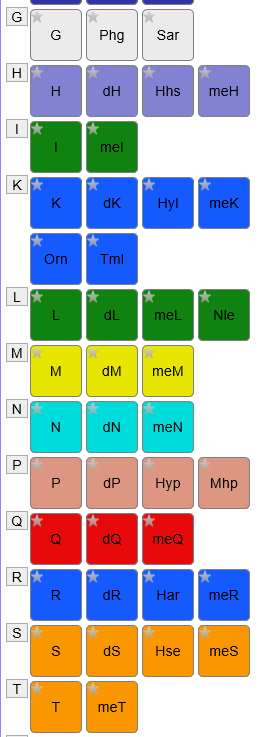
HELM polymers are made up of monomers. Each monomer has attachment points that are used to define how they are connected and there are two types of connection: a simple polymer connection that is defined as part of the polymer type and a complex polymer connection that is recorded in the second section of the HELM string.

For example: PEPTIDE1{H.E.L.M}$$$$V2.0 is a straightforward chain of 4 monomers: histidine, glutamic acid, leucine and methionine. The connection between each monomer is a standard chain extending connection, which for peptides is R2 of the left-hand monomer (the carboxylic acid) to the R1 of the right-hand monomer (the amine). HELM is directional, so the order cannot be reversed without changing the structure of the molecule.

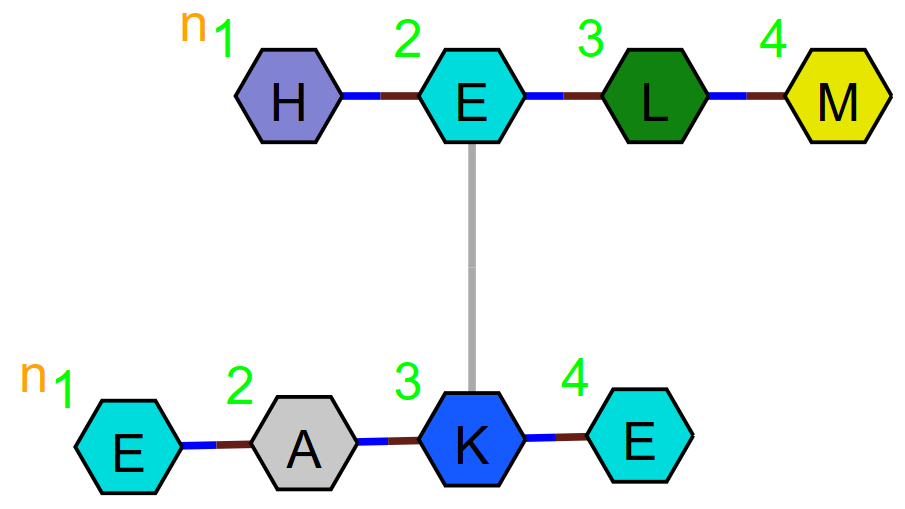


The biomolecule is displayed on the canvas as hexagons and the colour is determined by the natural amino acid. Analogues of natural amino acids share the same colour. The colours are only designed to be helpful rather than proscriptive, and some amino acids share the same colour since there are natural amino acids than colours used.





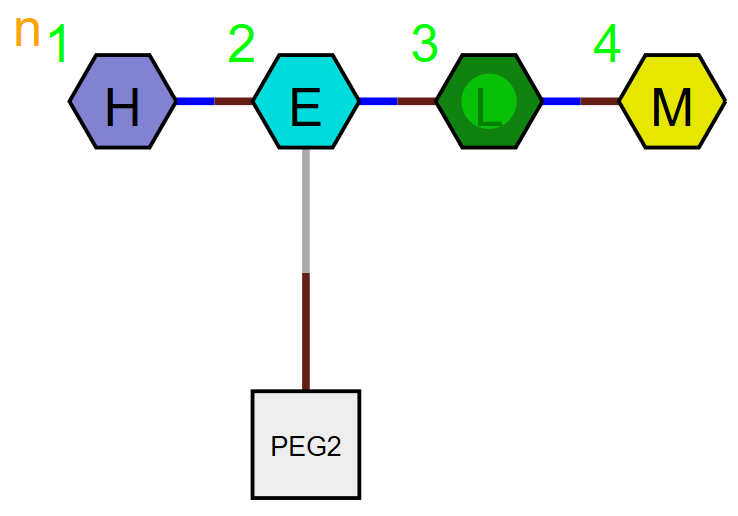
The display also numbers the monomers within a chain. Connected strands are numbered separately and the HELM string shows the connection point between them.



PEPTIDE1{H.E.L.M}|PEPTIDE2{E.A.K.E}$PEPTIDE1,PEPTIDE2,2:R3-3:R3$$$V2.0

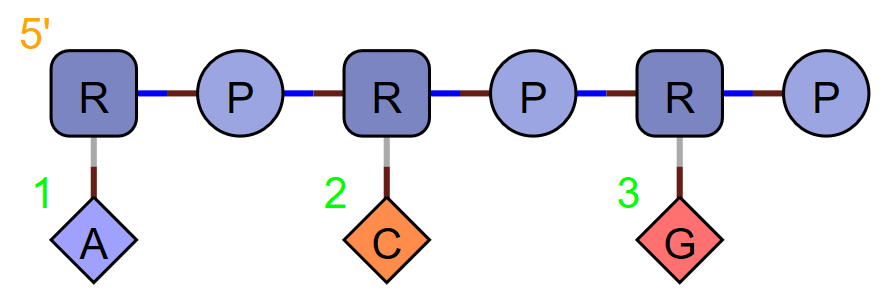
Inter-polymer connection points are defined after the first $ and include the polymer type and number, monomer number and the connection point number of each of the connected monomers.

The same section is used to connect other polymer types such as peptide to CHEM.



CHEM1{[PEG2]}|PEPTIDE1{H.E.L.M}$PEPTIDE1,CHEM1,2:R3-1:R1$$$V2.0

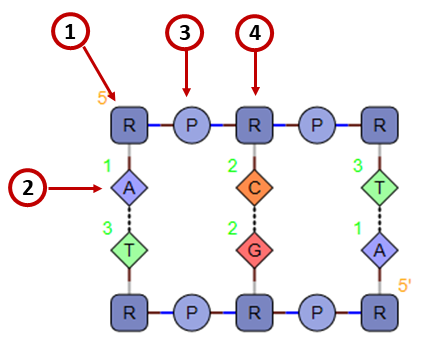
A Nucleotide monomer backbone is made up of the sugar and phosphate, and the base as a branch. The branch is indicated in the HELM string by placing the base inside brackets.



RNA1{R(A)P.R(C)P.R(G)P}$$$$V2.0

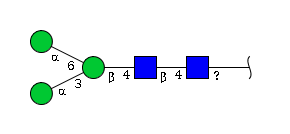
Nucleotides have a displayed numbering which counts the triad or base.

The monomer number used in the HELM string, starts from the left and counts all monomers in the molecule. In this example: the sugar, followed by the base then phosphate.



# CARB Polymer

Representing glycans and other polysaccharides involves addressing two major complexities. The n glycan motif below illustrates the issue.



Firstly. there are no consistent connection points. The anomeric position can connect in either an alpha or beta sense and may connection to any of the hydroxyls in the second monomer. This is unlike the peptide and nucleotide backbone types where the connections are consistent.

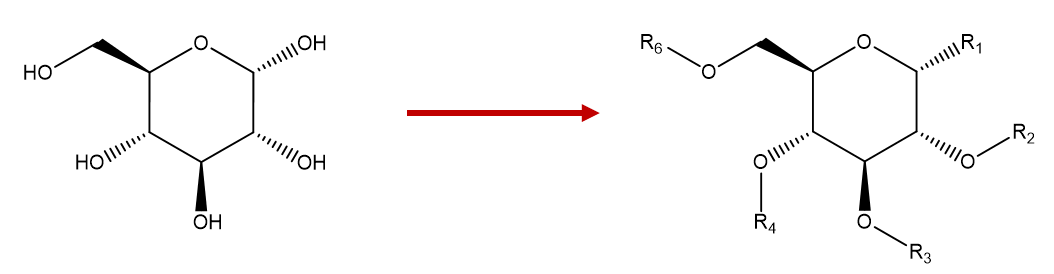
Secondly, there is significant branching. While HELM can handle branching, it does this by creating a new polymer strand and describing the connection in the second section. This is effective but could become cumbersome when there is extensive branching.

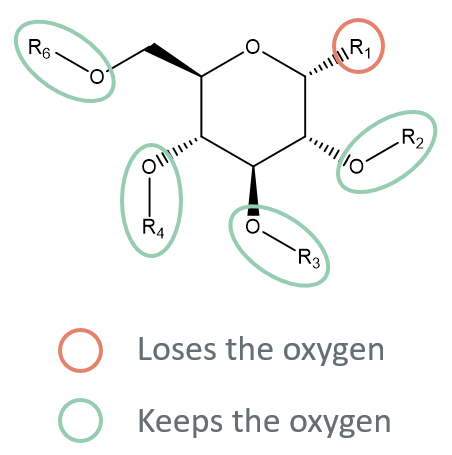
As a result, some changes are required to peptide and nucleotide polymer types.

## Monomer definition

The numbers used to indicate the connection between monosaccharides derive from a well-established numbering scheme. Essentially, numbering starts from the anomeric oxygen and continues clockwise around the ring.

For HELM monomers, we propose that monomer connection points are defined on all hydroxyls, and the R group number reflects the standard carbon numbering used in saccharide chemistry. This may mean that some R group numbers are not included, e.g. for glucose R4 and R6 are defined, but R5 is not. In other polymer types R groups must be consecutive, but this rule will be relaxed for CARB polymers.





A standard glycosidic linkage loses the oxygen at the anomeric position but retains the oxygen at the substitution point and we propose that HELM reflects the chemical process. This approach means that the polymer can form N (or other) links to peptides and other biopolymers, however substitutions elsewhere will retain the oxygen of the original hydroxyl. If a product contains an N (or other heteroatom) at that position, a different starting monomer must be used.

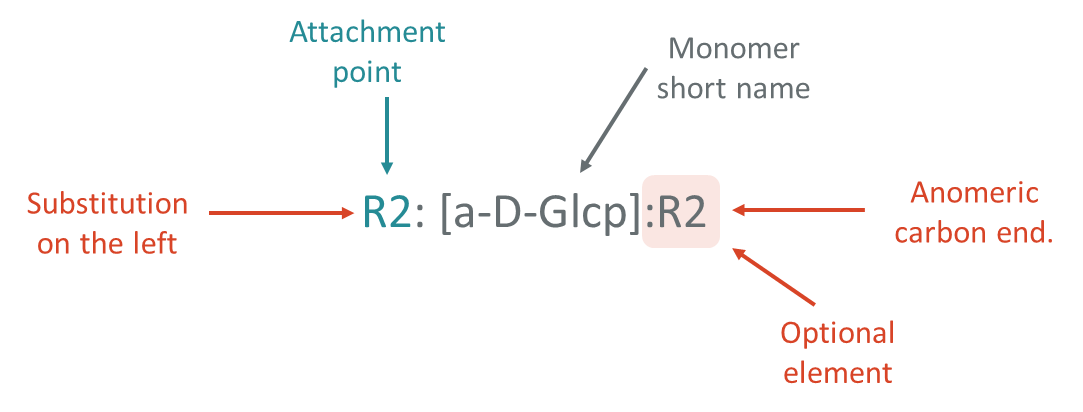
## Polymer backbone construction

The standard HELM string lists fragments separated by a ‘.’

For example:

PEPTIDE1{H.E.L.M}

We propose that CARB polymer also contains the substitution point of the non-anomeric hydroxyl and the connection orientation.



So (for aldoses) the HELM string looks like this:

CARB1{[a-D-Glcp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp]}$$$$

Non-aldose monomers, such as ketoses, will specify the anomeric centre after the monomer name. If nothing is specified, then R1 is assumed.

*Example containing a ketose:*

CARB1{[a-D-Glcp].R4:[a-D-Glcp]:R2.R6:[a-L-Galp].R3:[a-D-Glcp]}$$$$

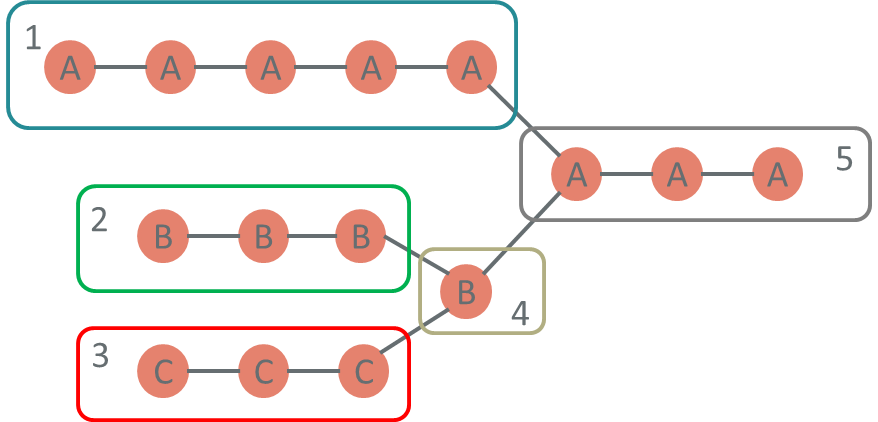
***Notes***

* The reducing end of a CARB is at the right, so the first monomer is connected to the second monomer at the anomeric carbon only, the last monomer is connected to the third monomer at the R3 position only and its anomeric end is not substituted. The middle monomers contribute both anomeric and non-anomeric attachments to the chain formation.
* The stereochemistry of the glycosidic linkage is given in the monomer definition.
* Unknown connection points are shown by a ‘?’

## Branching

HELM already allows branching in the nucleotide polymer type, where branch monomers are placed inside round brackets. However, until now branches have only contained single monomers. We propose that branches are extended to allow more than one monomer and branches within branches.

The following is an example multi-branched structure.



The CARB structure would look like this (HELM has been simplified to omit the connection points and stereochemistry).

CARB1{A.A.A.A.A(B.B.B(C.C.C).B).A.A.A}$$$$

and the full HELM string would look like this (indents are for clarity and are not required or significant)

CARB1{

[A].R3:[A].R2:[A].R2:[A].R2:[A]

([B].R2:[B].R2:[B]

([C].R2:[C].R2:[C].R6).

R2:[B].R6).

R3:[A].R4:[A].R4:[A]

}$$$$

Reading from left to right, the monomer branches converge (not diverge). This approach keeps the reducing end on the right-hand side of the notation in accordance with convention. The trisubstituted monomer at the convergence point is written after all converging branch monomers.

The connection points are still defined to the left of the monomer, but the beginning of a branch does not require a connection point, since it is only connected at the anomeric position. The monomer at the convergence point has multiple substitutions. The main chain substitution point is shown next to the monomer and the substitution points used by the branches are shown at the end of the branch section.

In full HELM notation and focussing on a single branch point, the atom/bond structure looks like this:

CARB1{[a-D-Glcp].R3:[a-D-Glcp] ([a-D-Glcp].R2)}$$$$



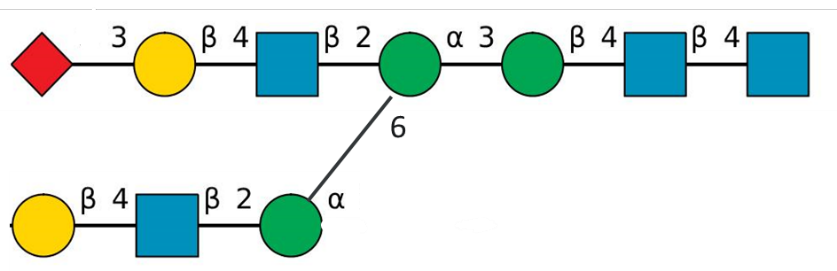
A more complex, but still real-life example would be

CARB1{[Sia].R3:[b-Gal].R4:[b-GlcNAc].R2:[a-Man]

([b-Gal].R4:[b-GlcNAc].R2:[a-Man].R6)

.R3:[d-Man].R4:[b-GlcNAc].R4:[b-GlcNAc]

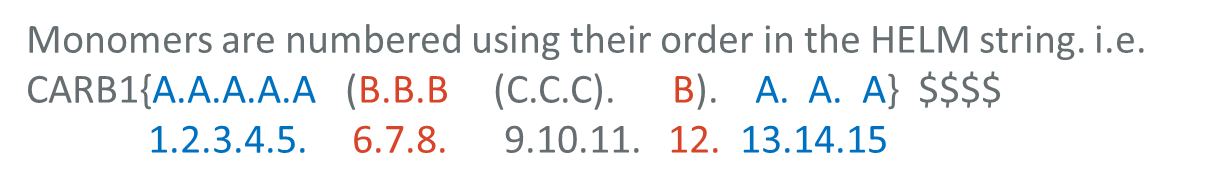
}$$$$



### Branch prioritisation

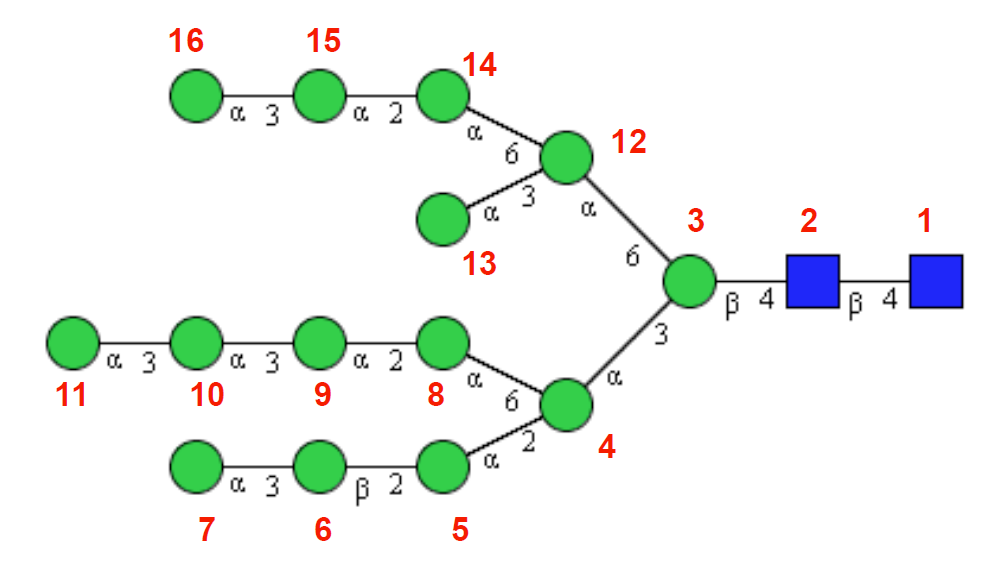
Monomers are numbered form left to right, as for other polymer types in HELM. i.e.

CARB1{A.A.A.A.A.(B.B.B.(C.C.C)B)A.A.A}$$$$ is numbered as:



This approach is useful for identifying monomers within a given HELM string, but since there are many ways of ordering the branches, the same molecule can have different HELM strings. This is acceptable for standard HELM, but canonical HELM must have rules to identify which branch should be defined first.

The linkage position is commonly used to define the priority of each branch. Under this scheme, you number from the branch with the lowest connection point number at the branch point. You then number the whole branch before returning to the more recent branch point and progressing along the branch with the next highest connection point. See below:



If the HELM string is ordered according to this priority the string will be unique and we recommend this for creating canonical HELM.

If the connection points are ambiguous, it may not be possible to determine the order. In this case the longest branch will be the higher priority followed by alphabetical order of the connected monomers.

An example of the full notation would look like this with the branch highlighted in yellow and the main chain in blue:

CARB1{[a-D-Glcp].R3:[a-D-Glcp]([a-D-Glcp].R6:[a-L-Galp].R2).R3:[a-D-Glcp]}$$$$

# Monomer set

Monosaccharide DB and SNFG have been used to guide the choice of monomers for an initial set.

The hexoses Glc, Man etc… generally have 10 different forms in monosaccharideDB and these can be replicated in the HELM monomer set.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **D/L** | **pyranose** | **furanose** | **Open chain** |
| **Alpha** | **D** |  |  |  |
|  | **L** |  |  |  |
|  | **achiral** |  |  |  |
| **Beta** | **D** |  |  |  |
|  | **L** |  |  |  |
|  | **achiral** |  |  |  |

However, not all forms of substituted or deoxy monosaccharides are present. For example, the following are the entries for AllNAc:

* b-D-AllpNAc
* a-D-AllpNAc6Me

Only 2 forms are present for the default substitution and one of them has been further substituted with a methyl at the 6 position.

Apart from the hexoses, the initial HELM monomer set will only list the monomers present the in SNFG monomer table. Further monomers can be added as needed and on request.

### Natural analogue

Monosaccharides have a very large number of natural monomers and therefore the concept of ‘natural analogue’ does not directly translate to this polymer type. The CARB monomer type will not have a natural analogue.

## Monomer naming

HELM will adopt the CarbBank naming convention for monomers. This is based on the IUPAC three letter codes and includes information about the ring size and substituents. Some examples are shown below:

|  |
| --- |
| [a-D-GalpA](http://www.monosaccharidedb.org/display_monosaccharide.action?id=21) |
| [b-D-Fucp](http://www.monosaccharidedb.org/display_monosaccharide.action?id=22) |
| [a-L-4-en-4-deoxy-thrHexpA](http://www.monosaccharidedb.org/display_monosaccharide.action?id=23) |
| [a-D-GlcpA](http://www.monosaccharidedb.org/display_monosaccharide.action?id=24) |
| [b-D-Galp3Me](http://www.monosaccharidedb.org/display_monosaccharide.action?id=25) |
| [b-L-Fucp](http://www.monosaccharidedb.org/display_monosaccharide.action?id=26) |
| [b-D-6-deoxy-Glcp](http://www.monosaccharidedb.org/display_monosaccharide.action?id=27) |
| [b-L-Rhap](http://www.monosaccharidedb.org/display_monosaccharide.action?id=28) |
| [b-D-2,6-deoxy-ribHexp](http://www.monosaccharidedb.org/display_monosaccharide.action?id=29) |
| [a-D-Rhap](http://www.monosaccharidedb.org/display_monosaccharide.action?id=30) |

These names are longer than typical for other HELM polymer types. Peptide, chem and nucleotide polymers display the names on the canvas, so shorter names are beneficial. Carb polymers will follow the SNFG notation so most symbols will be empty and the name length is less important.

## Substituents

Monosaccharides may be substituted by a variety of moieties.

Substitutions which are part of the SNFG approved list will be handled as new monomers. Other substitutions should be attached as CHEM monomers in the same way as other polymer types.

The list of substitutions at the time of writing are:

|  |  |
| --- | --- |
| **Abbreviation** | **Substituent** |
| Ac | acetyl |
| S | sulfate |
| P | phosphate |
| Me | methyl |
| Lt | lactyl |
| 4,6Py | 4,6-pyruvyl |

SNFG should be consulted to obtain the latest set.

### Note about displaying substituents

#### O Substituents

In accordance with note 7 of the SNFG nomenclature, abbreviations for approved substituents should be displayed outside the monomer shape. The abbreviation should be preceded by the position number.

#### N Substituents

N substituents are preceded by the letter N when displayed. If the position is not the most common position, the abbreviation is preceded by the position number.

## Display

HELM has established shape and colour conventions for the pictorial representation of each polymer type.

CHEMs are grey squares:



Peptides are hexagons coloured according to the natural analogue:



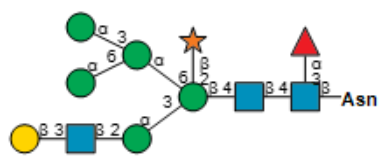
And for nucleotides: phosphates are circles, sugars are rounded squares and bases diamonds.



The monomer symbol is displayed within the shape.

The glycan community have a well-established convention for the pictorial display of structures: [Symbol Nomenclature for Glycans](https://www.ncbi.nlm.nih.gov/glycans/snfg.html) (SNFG). This is recommended for submission to major journal and other publications.

This nomenclature specifies the shape and colour for each monomer. It also specified bond labels to show the connection point and orientation. For example:



The appearance is sufficiently different from the other polymer types, since the monomer symbol is not displayed and the linkage information is shown, that it should be clear that this is a glycan and not one of the other polymers. We recommend that HELM uses SNFG and fully complies with the format.

There are detailed rules, some of which require characters inside and above/below the shape. Characters are used in the following situations.

*“Less common configurations need to be stated in a figure legend or by adding the letters inside the symbol (e.g., adding D or L to the symbol). Epimers at C8 of nonulosonates can be indicated by adding "8D" or "8L" inside the symbol. Furanose rings can be indicated by adding an italicized "f" inside the symbol, and alditols can be indicated with an italicized "o" inside the symbol.”*

*“For amino sugars in which the nitrogen is not at the most common carbon, add a number to the N (e.g., Rha4N is shown as a green triangle with 4N attached). Additionally, atypical acetamido groups may be represented using NAc (e.g. Fuc4NAc is shown using red triangle with 4NAc attached).”*

These complicate the implementation, but do not prevent the use of SNFG.

# Complex Polymers

## CARB polymer to non-CARB polymer connections

Connections between simple polymers are handled in the second section of the HELM string. CARB connections are subtly different since one of the connection points may already be used in the simple polymer, something that is not done for other polymer types.

A standard peptide to chem molecule would look like this;

CHEM1{[MCC]}|PEPTIDE1{H.E.L.M}$CHEM1,PEPTIDE1,1:R1-1:R1$$$V2.0

A carb to chem connection at monomer 2 would be:

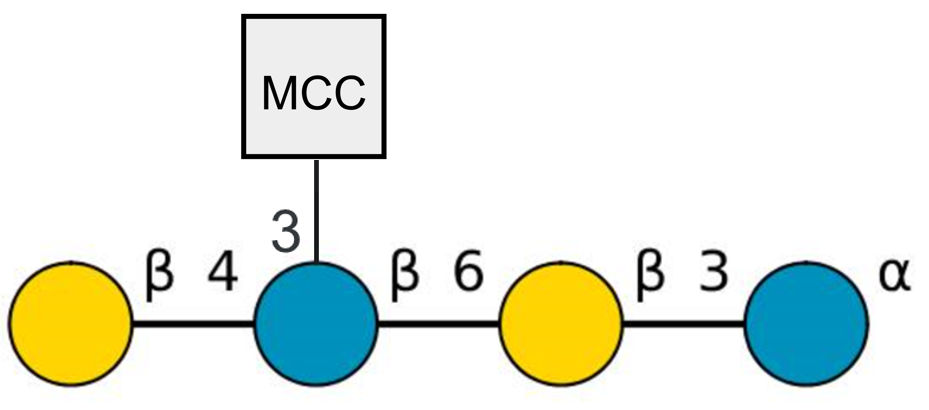
CHEM1{[MCC]}|

CARB1{[a-L-Galp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp]}

$CHEM1,CARB1,1:R1-2:R3

$$$V2.0

MCC is connected to the highlighted monomer via the R3 connection point and a-L-Galp is connected via the R4 connection point.



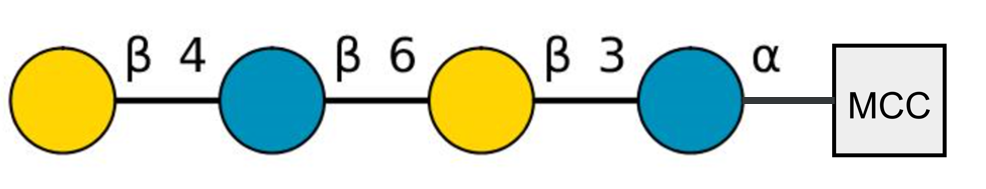
The connection may also be made to the anomeric carbon if it is not already used.

CHEM1{[MCC]}|

CARB1{[a-L-Galp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp]}

$CHEM1,CARB1,1:R1-4:R1

$$$V2.0



In this case you can think of EG as terminal sugar monomer without an anomeric attachment point, however, it is still defined as a CHEM monomer in HELM.

Care must be taken when substituting at the R1 position as there is no retained oxygen (unlike the other positions) and the HELM user must take this into account.

## CARB polymer to CARB polymer connection

Most CARB to CARB connections will be handled within the branching available in a simple polymer butCARB to CARB connectioncan be handled in the second section of the HELM string as for CARB to non-CARB polymer connections.

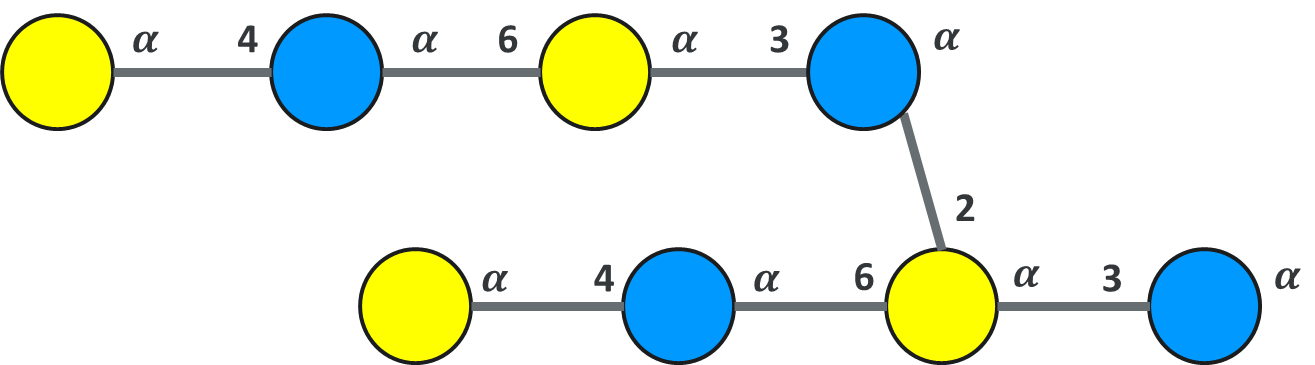
This example is valid HELM but would normally be written as a single simple polymer.

CARB1{[a-L-Galp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp]}|

CARB2{[a-L-Galp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp]}

$CARB1,CARB2,4:R1-3:R2

$$$V2.0



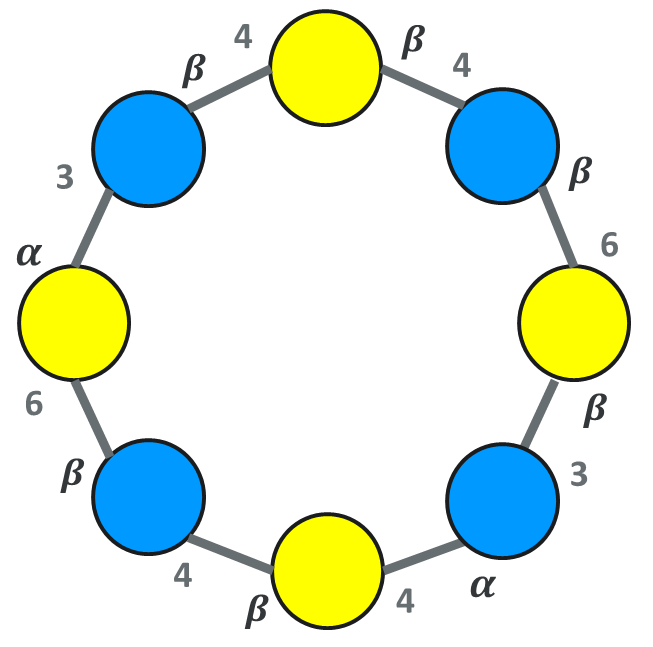
## Cyclisation

Cyclisation is a CARB to CARB connection, but the simple polymer is the same for both ends.

CARB1{[a-L-Galp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp].[a-L-Galp].R4:[a-D-Glcp].R6:[a-L-Galp].R3:[a-D-Glcp]}

$CARB1,CARB1,1:R4-8:R1

$$$V2.0



# Ambiguity

Ambiguity in HELM is defined in the HELM specification and takes several forms: component, connection and composition. Most principles can be adopted by the CARB polymer type directly, so the following sections describe the most relevant rules and any extensions required for CARB polymers. See the HELM specification for further information on HELM and ambiguity.

## Component Ambiguity

### Fully Unknown Monomer

X = A single unknown monomer

\* = 1-n unknown monomers

These should be shown as flattened white hexagons in line with the SNFG recommendations.

The X or \* may be included inside the hexagon.

Connection points are always unknown for unknown monomers and should not be included in the string.

e.g.

CARB1{[a-L-Galp].\*.R6:[a-L-Galp].R3:[a-D-Glcp].[a-L-Galp].X.R6:[a-L-Galp].R3:[a-D-Glcp]}

$CARB1,CARB1,1:R4-8:R1

$$$V2.0

### Partially specified monomers

Some monosaccharides have unknown stereochemistry, but other attributes such as the ring size and atoms are known. HELM can capture unknown stereochemistry in the atom/bond representation, and these monosaccharides are named in monosaccharide DB, so they should be used in the same way as fully specified monomers in HELM. White shapes are used for display in accordance with the rules set out in SNFG.



### Unknown polymers

An unknown CARB polymer should consist of single ‘\*’ monomer.

CARB1{\*}$$$$

This denotes 1-n unknown monomers and therefore can substitute for a polymer of type CARB where nothing is known about the individual monomers.

## Connection Ambiguity

Standard HELM rules about unknown monomer locations and connections to a specific monomer symbol apply. The following are specific to the CARB polymer type.

### Unknown connection points

If the connection point is unknown: ‘?’ is used instead of the number.

e.g. R?:[a-D-Glcp]

Standard SNFG recommendations will be used for the display I.e. the ‘?’ will be displayed in place of the connection point number

### Unknown stereochemistry at the anomeric carbon.

Unknown stereochemistry should be shown in the monomer structure and is not part of the HELM notation.

Standard SNFG recommendations will be used for the display I.e. the ‘?’ will be displayed in place of the alpha/beta.

### Connections between a polymer and an unknown CARB polymer

Bonds to unknown polymers must have unknown connection points i.e. you cannot have a specified R group. The monomer number can be specific or unknown.

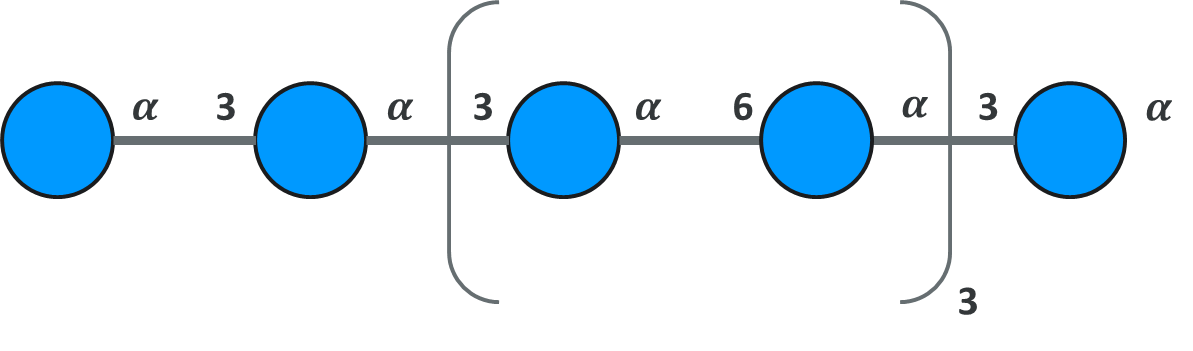
CARB1{\*}|PEPTIDE1{A.C.D}$PEPTIDE1,CARB1,2:R3-1:?$$$V2.0

## Composition Ambiguity

### Repeating groups – simple polymer (inc branches)

In HELM repeating groups are contained within round brackets followed by the number of repeats inside single quote marks. E.g.

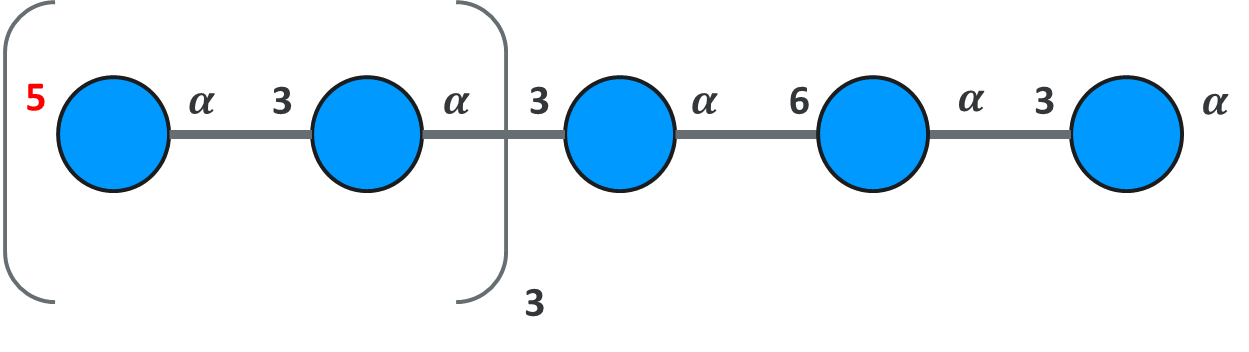
CARB1{[a-D-Glcp].R3:[a-D-Glcp].(R3:[a-D-Glcp].R6:[a-L-Galp])’3’.R3:[a-D-Glcp]}$$$$



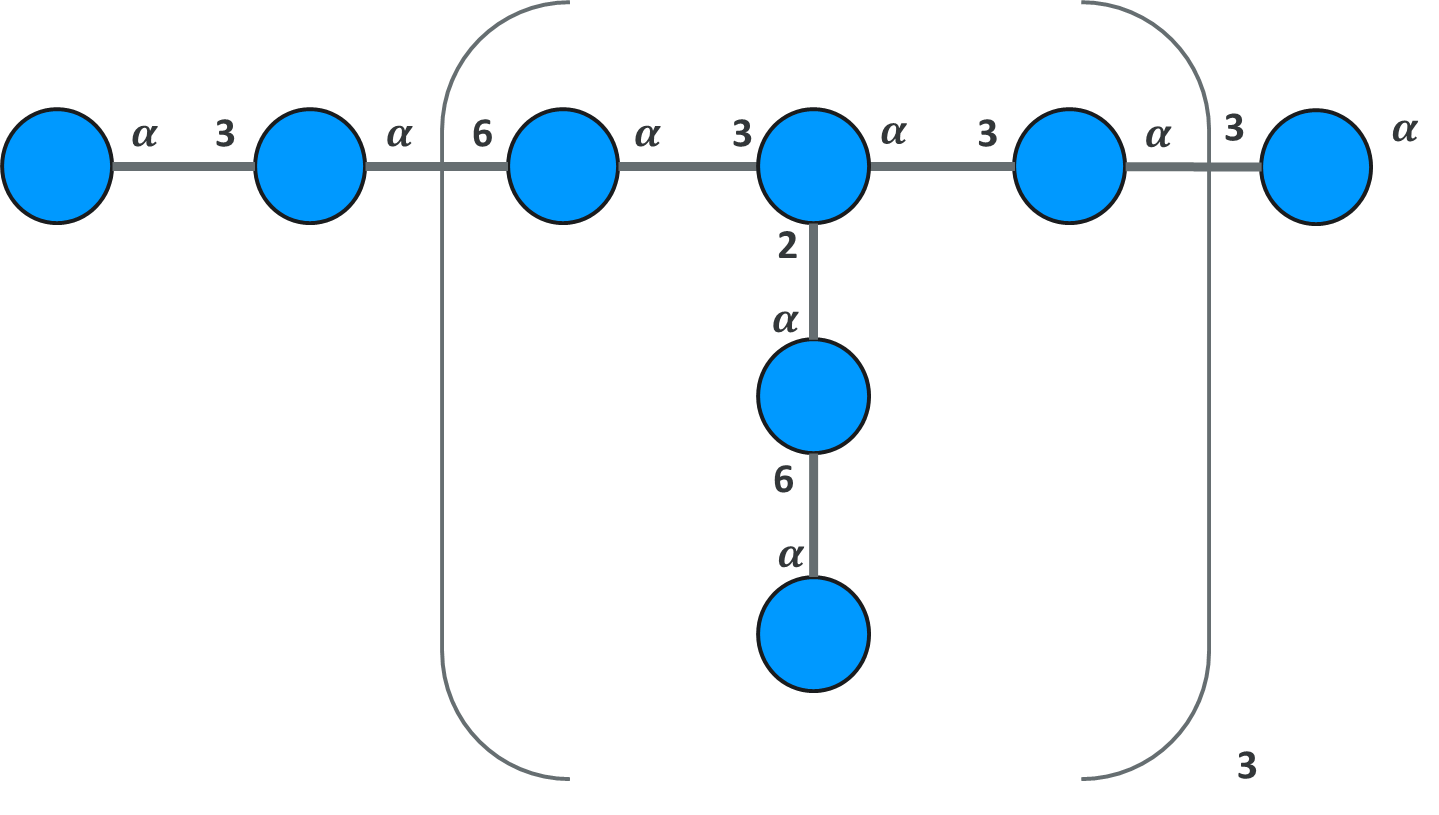
The branch is highlighted in blue, with the repeating group in yellow. Note that the only differentiation between the branch and the repeating group is the number in quotes following it.

If a repeating group includes the start of a polymer, the first monomer must have a connecting group defined, so we know how to connect the monomer at the end of the repeat to the beginning. This connection point would not normally be defined.

CARB1{(R2:[a-D-Glcp].R3:[a-D-Glcp].(R3:[a-D-Glcp].R6:[a-L-Galp])’3’.R2)R3:[a-D-Glcp]}$$$$



CARB polymers are unique in HELM in that the simple polymer can contain branches with multiple monomers. It is possible to repeat a section of backbone that also contains branches.



CARB1{[a-D-Glcp].R3:[a-D-Glcp].

(R6:[a-D-Glcp].

([a-L-Galp].R6:[a-L-Galp].R2)

R3:[a-D-Glcp]) ’3’.

R3:[a-D-Glcp]. R3:[a-D-Glcp]}$$$$

A branch must be entirely contained within the repeating group.

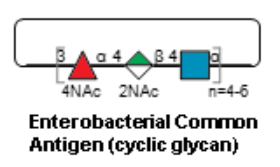
Repeats can be integers, integer ranges (n-m), unspecified or a range where one of the values is unspecified (e.g. 2-n).

The exact suffix differs slightly from SNFG, but is identical in meaning.

|  |  |  |
| --- | --- | --- |
| **Type** | **HELM suffix example** | **SNFG example** |
| Specific Integer | 2 | n=2 |
| Non-specific integer | n | n |
| Specific range | 2-4 | n=2-4 |
| Non-specific range | 9-n | n>8 |

### Repeating groups – cyclisation

Repeating groups can be defined within a cyclic structure even if they represent the whole of the monomers within the ring. For example:



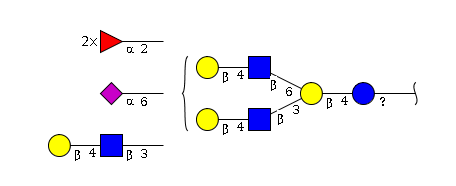
Can be represented as:

CARB1{(R3:[b-D-Fucp].R4:[b-D-ManpA].R4:[b-D-GlcpNAc])‘4-6’}$CARB1,CARB1,1:R4-3:R1$$$V2.0

HELM numbers monomers according to their position in the HELM string, so [b-D-GlcpNAc] is monomer 3 despite being part of a repeated cyclic structure.

## Connection ambiguity

A common feature of glycans is that the position of an substituent may be unknown see below:



We propose HELM handles this by creating a group for the main section, listing the possible attachment points. Substituents may be attached to any of the listed attachment points in the group. NB Listing attachment points is an extension to the HELM notation.

Examples below:

Specific fragment attachment point to undefined position on the main glycan

CARB1{[Sia]:R2}|

CARB2{[Neu5Ac]:R2}|

CARB3{[b-L-Gal].R4:[b-L-GlcNAc]}|

CARB4{[b-L-Gal].R4:[b-L-GlcNAc]

([b-L-Gal].R4:[b-L-GlcNAc].R6)

.R3:[b-D-Gal].R4:[b-Glc].}

$CARB1,CARB4, 1:R2-[?:?]

|CARB2,CARB4,1:R6-[?:?]

CARB3,CARB4,2:R3-[?:?]

$G1{(CARB1:2+CARB4:1)$$V2.0

# Bibliography

Symbol Nomenclature for Graphical Representation of Glycans, *Glycobiology* 25: 1323-1324, 2015.

Zhang T, Li H, Xi H, Stanton R V, Rotstein S H, HELM: A Hierarchical Notation Language for Complex Biomolecule Structure Representation. *J. Chem. Inf. Model*. 52, 2796–2806(2012).

Böhm M, Bohne-Lang A, Frank M, Loss A, Rojas-Macias MA, Lütteke T Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update) *Nucleic Acids Res.* 2019, 47(D1):D1195-D1201

Frank, Martin and von der Lieth, Claus-Wilhelm and Lütteke, Thomas. (2009). *Bioinformatics for Glycobiology and Glycomics: An Introduction*