Antibody Markup Language (AbML) Format Description V1.0

10th February 2022

General

- Whitespace (including line breaks) is ignored except within comments
- The system is case insensitive except for the comments
- The term 'domain', as used in this document, is a general term for a region of the protein and can refer to flexible linkers and hinge regions as well as formal protein domains.

Domain Types

VL Variable Light
CL Constant Light
VH Variable Heavy
VHH Camelid single VH domain
CH1 Constant Heavy 1
H Hinge
CH2 Constant Heavy 2
CH3 Constant Heavy 3
CH4 Constant Heavy 4
L Linker
X Extra domain
C Chemical conjugation

Domain Peptide Connectivity

Working from N-terminus to C-terminus, connectivity is indicated with a '-'. Chains are separated by a '|'

e.g.

```
VL-CL|VH-CH1-H-CH2-CH3
```

Domain Identifiers and Interactions

After any Domain Type, a numeric domain identifier may be indicated in parentheses. These will normally be used sequentially.

e.g. for a normal antibody:

```
VL(1)-CL(2) | VH(3)-CH1(4)-H(5)-CH2(6)-CH3(7) |
VL(8)-CL(9) | VH(10)-CH1(11)-H(12)-CH2(13)-CH3(14)
```

Interacting domains are indicated by a ':' followed by a comma-separated list of interacting domain IDs.

e.g. for a normal antibody

```
VL(1:3)-CL(2:4) | VH(3:1)-CH1(4:2)-H(5:12)-CH2(6:13)-CH3(7:14) |
VL(8:10)-CL(9:11) | VH(10:8)-CH1(11:9)-H(12:5)-CH2(13:6)-CH3(14:7)
```

Disulfides

The number of disulfides occurring between interacting domains can be indicated in curly brackets. Note that disulfides must follow an domain interaction indicator.

e.g. for a normal antibody

```
VL(1:3)-CL(2:4){1} | VH(3:1)-CH1(4:2){1}-H(5:12){2}-CH2(6:13)-CH3(7:14) |
VL(8:10)-CL(9:11){1} | VH(10:8)-CH1(11:9){1}-H(12:5){2}-CH2(13:6)-CH3(14:7)
```

Specificity

For multi-specific antibodies, the specificity is indicated with a '.x' after the Domain Type. e.g. `VL.a`, `VL.b`. A domain having multiple specificities is indicated with '.x...' e.g. `VL.ab` for two specificities, etc.

Linkers

Linkers (indicated by an 'L') simply occur within the sequence of domains. A Linker may be followed by (domain ID / interaction) information optionally followed by disulphide information and/or a comment.

The comment keyword LENGTH: is reserved for indicating the length of a linker.

e.g.

- L(5), L(5:10), L(5:10){1}
- L[LENGTH:20]
- L(5:10){1}[LENGTH:15]

Extra Domains (X)

An Extra Domain is indicated with the Domain Type X. An Extra Domain may be followed by (domain ID : interaction) information, optionally followed by disulphide information and/or a comment. Typically a [TYPE:xxx] comment will be included to indicate the type of the extra domains. (See **Comments**, below)

Chemical Moieties (C)

Chemical Moieties are chemical cross linkers indicated with the Domain Type C and used to join two or more protein domains (note that these are *not* conjugation linkers for ADCs).

Chemical Moieties may be followed by (domain ID : interaction) information, optionally followed by a comment. Typically a [TYPE:xxx] comment will be included to indicate the type of the chemical moiety. (See **Comments**, below)

Modifications

Specific and general domain modifications can be indicated with the following symbols which must appear immediately after a Domain Type:

- ^ specific ADC site
- > a 'knob' for domain pairing
- @ a 'hole' for domain pairing
- + a positive charge for domain pairing
- _ a negative charge for domain pairing (note this is an underscore, since the - is reserved for connections between domains)
- ! used after a CH2 domain to indicate that it is not glycosylated
- * a general modification (which may then be explained by a comment)

e.g.

```
CH3>(7:14)
CH3@(14:7)
```

Where a modification occurs to a variable domain where the specificity is also indicated, the modification is described before the specificity.

e.g.

```
VL*.a
```

Thus a bispecific antibody using a knob-into-hole for heavy chain pairing and charges for light chain pairing might be:

```
VL.a(1:3)-CL+(2:4){1} |
VH.a(3:1)-CH1_(4:2){1}-H(5:12){2}-CH2(6:13)-CH3>(7:14) |
VL.b(8:10)-CL_(9:11){1} |
VH.b(10:8)-CH1+(11:9){1}-H(12:5){2}-CH2(13:6)-CH3@(14:7)
```

A modification in CH2 to enhance FcRn binding would be:

```
CH2*[MOD:ENHANCEFCRN]
```

Comments

Comments (each preceded by a keyword) may be added in square brackets and appear last in the set of qualifiers after a Domain Type. e.g. `VL.a(1:3)[ANTI:CD3]`

Multiple comments may appear as a comma-separated list, or in separate sets of square brackets. e.g. `VL*.a(1:3)[ANTI:CD3,MOD:PI]` or `VL*.a(1:3)[ANTI:CD3][MOD:PI]`

The following keywords are currently allowed for comments:

- ANTI: Gives the specificity (free text)
- MOD: Used to indicate the type of a modification - only a restricted list is allowed
- TYPE: Used with Extra Domains and Chemical Moieties to indicate what they are - only a restricted list is allowed
- LENGTH: The length of a domain (typically of a Linker)
- NOTE: Any other comment (free text, must appear last in a list of comments)

TYPE - allowed keywords

The following keywords are reserved for Extra Domain types:

- TYPE:ZIPPER - a leucine zipper
- TYPE:FUSION - a fusion protein
- TYPE:OTHER - a type of extra domain not explained by any reserved keywords (explained in a NOTE comment)

The following keywords are reserved for Chemical Moiety types:

- TYPE:OPDM - a thiol-thiol chemical crosslinker (orthophenylenedimaleimide)
- TYPE:SPDP - an amine-amine chemical crosslinker (succinimidyl 3-(2-pyridyldithio)propionate)
- TYPE:SMCC - a thiol-amine chemical crosslinker (succinimidyl 4-(N-maleidomethyl)cyclohexane-1-carboxylate)
- TYPE:OTHER - a type of extra domain not explained by any reserved keywords (explained in a NOTE comment)

MOD - allowed keywords

- MOD:ENHANCEFCRN - a modification to enhance FcRn binding
- MOD:ENHANCEADCC - a modification to increase antibody dependent cell-mediated cytotoxicity
- MOD:STRANDEXCHANGE - a modification for strand exchange engineered domains
- MOD:DISULPHIDE - a modification for additional disulphide bonds
- MOD:DISULFIDE - a modification for additional disulphide bonds
- MOD:PI - a modification to alter the isoelectric point
- MOD:CONJUGATION - a modification for a specific conjugation site
- MOD:HEXAMER - a hexamer formation of IgG
- MOD:NOFCGR - a modification to reduce FcRn binding
- MOD:NOPROTEINA - a modification to reduce ProteinA binding
- MOD:NOOX - a modification to reduce oxidation
- MOD:NOADCC - a modification to reduce antibody dependent cell-mediated cytotoxicity
- MOD:NOCDC - a modification to reduce complement dependent cytotoxicity
- MOD:NOADCP - a modification to reduce antibody dependent cellular phagocytosis
- MOD:NOADCCCDC - a modification to reduce ADCC and CDC
- MOD:NOGLYCOS - a modification to remove glycosylation site (other than the CH2 one which has its own symbol)
- MOD:NOADE - a modification to remove prevent antibody dependent enhancement of viral uptake
- MOD:NOAGG - a modification to reduce aggregation
- MOD:NOPROT - a modification to reduce proteolysis
- MOD:REMCYS - a modification to remove free cysteine or a disulphide
- MOD:STABILIZATION - a modification for stabilization
- MOD:AFFINITY - a modification to increase or decrease affinity
- MOD:OTHER - a modification not explained by any reserved keywords (explained in a NOTE comment)