HELM Line Notation Extension - Ambiguity

Contents

Introduction 3
HELM Notation – new elements 3
Inline annotations 6
  Monomer annotation 6
  Simple polymer annotation 6
  Connection annotation 6
Monomer ambiguity 6
  Missing monomer 6
  Single Monomer – no probability 7
  Single Monomer with probability 7
  Monomer mixture 7
  Monomer mixture with ratios 7
  Unknown monomer 8
  Repeating monomers with defined count 8
  Repeating monomers with range count 8
Connection ambiguity 8
  Connected monomer type is known, position unknown 8
  Connection partner/monomer is undefined 9
  Connection involves a simple polymer OR group 9
  Connection involves a simple polymer AND group 9
  Binding ratio – Composition Ambiguity 10
Component ambiguity 11
  Sequence unknown, type of polymer not defined 11
  Sequence unknown, type of polymer known 11

Sequence partially known 11
Bead coupling 11
Nanoparticles 12
Component and connection and composition ambiguity 12
Glycosylation 12
No ambiguity 12
Component ambiguity - Glycosylation not fully defined. 13

Introduction

The HELM string is extended to host all structurally relevant information within the line notation. Besides direct annotations of structurally important parts (formed by "") all extended annotations reside in the annotation section.

Section 4 of HELM notation is intended to capture additional annotations for the macromolecule structure, but these annotations are not considered structurally important. There is a proposal to standardize the annotations on JSON format, and any structure elements can be referenced from JSON. However, from an ambiguity support point of view, this is considered optional.

HELM Notation – new elements

The following new elements are used within this document:

- Unknown Monomers:
  - Unknown monomers "*": represents 0..n monomers, monomer count = 0..n
  - Unknown single monomer "X"/"N": "X" represents single amino acid monomer in PEPTIDE, "N" represents single base monomer in RNA, count =1
  - Single deleted/missing monomer is marked using an underscore "_": monomer count =0

- List elements:
  - Both monomers and simple polymers can be put into a list
  - List elements are grouped using parentheses ()
  - Comma "," represents OR relationship, only one single element of the list is possible i.e no mixture but undetermined identity of a element. Probabilities of occurrence can be assigned to the element of the list. These probabilities are separated by colons.
  - Plus "+" represents AND relationship, all elements in the list are possible and thus form a mixture. Ratio of each element can be specified after colon, and default is 1 when omitting.

- List of Monomers
  - List of Monomers are handled in the simple polymer notation using above syntax
  - (K,R,H), at the given monomer position, only one of the three monomers in the list is possible. The default probability is 1/3
  - (_,K), at the given position, we could either have monomer K or nothing. The probability of either one is 50%

    - (K:45, C:55), at the given monomer position, we could have either K or C, and the probability for K is 45%, and the probability for C is 55%

- List of Simple Polymers
    - Simple polymer grouping info will be handled in section three of HELM notation, which was used to handle hydrogen bonding info in HELM 1. Groups are separated by vertical pipe | , which is the same as in simple polymer list and connection list.
    - Hydrogen bonding connections will be moved to the connection section (Section 2) in HELM 2, as it is just a special kind of connection.
    - Each simple polymer group can contain two or more simple polymers, and will be assigned a group ID such as G1, G2…, which can be referenced in HELM2 notation. Grouping description follows the same syntax as for monomer list
    - G1(PEPTIDE1+PEPTIDE2+PEPTIDE3+PEPTIDE3), in G1, we have a mixture of four PEPTIDE polymers. The default ratio for the mixture is 1:1:1:1
    - G2(PEPTIDE1:2+PEPTIDE2:2+CHEM3), in G2, we have a mixture of two PEPTIDE polymers and one CHEM polymer. The ratio is 2:2:1
    - G1(PEPTIDE1+CHEM1)|G2(RNA1+CHEM2)|G3(G1:2.5+G2) we have three groups here. In G1, we have a mixture of PEPTIDE1 and CHEM1, 1:1 ratio. In G2, we have a mixture of RNA1 and CHEM2, 1:1 ratio. In G3, which is a super group, we have a mixture of G1 and G2, with a ratio of 2.5:1.
    - Simple polymer grouping will be used to represent nanoparticle formulations.

- Inline Annotations
    - Inline annotations are marked with quotation marks “”. They are always located after the element, i.e. before the separator of the next element or section
    - Inline annotations can be applied to monomers, simple polymers and connections
    - Extended annotations are using a JSON format as described in a separate document

- Unknown simple polymer types
    - Use BLOB as unknown polymer type
    - Specify polymer type inside curly braces
    - Specify polymer name with inline annotation
    - BLOB1{Bead}”Aminated Polystyrene”$$$$ represents animated polystyrene bead. Simple polymer type is “Bead”, and simple polymer name is “Animated Polystyrene”

- Simple polymer with unknown structure/sequence
    - Put * inside curly braces

    - Specify polymer name with inline annotation
    - PEPTIDE1{*}”IL6”$$$$ represents a peptide chain with the name IL6.

- Monomer repeating units
    - In single polymer, a monomer or a fragment can repeat itself , and sometimes the repeating units could be a range
    - We can add the repeat unit count immediately after the monomer or fragment and enclosed it with single quote ‘ ‘
    - A’4’, we have 4 monomer A in the chain, equivalent to A.A.A.A
    - A’23-35’, we have 23-25 repeating As in the chain, exact number unknown.
    - (R(A)P.R(G)P)’15’, there are 15 repeating fragments of R(A)P .R(G)P

- Connection Ambiguity
    - Source and target polymers, instead of just one simple polymer from the simple polymer list in HELM 1, can be simple polymer groups as specified in HELM notation section 3
    - Monomer position can be text to describe monomer type, such as K for lysine, or * for unknown
    - R group can be * for unknown

- Extended Annotation
    - In HELM 2, section 4 of HELM string has to be in valid JSON format
    - Compare the difference between HELM 1 and 2 in annotation section:
      HELM 1: PEPTIDE1{hc}|PEPTIDE2{lc}
      HELM 2: {“PEPTIDE1”:{
      “ChainType”:”hc”},”PEPITDE2”:{“ChainType”:”lc”}}

- Version Number
    - Add V2.0 to the end of HELM notation for easy parsing.
    - Missing version indicates V1.x
    - PEPTIDE1{}$$$$V2.0

Inline annotations

While most inline annotations are not structurally important, they are important for BLOB polymer types, and for polymers without structure info. The annotations are marked with quotation marks (“”). They are located at the end of a monomer, simple polymer or connection – directly in front of the separator used there.

Monomer annotation

A Cystein is annotated as a mutation, annotation is not structurally important.

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C”mutation”.D.D.D.D.D.D.D.D.D.D.D.
D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}
```

Simple polymer annotation

The peptide chains are annotated as LC and HC, annotations are not structurally important

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}”HC”|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.X.X.X.X.X.X.K.K.
K}”LC”$$$$
```

A undefined peptide is annotated as IL6. Annotation is structurally important in case structure info is unknown. “IL6” is all you know about this PEPTIDE polymer.

```
PEPTIDE1{*}”IL6”$$$$
```

Connection annotation

The connection between PEPTIDE1 and CHEM 1 is annotated as “Specific Conjugation”.

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHE
M1{[*]SCCCc1ccccc1 |$_R1;;;;;;;;;;$|}$PEPTIDE1,CHEM1,C:R3-1:R1”Specific Conjugation”$$$
```

Monomer ambiguity

Missing monomer

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.

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.(_,K)}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}
$$$$
```

Single Monomer – no probability

Any single monomer of a list may be found at given position. The list elements are separated by a comma and grouped using parentheses. No probability for the occurrence of a distinct monomer is given.

```
PEPTIDE1{A.A.A.A.(A,G,W,[Aha]).A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.
D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.
K.K.K.K}$$$$
```

Single Monomer with probability

Any single monomer of a list may be found at given position, probability of occurrence of given monomer within that list is known, but product contains exactly one monomer.

```
PEPTIDE1{A.A.A.A.(A:1.1,G:69.5,W:25.5,[Aha]:3.9).A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.
D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2
{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.
K.K.K.K.K.K.K.K.K.K.K.K.K.K.K.K}$$$$
```

The chromatogram of a sequence doesn't allow to define exactly which monomer is available a certain position. But it allows giving information about the probability for the potential monomers.

Monomer mixture

A mixture of monomers of a list may be found at given position.

```
PEPTIDE1{A.A.A.A.(A+G+W+[Aha]).A.(A+G).A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.
D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.
K.K.K.K.K.K}$$$$
```

Monomer mixture with ratios

A mixture of monomers of a list may be found at given position, ratio of monomers within that list is known.

```
PEPTIDE1{A.A.A.A.(A:1.1+G:69.5+W:25.5+[Aha]:3.9).A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.
D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2
{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.
K.K.K.K.K.K.K.K.K.K.K.K.K.K.K.K}$$$$
```

Mixture of 4 proteins with differing monomers at positions 5 and 10.

```
10% PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.
40% PEPTIDE1{A.A.A.A.G.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.
10% PEPTIDE1{A.A.A.A.A.A.A.A.A.G.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.
40% PEPTIDE1{A.A.A.A.G.A.A.A.A.G.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.
```

Resulting HELM string with ambiguity notation:

```
PEPTIDE1{A.A.A.A.(A+G).A.A.A.A.(A+G).A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.
```

Unknown monomer

Unknown monomers (0..n monomers) may be found at given position.

```
PEPTIDE1{A.A.A.A.*.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}$$$$
```

Unknown monomers (1..n monomers) may be found at given position.

```
PEPTIDE1{A.A.A.A.(X,*).A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}
$$$
```

Unknown monomers (1) may be found at given position in PEPTIDE.

```
PEPTIDE1{A.A.A.A.X.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}$$$
```

Unknown monomers (1) may be found at given position in RNA.

```
RNA1{R(A)P.R(G)P.R(N)P.R(A)P}$$$
```

Repeating monomers with defined count

If a simple polymer contains repeating monomers at a defined position, it can be described using number in single quotes. This is a short hand notation, as it does not introduce structure ambiguity.

```
PEPTIDE1{A.G.D.A'55'}
```

There are 55 A monomer at the end of the Peptide polymer.

Repeating monomers with range count

If a simple polymer contains repeating monomers at a defined position but without a defined count, this can be described using single quotes.

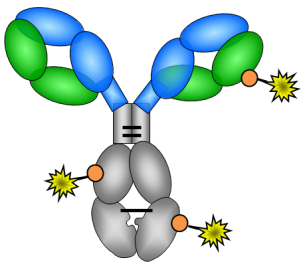
```
PEPTIDE{A.G.[repeatingMonomer]'70-100'}
```

repeatingMonomer is contained in a range of 70-100 single monomers.

Connection ambiguity

For Antibody-Drug-Conjugates, there are several use cases where connection ambiguities are important. These are described below.

Connected monomer type is known, position unknown



A small molecule binds to one Cysteine at peptide 1. This is denoted by a monomer name instead of the position in the connection section.

```
PEPTIDE1{A.A.A.A.A.A.C.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHE
M1{[*]SCCCc1ccccc1 |$_R1;;;;;;;;;;;$|}$PEPTIDE1,CHEM1,C:R3-1:R1$$
```

A small molecule binds to a either a Cysteine or a Lysine at peptide 2

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHE
M1{[*]SCCCc1ccccc1 |$_R1;;;;;;;;;;;$|}$PEPTIDE2,CHEM1,(C,K):R3-1:R1$$$$
```

Connection partner/monomer is undefined

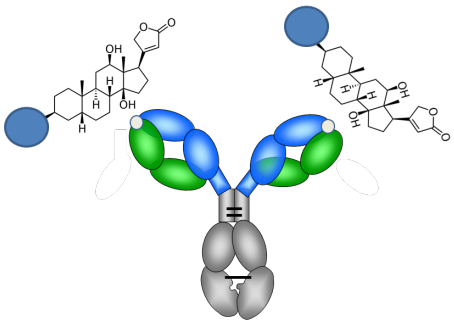


Figure 1: Anti-digoxigenin antibody with complex bound digoxigenin derivatives

We know PEPTIDE1 and CHEM1 are connected, and that's all we know.

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHE
M1{Dig}$PEPTIDE1,CHEM1,*:*-*:*$$$
```

Connection involves a simple polymer OR group

A small molecule binds to one Cysteine at peptide 1 or 2. This is denoted by a monomer name instead of the position in the connection section, and used polymer OR group.

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHE
M1{[*]SCCCc1ccccc1 |$_R1;;;;;;;;;;;$|}$G1,CHEM1,C:R3-1:R1$G1(PEPTIDE1,PEPTIDE2)$$
```

Connection involves a simple polymer AND group

A small molecule binds to Cysteines at peptide 1 and 2. This is denoted by a monomer name instead of the position in the connection section, and used polymer AND group.

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.
D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G
.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHE
M1{[*]SCCCc1ccccc1 |$_R1;;;;;;;;;;;$|}$G1,CHEM1,C:R3-1:R1$G1(PEPTIDE1+PEPTIDE2)$$
```

Binding ratio – Composition Ambiguity

The ratio between the conjugated drug and a protein (antibody) is usually determined by analytical methods like MS. This ratio is relevant to describe the ADC properly.

There are 3 types of ratios to be considered:

1. no ratio defined
2. ratio given as decimal number (e.g. 2.1)
3. ratio given as interval (e.g. 2.1 - 2.3)

If needed, a standard deviation can be added in the annotation section:

Binding ratios will be described not in the connection section, but in the simple polymer group section as AND groups.

If the same simple polymer is connected in different ratios to different chains, it has to be added multiple times to the simple polymer list. The connection ratio is then added to the simple polymer group section as described above.

```
Example: 2.5 equivalents of CHEM1 connected PEPTIDE1, and 1.5 equivalent CHEM2 connected
to PEPTIDE2
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.K.K.K.K.K.K}|CHEM1{Dig}|CHEM2{Dig}$PEPTIDE1,CHEM1,C:R3-1:R1|PEPTIDE2,CHEM2,C:R3-1:R1$G1(PETPDIE1+CHEM1:2.5)|G2(PEPTIDE2+CHEM2:1.5)$$
```

Example: 2.5 equivalents of CHEM1 bound to 1 equivalents of PEPTIDE1

```
PEPTIDE1{Flu}|CHEM1{*}$PEPTIDE1,CHEM1,C:R3-1:R1"annotation"$ G1(PETPDIE1+CHEM1:2.5)|$$
```

Example: 1 equivalents of CHEM1 bound to 2.5 equivalents of PEPTIDE1

```
PEPTIDE1{Flu}|CHEM1{*}$PEPTIDE1,CHEM1,C:R3-1:R1"annotation"$G1(PETPDIE1:2.5+CHEM1)|$$
```

Example: 1 equivalents of CHEM1 bound to a range of 1.5 to 2.5 equivalents of PEPTIDE1

```
PEPTIDE1{Flu}|CHEM1{*}$PEPTIDE1:1.5-2.5,CHEM1,C:R3-1:R1"annotation"$G1(PETPDIE1:1.5-2.5+CHEM1)|$$
```

Example: ADC with a given antibody drug ratio of 4.5 (statistical binding):

```
PEPTIDE1{*}"LC"|PEPTIDE2{*}"HC"|PEPTIDE3{*}"HC"|PEPTIDE4{*}"LC"|CHEM1{*}$
G1,CHEM1:4.5,K:R3-1:R1|
PEPTIDE2,PEPTIDE3,250:R3-250:R3"Hinge S-S connection"|
PEPTIDE2,PEPTIDE3,252:R3-252:R3"Hinge S-S connection"|
PEPTIDE1,PEPTIDE2,120:R3-248:R3"LC Hinge S-S connection"|
PEPTIDE4,PEPTIDE3,120:R3-248:R3"LC Hinge S-S connection"
$G1{PEPTIDE1+PEPTIDE2+PEPTIDE3+PEPTIDE4)|G2(G1+CHEM1:4.5)$$
```

Component ambiguity

Sequence unknown, type of polymer not defined

For undefined simple polymer type, i.e. not RNA, PEPTIDE or CHEM, use BLOB. Put polymer type inside curly braces, and put name/description via inline notation

```
BLOB1{Antibody}"HER2"$$$$ represents a HER2 antibody
BLOB1{Gold Particle}"Gold 2.5nm"$$$$ represents a gold particle with 2.5nm diameter.
```

Sequence unknown, type of polymer known

use * to indicates that unknown polymer structure.

```
PEPTIDE1{*}$$$$
```

Sequence partially known

Sequence partially known, exact residue count known (e.g. H3 hidden).

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.X.X.X.X.X.X.K.K.K}$$$$PEPTIDE1{hc}|PEPTIDE2{lc}$
```

Sequence partially known, exact residue count known (e.g. H3 hidden).

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.X'6'.K.K.K}$$$PEPTIDE1{hc}|PEPTIDE2{lc}$
```

Sequence partially known, exact residue count almost unknown (e.g. hidden domain).

```
PEPTIDE1{A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.A.C.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.D.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E.E}|PEPTIDE2{G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.C.S.S.S.S.S.S.S.S.P.P.P.P.P.P.P.P.K.K.K.K.K.(X,*)"Note:_at_least_one_residue".K.K.K}$$$$
```

Bead coupling

Peptide sequence conjugated to gold particle

```
PEPTIDE1{C.C.C.C.C.C}|BLOB1{Gold Particle}"Au10,
Diameter:10nm"$PEPTIDE,BLOB1,C:R3-?:R1$G1(PEPTIDE1:20-34+BLOB1)${"Name":"Gold particle conjugated with peptides","Load":26}$
```

One out of two sequences conjugated in given ratio with not known structure (bead).

```
PEPTIDE1{C.C.C.C.C.C}|PEPTIDE2{A.C.A}|BLOB1{Gold Particle}"Au10,
Diameter:10nm"$G1,BLOB1,C:R3-?:R1$G1(PEPTIDE1,PEPTIDE2)|G2(G1:24+BLOB1)$V2.0
```

```
PEPTIDE1{C.C.C.C.C.C}|PEPTIDE2{A.C.A.A.A.A}|BLOB1{}$(PEPTIDE1,PEPTIDE2):24,BLOB1,2:R3-?:R1$$PEPTIDE1{Type:Peptide,Name:Gold-conjugated peptide}|BLOB1{Type:Gold particle,Name:Au10,Diameter:10nm}$
```

Total of two sequences in given distribution conjugated in given ratio with not known structure (bead).

```
PEPTIDE1{C.C.C.C.C.C}|PEPTIDE2{A.C.A}|BLOB1{Gold Particle}"Au10,
Diameter:10nm"$G1,BLOB1,C:R3-?:R1$G1(PEPTIDE1:49+PEPTIDE2:51)|G2(G1:24+BLOB1)$V2.0
```

Nanoparticles

A nanoparticle contains RNA1 and RNA2 as payload, PEPTIDE1 as surface ligand, and uses Lipid A for nanoparticle formation. Component ratios are specified in section 3: Simple Polymer Groups

RNA1{R(A)P.R(G)P}| RNA2{R(A)P.R(G)P}||PEPTIDE1{A.G.C.H.E}|CHEM1{*}"Lipid A"$$G1(RNA1+RNA2:2)|G2(G1+PEPTIDE1:5.0)${"Name":"lipid nanoparticle with RNA payload and peptide ligand"}$V2.0

Component and connection and composition ambiguity

In the example below, we provide an example where ambiguity is everywhere. First, we have a component ambiguity where CHEM1 is a payload without structure, and BLOB1 is a HER2 antibody without sequence info. Second, we have connection ambiguity where CHEM1 and BLOB1 is connected, but the connecting monomer and/or R group is unknown. Third, we have composition ambiguity that the ratio between CHEM1 and BLOB1 is 4.5 as shown in G1. Finally, this ADC is connected with PEPTIDE1. The connection information from the ADC (G1) is unknown (*:*), but connection info from PEPTIDE is partially known (K:R3)

```
PEPTIDE1{A.G.C}|CHEM1{*}"Payload"|BLOB1{Antibody}"Her2"$BLOB1,CHEM1,*:*
-1:*|G1,PEPTIDE1,*:*-K:R3$G1{BLOB1+CHEM1:4.5}$$V2.0
```

Glycosylation

No ambiguity

Glycosylation fully defined. This is HELM 1.x.

```
CHEM1{BMA}|CHEM2{MAN}|CHEM3{MAN}|CHEM4{NAG}|CHEM5{FUL}|CHEM6{NAG}|CHEM7{NAG}|CHEM8{NAG}|C
HEM9{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.[N
gly].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}|CHE
M39{BMA}|CHEM40{BMA}|CHEM41{MAN}|CHEM42{NAG}|CHEM43{NAG}|CHEM44{BMA}|CHEM45{NAG}|CHEM46{N
AG}|CHEM47{FUL}|CHEM48{NDG}|CHEM49{NAG}|CHEM50{MAN}|CHEM51{BMA}|CHEM52{BMA}|CHEM53{NAG}|C
HEM54{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$$$
```

In annotation a group containing the sugar moieties should be defined. That group can have further attributes, such as name, composition, symbol etc. This will fit into the concept of extended annotations.

Component ambiguity - Glycosylation not fully defined.

The glycosylation group could be replaced by a CHEM1{*}"Glycosylation" component, if glycosylation is unknown. But be aware, that the group itself is just an annotation facilitating the addition of metadata to a predefined part of the macromolecule and facilitates replacement of a particular molecule part; other example could be group referencing the peptide chains of an antibody.