A Notation Language for Bispecific Antibody Formats (Antibody Markup Language) and Software for Obtaining Expressions for Desired Antibodies (abYdraw)

James Sweet-Jones, Maham Ahmad, Andrew C.R. Martin
Institute of Structural and Molecular Biology,
Division of Biosciences, University College London,
Darwin Building, Gower Street,
London, WC1E 6BT, UK

February 2, 2022

Abstract

Bispecific antibodies (BsAbs) are an up-and-coming class of biologic drugs that differ from monoclonal antibodies through their ability to bind to more than one type of antigen. As techniques to generate such molecules have diversified, so have their formats and need for standard notation. Previous efforts for developing a notation language for macromolecule drugs have been insufficient for describing BsAbs. Here, we present Antibody Markup Language (AbML), a new notation language specifically for antibody formats which overcomes the limitations of existing languages and can notate a great diversity of current BsAb formats. To assist users in this language we also developed a tool, abYdraw, that may draw antibody schematics from AbML descriptor strings or generate a descriptor string from a drawn antibody schematic. AbML has potential to become a standardised notation for describing new BsAb formats entering clinical trials.

1 Introduction

Immunoglobulins (IgG), or antibodies, have become useful molecular tools in biology and medicine due to their natural ability to bind a specific antigen. When clonally expanded, monoclonal antibodies (mAbs) have applications spanning molecular diagnostic assays, to medical imaging and of course, as drug therapeutics [1]. Bispecific Antibodies (BsAbs) are specially engineered proteins that differ from naturally occurring mAbs in their ability to bind to more than one type of antigen. Depending on the number of different antigens the molecule in question reacts with, antibodies can be considered trispecific or tetraspecific, but these will all be called BsAbs in this report. This makes them a versatile class of molecules which has become a keen focus of therapeutics in clinical trials because their bispecificity allows proximally gathering two molecules, two cells, or one of either to initiate a reaction or neutralise an infectious agent [2]. Logically this has given them great interest in the field of immunomodulatory cancer treatments [3], which there are currently two FDA-approved BsAb drugs: blinatumomab and catumaxomab [4, 5] with another, emicizumab, approved for Factor VIII deficiency haemophilia [6].

The engineering of these molecules has evolved over time since their inception in the 1970s. At first, the quadroma was to fuse two hybridoma cell lines, used for generating mAbs which would then result in some cases where two halves of different Fab fragments form heterodimers which results in a molecule with two specificities [7, 8]. These techniques offer poor yield due to the disfavoured formation of the desired heterodimers and so efforts for more scalable synthesis have led to new techniques of BsAb generation [9].

DNA recombination has allowed greater flexibility in designing BsAbs with IgG-like formats, which can be done by appending additional Fv fragments at the N- or C- terminus of IgG light or heavy chains [10]. This may generate a BsAb fragments can dimerise to give a symmetrical molecule which is favoured in the reaction. Additional residue mutations for knobs-in-holes (KIH) formats [11] as well as positively and negatively charged formats [12] have assisted in chain pairing to make the desired antibody format more favourable when pairing unsymmetrical formats [9].

Furthermore, recombination also allows linking of VH and VL domains to give single chained Fv (scFv) fragments or camelid single domain Fv fragments (nanobodies), which may be sequentially added by engineered linkers to allow for more specificities along one protein chain [13]. Protein engineering allows for generation of smaller fragment based BsAbs including 2-chained diabodies or a single chained sequence of scFvs. These Non-IgG-like molecules are advantageous because they are easier to produce, and have lower immunogenicity risk, but these molecules are limited by short half-lives, which can be extended through human serum albumin (HSA) ligation or polyethelene glycol(PEG)-ylation or the addition of disulphide bonds [14, 1].

Most recently chemical conjugation allows modularity of domains which has given rise to great diversity in structures and presentation of these molecules [9]. In addition to ligating antibody fragments as seen in the "Dock and Lock" format, antibody-drug-conjugates (ADCs) and fusion proteins have become popular as methods of delivering small molecule drugs to an intended target [15]. Finally, the potential of chemical ligation has demonstrated production of BsAbs by appending two IgG molecules to give IgG-IgG molecules [16].

Improvements in synthesis now leading to be more successful drugs and because of the increased diversity of BsAbs available, they require standardised formats for description and annotation when they are submitted for clinical trials. HELM notation is a useful tool for noting biologics which tend to be complex, large macromolecules, including antibodies and an editor was also developed to assist in generating expressions that display the molecule described [17]. Whilst this is suitable for standard IgG molecules, it has shortcomings when it is used for BsAbs. The greatest point which makes HELM unsuitable for BsAbs is that it does not allow for notation of Fv fragment specificities or to add comments or notes about the type of extra domains that can be added to an IgG molecule. Furthermore, the HELM editor does not have specific expressions for antibody-based drugs as it requires amino acid sequences to draw a schematic of a molecule, which is not suitable when describing fragmented BsAbs. Due to the gaining popularity of BsAbs and the unsuitability of current macromolecule notation, it presents a case to develop a new notation language that adequately describes the ever-increasing diversity of BsAbs.

In this paper, we present a new BsAb notation language, Antibody Markup Language (AbML), which we demonstrate can be used to describe a variety of current BsAbs formats. Furthermore we apply our language in a graphical programme, abYdraw, which uses AbML to render schematics of BsAbs, as well as produces expressions from drawn antibody schematics.

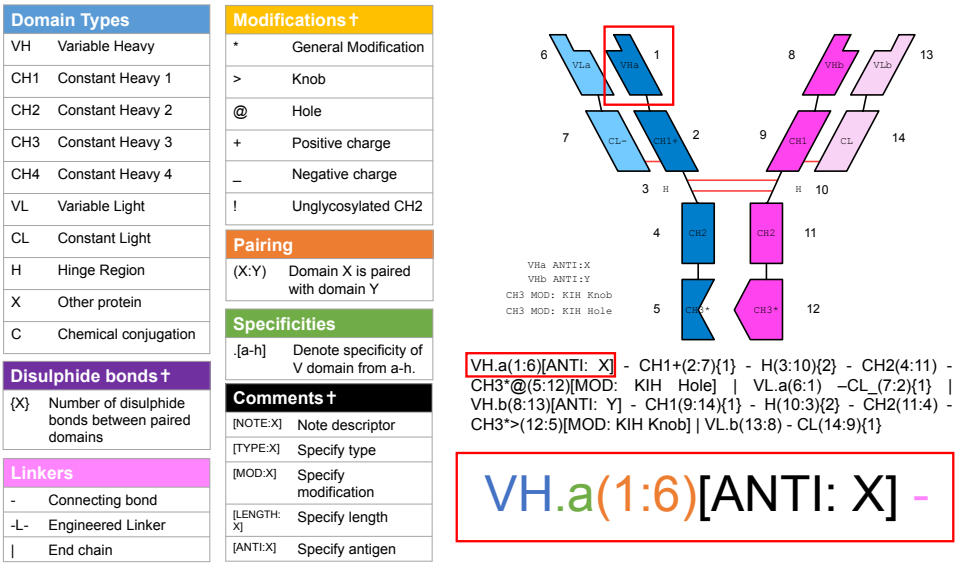

Figure 1: AbML Guidesheet Explaining the Properties of the Language. All possible domain types, modifications, linkers and comment types as well as how to notate pairings and dislphide bonds are given in colour-coded fashion to the example antibody domain highlighted in red. Antibody schematic was rendered with abYdraw and numbers represent the numbering of each domain given in AbML and labelled on the schematic.

2 Methods

2.1 Devising a descriptor language

We used literature describing the formats of over 60 BsAbs to develop our description language [9]. An effective notation language must be simply structured but can carry as much information as needed. Using the documentation of the HELM descriptor language [17], it was decided that our language would be structured similarly, but correct for instances where HELM is not appropriate for BsAbs. We adopt similar domain type characters in AbML but remove the need to specify a constant heavy (CH) domain and comment it as to which CH type it is. Similarly we decided to use `"-"` characters to indicate connections between domains but we also include hinges and linkers as a type of connector-based domain rather than having to specify linkers as an extra type of domain. Furthermore, disulphide bonds between interacting domains were to be shown as these are important in stabilising chain interactions. Finally it was decided to incorporate a comment system would to describe modifications made to the BsAb and denote `"X"` domains which are not in the standard domain types.

2.2 Developing abYdraw

When it had been shown the descriptor language could be successfully applied to a variety of BsAb types, our next step was to make this language accessible. The most suitable way for that was to design a programme which would give researchers the expression for their antibody of interest and allow them to save both the expression and a rendering of the schematic. abYdraw was written in Python3 and given a graphical user interface with TKinter, a standard Python package.

3 Results

3.1 Antibody Markup Language

AbML is based on describing protein domains, arranged in a string and separated by connecting bonds, which represents a chain from a BsAb from N-terminus to C-terminus. Each domain is numbered sequentially in order of its appearance in the expression unless referring to a previously notated domain. Domains are connected by `"-"` characters posing as linkers between domains. Chains are separated by `"|"` characters and those which are part of the IgG molecule can be presented in any order but additional fragments that interact with it typically are noted last. Each chain must have at least one domain that interacts with another domain on a different chain. Chains that do not interact with the others are considered different antibodies. Whitespace, including line breaks are ignored in AbML except for comments given in square braces. What follows is an in-depth description of how units link together to generate a complete expression.

Domains of the antibody are given in a format that aims to convey as much information about the domain without becoming overly complex. Notating domains always begins with the type of domain in question. Possible domains include [`"VH"`, `"VL"`, `"CH1"`, `"CH2"`, `"CH3"`, `"CH4"`, `"CL"`, `"X"`, `"C"`, `"H"`] which are explained in the language guide sheet (Figure 1). Domains containing `"X"` are considered extra domains that are not part of a standard immunoglobulin which will usually be explained by comments.

Specific modifications notated by special characters can follow the domain type which include `"@"` and `">"` for KIH pairing with holes and knobs given respectively or `"+"` and `"_"` denoting a positively or negatively charged domain. General modifications are given with a `"*"` which can be elaborated by a comment. Modification symbols can appear in any order when they immediately succeed the domain type in the expression however AbML does not accommodate combinations where `"@"` appears with `">"` or `"+"` appears with `"_"` because these are contradictory adaptations. Similarly, `"!"` can only appear in domains of type `"CH2"` as it specifies this domain is unglycosylated.

If applicable, Fv specificities can be notated by `"."` followed by a letter corresponding to one type of Fv specificity. We provide letters a-h, with a single letter representing a different antigen specificity or multiple letters can indicate a combination of specificities. Typically, an interacting pair of VH and VL domains should both be assigned identical specificity descriptors but exceptions apply when two different light chains share a common heavy chain.

Following specificities, a bracketed list of the number on that domain a colon and its interacting domain, although it is optional to exclude interacting domains e such as the case of nanobodies or to include a list of interactions when describing an `"X"` domain multimer. Possible domain pairings include `VH:VL`, `CH1:CL`, `CH2:CH2`, `CH3:CH3`, `CH4:CH4`, `H:H`, `X:X`, and `C:C`. Further optional curled braces notating the number of disulphide bonds between the previously specified interacting domains.

Each domain may be given an optional comma-separated list of comments within a set of squared braces. These comments can denote types of other protein domains not notated in the language as well as antigen specificities or the length of a domain or linker. A full list of keywords and modifications can be found on the AbML descriptor sheet (Figure 1).

Domains are linked by `"-"` which represents an ordinary peptide linker between bonds, however these can be changed to hinge regions or artificially engineered linkers (`-L-`) which may carry more information about pairing. disulphide bond connections noted by sets of round or curled braces. It is conventional that hinge regions are placed between CH1 and CH2 domain types and linkers are placed where

a different antibody chain or domain is appended to the IgG molecule, and between the VH and VL domains of a scFv.

3.2 abYdraw

abYdraw is a graphical programme written in Python3 where users may input expressions in AbML to obtain a schematic of their designed antibody by clicking the `"Get Structure"` button. However, the user is also able to draw antibodies by arranging standard antibody domains and connecting them with linkers to obtain the appropriate expression for their design by using the `"Get Sequence"` button. Once the sequence is obtained for the drawing, using `"Get Structure"` will re-render the schematic automatically. Both functions can run in sequence using the `"Tidy"` button. The programme will also print out comments made in the expression and highlight the domain linked to those comments. abYdraw can be used to export these schematics as figures for future publications and generate a standardised expression that may be used in BsAb annotations.

The interface draws domains as blocks labelled with their domain type and modifications given in the expression. Character substitutions include `"_"` to `"-"` in labels with `"@"` and `">"` characters also being omitted as these modifications are used to affect the shape of the domain. Domains are coloured according to their specificities descriptor. It is possible that chains will have blocks of different colours when domains of different specificities are given in the same chain. Variable domains appear with a cut-out at the top of the domain referring to its antigen-interacting site which pairs with another to give a complete Fv fragment. Nanobodies have a unique domain shape reflecting their single domain Fv fragment as KIH adaptations are displayed by constant domains with either a cut-out or an extension to their side which slots together to demonstrate how these domains are paired.

Normal Connections between each domain are given by black lines that are drawn from the bottom of one domain to the top of the next domain. Artificial linkers are shown in purple lines, disulphide bonds are shown by red lines and hinges are shown in dark green. A default list of colours for all domain and bond types is given in the programme but these may be changed in settings for the user. Any comments given in the expression are printed with the domain symbol of the domain which it applies to the antibody. If multiple comments are given for a domain, they are printed on the same line.

Users may draw BsAbs from scratch or begin with a template design of common BsAb formats that may be manipulated by the user. To Draw domains, a user must select a specificity and any mods for that domain and then place it on the canvas. Both specificities and modifications can be updated whilst on the canvas by selecting a specificity or modification, but not a domain type. Once drawn, domains may be translated to a space where they interact with other domains to be paired. VH and VL domains must face each other to be considered as interacting. Users can right-click newly drawn domains to change the direction they are facing. Nanobodies cannot be paired with other domains as these are single-domain Fv fragments. Bonds connecting domains must be drawn starting on the previous domain and ending on the current domain. Disulphide bonds can be drawn starting from either domain that are supposed to be interacting. To Insert a comment, users will select which comment type they want, type their comment into the keypad entrybox and then place the comment inside the domain it is applicable to.

Figures 2 and 3 demonstrate that AbML may be applied to numerous antibody formats described by Spiess *et al.*[9] and then rendered using abYdraw.

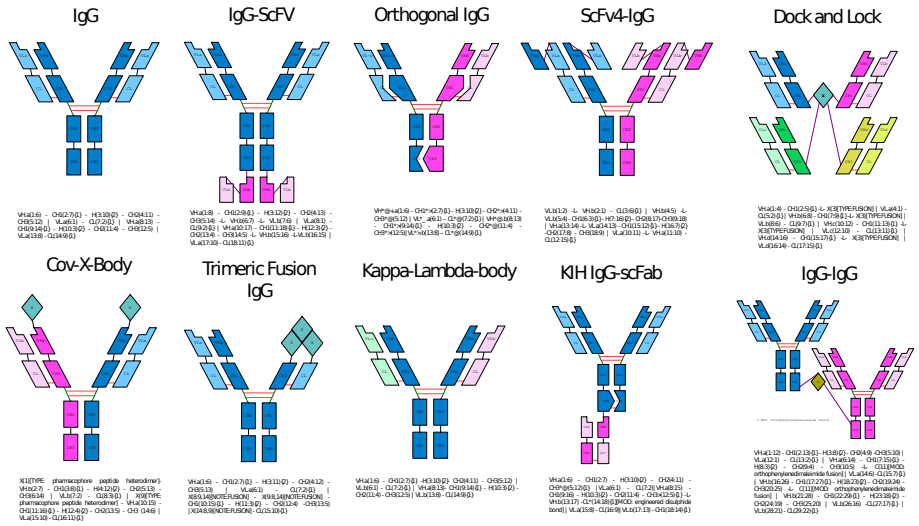

Figure 2: AbML descriptor strings of commonly-used 4-chain bispecific antibodies. Schematics of antibodies were rendered in abYdraw.

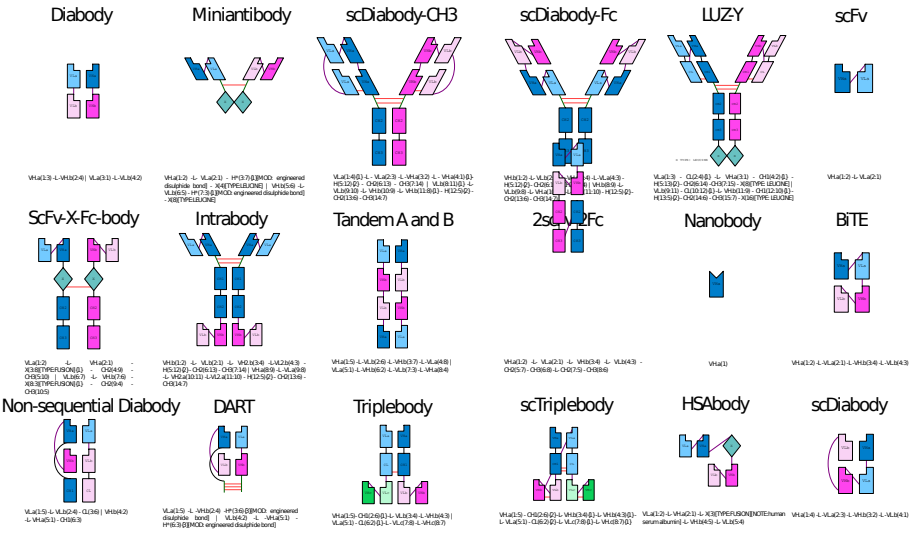

Figure 3: AbML descriptor strings of commonly-used 2-chain and 1-chain bispecific antibodies. Schematics of antibodies were rendered in abYdraw.

4 Discussion

By addressing the pitfalls of currently available notation languages, we have developed AbML which can be applied to existing BsAbs. Our language was based on the established HELM notation for macromolecule biologics but simplified and adapted to specifically describe antibody formats. The simplicity of AbML over HELM allows greater accessibility as well as allowing the potential to insert additional modifications symbols and domain types that will futureproof the language to cope with the inevitably expanding formats of recombinant and chemically conjugated BsAbs. This will not always be necessary as `"X"` and `"C"` domains can be described as a multitude of possible fusion proteins, drug conjugates and chemical bonds using the comments system, and so the language should not require constant updating. Despite this anticipation, it is expected that AbML and abYdraw will require updating with new formats that emerge.

We hope that by developing abYdraw, a compiled application which may generate AbML from an antibody schematic, where the schematic and expression may be used to describe the BsAb format. Tools such as abYdraw are intended to make AbML more accessible and promote the use of AbML to its potential in becoming a widely-used, standardised notation for describing BsAb formats. To aid this usage, abYdraw includes a library of commonly used BsAb formats complete with AbML strings and diagrams that can be used as starting points for researchers to design new drugs. While the AbYdraw software is available from GitHub repositories, further developments could be made to promote it on online by porting it to JavaScript or by installing a command-line interface that could be installed onto a webpage to make it more accessible.

5 Conclusion

To Conclude, our annotation language AbML is a new descriptor language for BsAb formats and has demonstrated its ability application to existing BsAb formats. We envision this language and its corresponding tool abYdraw to become useful in the development of future BsAb drugs, allowing for standardisation of BsAb description as part of ushering in a new era of BsAb development. Improved descriptions of their formats will demonstrate the most popular formats and those which are most likely to work as drugs, therefore prompting greater development in the bispecific field.

6 Software Availability

Compiled apps for Mac OS and windows are made free to download at: Source code for this project is also made available at `https://github.com/JamesSweetJones/abYdraw`

References

[1] J. Ma, Y. Mo, M. Tang, J. Shen, Y. Qi, W. Zhao, Y. Huang, Y. Xu, and C. Qian. Bispecific antibodies: From research to clinical application. *Frontiers in Immunology*, 12:626616, 2021.

[2] G. Fan, Z. Wang, M. Hao, and J. Li. Bispecific antibodies and their applications. *Journal of Hematology and Oncology*, 8:130, 2015.

[3] A. F. Labrijn, M. L. Janmaat, J. M. Reichert, and P. W. H. I. Parren. Bispecific antibodies: a mechanistic review of the pipeline. *Nature Reviews Drug Discovery*, 18:585–608, 2019.

[4] A. C. Wilke and N. Gökbuget. Clinical applications and safety evaluation of the new CD19 specific T-cell engager antibody construct blinatumomab. *Expert Opinion on Drug Safety*, 16:1191–1202, 2017.

[5] D. Seimetz. Novel monoclonal antibodies for cancer treatment: the trifunctional antibody catumaxomab (removab). *Journal of Cancer*, 2:309–316, 2011.

[6] C. Schmitt, J. I. Adamkewicz, J. Xu, C. Petry, O. Catalani, G. Young, C. Negrier, M. U. Callaghan, and G. G. Levy. Pharmacokinetics and pharmacodynamics of emicizumab in persons with hemophilia A with factor VIII inhibitors: HAVEN 1 study. *Thrombosis and Haemostasis*, 121:351–360, 2021.

[7] C. Milstein and A. C. Cuello. Hybrid hybridomas and their use in immunohistochemistry. *Nature*, 305:537–540, 1983.

[8] R. E. Kontermann and U. Brinkmann. Bispecific antibodies. *Drug Discovery Today*, 20:838–847, 2015.

[9] C. Spiess, Q. Zhai, and P. J. Carter. Alternative molecular formats and therapeutic applications for bispecific antibodies. *Molecular Immunology*, 67:95–106, 2015.

[10] U. Brinkmann and R. E. Kontermann. The making of bispecific antibodies. *MAbs*, 9:182–212, 2017.

[11] J. B. B. Ridgway, L. G. Presta, and P. Carter. 'knobs-into-holes' engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein Engineering, Design and Selection*, 9:617–621, 1996.

[12] K. Gunasekaran, M. Pentony, M. Shen, L. Garrett, C. Forte, A. Woodward, S. Bin Ng, T. Born, M. Retter, K. Manchulenko, H. Sweet, I. N. Foltz, M. Wittekind, and W. Yan. Enhancing antibody Fc heterodimer formation through electrostatic steering effects: Applications to bispecific molecules and monovalent IgG. *Journal of Biological Chemistry*, 285:19637–19646, 2010.

[13] F. Le Gall, S. M. Kipriyanov, G. Moldenhauer, and M. Little. Di-, tri- and tetrameric single chain Fv antibody fragments against human CD19: effect of valency on cell binding. *FEBS Letters*, 453:164–168, 1999.

[14] R. E. Kontermann. Strategies for extended serum half-life of protein therapeutics. *Current Opinion in Biotechnology*, 22:868–876, 2011.

[15] S. Sau, H. O. Alsaab, S. K. Kashaw, K. Tatiparti, and A. K. Iyer. Advances in antibody-drug conjugates: A new era of targeted cancer therapy. *Drug Discovery Today*, 22:1547–1556, 2017.

[16] P. Szijj and V. Chudasama. The renaissance of chemically generated bispecific antibodies. *Nature Reviews Chemistry*, 5:78–92, 2021.

[17] T. Zhang, H. Li, H. Xi, R. V. Stanton, and S. H. Rotstein. HELM: A hierarchical notation language for complex biomolecule structure representation. *Journal of Chemical Information and Modeling*, 52:2796–2806, 2012.