Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

VERITAS: Harnessing the power of nomenclature in biologic discovery

Riti Biswas [a], Ed Belouski[b], Kevin Graham [b], Michelle Hortter[b], Marissa Mock[b], Christine E. Tinberg [a], and Alan J. Russell [b]

aBiologic Therapeutic Discovery, Amgen Research, South San Francisco, CA, USA; bBiologic Therapeutic Discovery, Amgen Research, Thousand Oaks, CA, USA

ABSTRACT

We are entering an era in which therapeutic proteins are assembled using building block-like strategies, with no standardized schema to discuss these formats. Existing nomenclatures, like AbML, sacrifice human readability for precision. Therefore, considering even a dozen such formats, in combination with hundreds of possible targets, can create confusion and increase the complexity of drug discovery. To address this challenge, we introduce Verified Taxonomy for Antibodies (VERITAS). This classification and nomenclature scheme is extensible to multispecific therapeutic formats and beyond. VERITAS names are easy to understand while drawing direct connections to the structure of a given format, with or without specific target information, making these names useful to adopt in scientific discourse and as inputs to machine learning algorithms for drug development.

The biopharmaceutical industry is increasingly realizing the promise of antibody formats which bind two or more targets, i. e., multispecific antibodies (MsAbs). At the end of 2021, 17% of antibody-based therapeutics in late-stage clinical trials used these non-traditional mAb formats (as compared to only 6% in 2013).[1,2] MsAbs can engage in various therapeutic modes, such as enhancing cancer cell-specific T cell-mediated cell killing, that small molecules and monospecific antibodies cannot leverage as easily or effectively.[3]

Valency and the orientation of the targeting domains of MsAbs affect the characteristics and biological properties of these molecules. For example, CD3 binding valency in tumor-targeted T cell-activating antibody formats is generally monovalent, which avoids clustering of the receptor and subsequent off-target T cell activation. However, increasing tumor-associated antigen binding valency in these molecules can improve affinity and cell killing (as in the case of the trivalent anti-EGFR ATTACK format).[4] Thus, as the number of therapeutically relevant biological mechanisms expands, so does the variety of antibody-based formats used to explore them. Many of these formats have widely accepted names, such as CODV, diabody, TriBi minibody, DART, CrossMab, DVD-Ig, and BiTE®, which are generally created by their inventors.[3,5–10] However, it is unreasonable to memorize every structure that corresponds to each format name. The expansion of the "zoo" described by Brinkmann and Kontermann[11] in 2017 is now stampeding over reliable scientific discourse between and within large research teams (Figure 1).[12,13]

To address this issue, Sweet-Jones et al.[14] developed AbML, a standardized notation language specialized for antibodies. The associated software, abYdraw, allows one to draw the schematic diagram of a specific antibody- or T cell receptor-based molecule and automatically generate the AbML text string which describes that molecule. These strings, however, are quite complex. For example, the AbML representation of a monoclonal antibody is:

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

One can input a given AbML string into abYdraw and retrieve the schematic diagram of the corresponding molecule. This approach addressed a critical issue, though the solution itself generated complex names that, at least in our environment, have been a barrier to adoption of AbML strings for common scientific discourse, especially given the popularity of convenient (but unstandardized) names like DVD-Ig and DART.[8,10,11,15–22] The complexity of AbML strings also obfuscates similarities and differences between molecules. For example, compare the AbML strings of a single-chain diabody fused to an Fc versus a single-chain diabody fused to a CH3:

Fc-fused:

VL.a(1:7)-L(2)-VL.a(3:5)-L(4)-VH.a(5:3)-L(6)-VH.a(7:1)-H(8:18) {2}-CH2(9:19)-CH3(10:20)|VL.b(11:17)-L(12)-VL.b(13:15)-L(14)-VH.b(15:13)-L(16)-VH.b(17:11)-H(18:8){2}-CH2(19:9)-CH3 (20:10)

CH3-fused:

VL.a(1:7)-L(2)-VL.a(3:5)-L(4)-VH.a(5:3)-L(6)-VH.a(7:1)-H(8:17) {2}-CH3(9:18)|VL.b(10:16)-L(11)-VL.b(12:14)-L(13)-VH.b (14:12)-L(15)-VH.b(16:10)-H(17:8){2}-CH3(18:9)

Ignoring the numbering indicating interacting partners, the two strings describe a similar list of domains; it is easy to miss
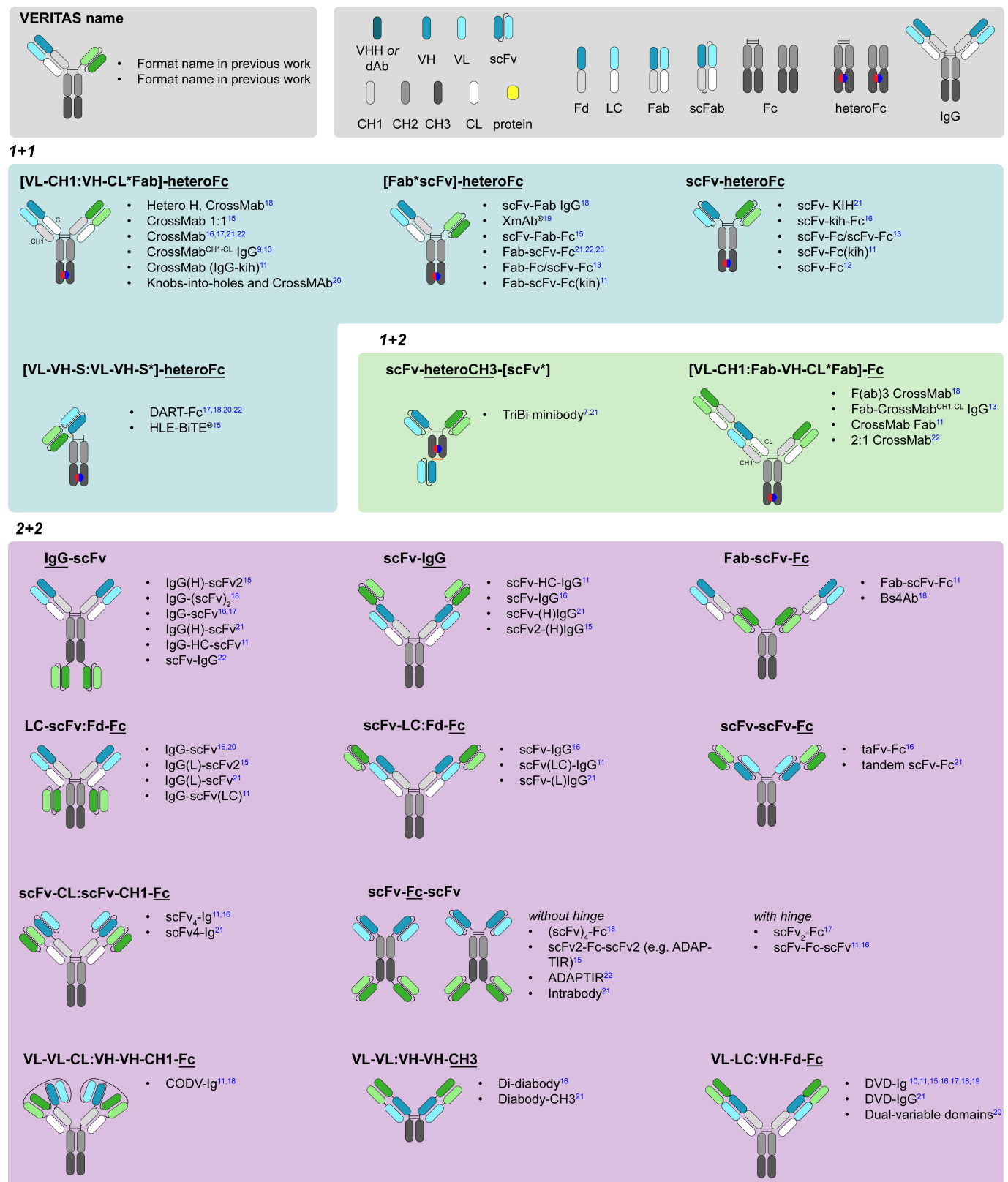

Figure 1. A lack of standardized nomenclature turns the "zoo" of multi-specific antibodies into a stampede. Diagrams from various prior works describing MsAbs are presented. VERITAS names of these molecule formats are bolded, while the various names assigned to them by prior work are listed as bullet points to the right of each diagram. Note that the same format is often referred to with different names across various works.

the absence of "CH2" in the second string. Regarding the numbering, there is a shift in numbering of the interacting partners between the two strings that belies the fact that these two molecules contain very similar interacting domains.

Hence, there remains an urgent need for a nomenclature system that: 1) is human-readable, 2) is standardized, and 3) relays general structure information at a customizable level of detail (either with or without specific target information).

Such a system would remove the necessity to research background information on a format to understand its structure, facilitate grouping of similar formats by their names, and aid in identifying formats based on specific features in which one is interested. Finally, having a systematic approach to multimeric proteins overall allows for consistency between projects and provides a more organized approach to discovering and developing fit-for-purpose protein therapeutics. In an ideal solution, schematic diagrams could be easily derived using the rules of the nomenclature system, and vice-versa, so that scientists who find it easier to communicate in images can do so with equivalent precision.

Here, we present such an approach for MsAbs. Formats are broken down into modular subunits and represented as a multimerization center plus N- and C-terminal appendages. We formalize this paradigm in text form with a simple set of symbols, leading to a systematic nomenclature scheme that is still easily understood. This scheme, VERITAS, is extensible beyond IgG-based formats and can theoretically produce systematic names for any multimer. Furthermore, its specificity can be customized: targets can be specified to differentiate between molecules, or unspecified to investigate correlations between molecular structure and attributes or functional properties. These text-based names can therefore be inputs to machine learning algorithms or used for automated format classification because they are rooted in structure.

The following example illustrates the VERITAS scheme and its ability to describe structure. Consider the molecule format called [Fab*scFv]-heteroFc in the 1 + 1 section of Figure 1. When the heteroFc is designated as the central focus, i.e., "multimerization center", of the molecule, the Fab and scFv are both attached to the N-terminus of the center. Amino acid sequences are read N-terminus to C-terminus and are therefore written left to right. Thus, it is logical to place "Fab" and "scFv" to the left of "heteroFc". Now, how can we relay the relationship between the Fab and the scFv? These modules are attached to two separate chains of the heteroFc, so let us use an asterisk ("*") between them. "Fab*scFv", which is the description of the N-terminal appendages to the "heteroFc" center, is enclosed in square brackets ("[]") for readability and to quickly indicate that this molecule format is asymmetric.

With the VERITAS scheme, all antibody-based formats can be broken down into various modules. The modules of every format can be classified as one of the following: 1) N-terminal appendages (e.g., VHH, VH, VL, CH1, CH2, CH3, CL, Fd, LC, scFv, scFab, protein); 2) C-terminal appendages (same as N-terminal appendages); and 3) Multimerization center (e.g., IgG, heteroIgG, Fc, heteroFc, Fab). Formats have only one multimerization center, but this center can have one, multiple, or no modules attached at its N and C termini (appendages). Any type of module can be an N- or C-terminal appendage, but only module types which are multimeric can be considered the center of a format. For example, a single "VH" can never be the center of a format because a VH domain alone is not composed of multiple amino acid chains. In contrast, "Fab" is composed of two chains and therefore can be the center of one format, but in a different format, "Fab" may be an N-terminal appendage.

As described in the example, the VERITAS scheme uses descriptors and symbols in specific ways to denote specific relationships. A dash ("-") is used between modules (or sets of modules) that occur on the same amino acid chain, whereas an asterisk ("*") is used between modules that are on separate chains. Figure 2A shows a selection of modules and their standard descriptors.

The most comprehensive descriptor is always used. For example, the definition of "LC" (light chain) is "VL-CL". Therefore, any time the sequence of modules "VL-CL" occurs in a format, they can be captured with the descriptor "LC". A similar logic applies to other compound modules like "Fab", "IgG", and "Fc".

For asymmetric formats (where the chains of the multimerization center do not have the same appendages), square brackets contain information about N- and C-terminal appendages. If the brackets occur before the center description, their contents describe the N-terminal appendages, whereas brackets after the center description contain information about the C-terminal appendages. The asterisk is used within these square brackets to separate modules which are attached to separate chains of the center, as shown in Figure 2Bi.

The scheme can also handle cases where there is a module appended to only one chain of the center but not the other, as demonstrated in Figure 2Bii. In this example, the C-terminal brackets (occurring after the center, "heteroFc") contain only "scFab" and an asterisk, with nothing on the other side of the asterisk. This denotes that only one chain of the heteroFc has a scFab appended to the C-terminus; the other chain of the heteroFc has nothing appended at the C-terminus.

When the appendages at one or both termini of the center are symmetric (i.e., both chains of the center have the same modules appended in the same order), square brackets and asterisks are omitted (Figure 2Biii).

If both the N- and C-terminal sets of appendages of a format are asymmetric, there will be two sets of square brackets in the VERITAS names. In such cases, we define the "left" and "right" sides of the molecule, and this distinction is maintained in both square brackets. That is: all modules described to the left of the asterisk in the N-terminal square brackets occur on the same chain ("left chain") of the center as the modules described to the left of the asterisk in the C-terminal brackets. The modules on the opposite chain ("right chain") are described to the right of the asterisk in both sets of square brackets. Figure 3 depicts the method used to decide which side of a format should be described to the left of the asterisk. This method prevents variations in names of formats that are physically identical but are named differently simply because there are two possible ways to draw their diagrams.

A colon (":") represents a noncovalent interaction between two chains which are part of the N- or C-terminal appendages (i.e., noncovalent interactions aside from those in the multimerization center). The rule of using the simplest descriptor extends to the usage of the colon. The definition of "Fab" is "LC:Fd" (where LC = light chain or VL-CL and Fd = heavy chain of the Fab or VH-CH1), so when such a complex is present (either as an appendage or as the multimerization center), it suffices to refer to it as
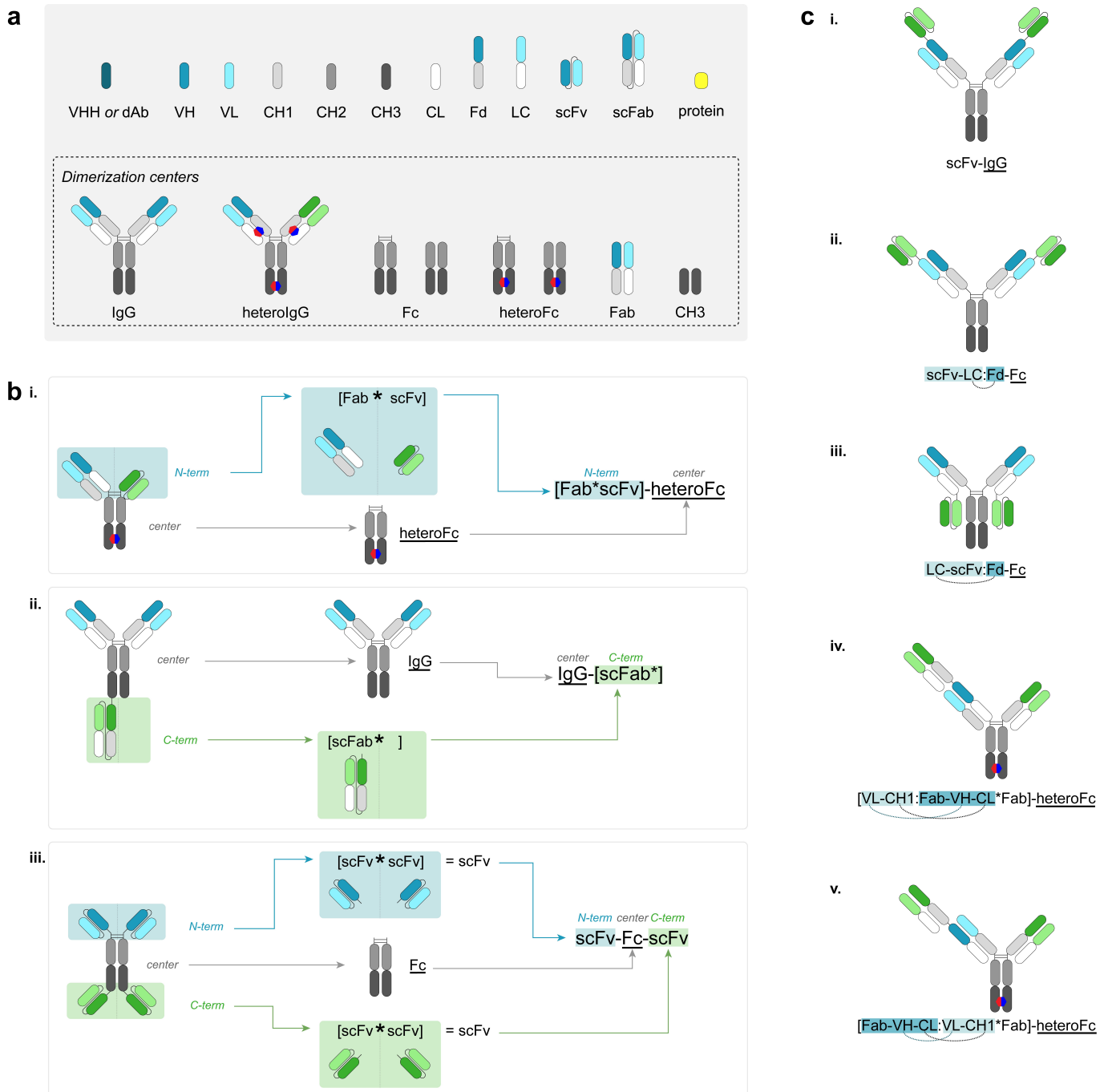
4 R. BISWAS ET AL.

Figure 2. An overview of the VERITAS nomenclature scheme. (A) Diagrams of various standard antibody parts that are used as "modules" in the VERITAS nomenclature scheme, as well as the standard descriptor text for these modules. The inset box shows modules that can serve as dimerization centers. Dimerization centers are underlined for readability, but underline can be omitted in applications where plaintext names are necessary. (B) Examples of implementations of the VERITAS scheme to (i) a format with only N-terminal appendages, (ii) a format with a C-terminal appendage on only one chain of the center, and (iii) a symmetric format. (C) Noncovalent interactions in the *N*- or C-terminal appendages are denoted with a colon (":") separating two interacting chains. In (i), scFv is appended to the heavy chain of the Fab (i. e., Fd). "scFv-Fab" implies that the scFv is appended to the Fd. (ii) shows a case where colon must be used. "scFv-LC:Fd" specifies that the scFv is appended to the light chain. Interacting pairs in chains may not be written directly before and after the colon, as shown in (iii)-(v), so the most biologically likely pair of modules is assumed to interact. (iv) shows an example of a molecule where there are multiple interacting partners in a pair of noncovalently interacting chains at the N-terminus. The modules described closest to the asterisk are directly attached to the center. Thus, while "VL-CH1:Fab-VH-CL" and "Fab-VH-CL:VL-CH1" describe the same set of modules, the former should be used in this molecule. If the latter were used, it would describe the slightly different molecule shown in (v).

"Fab", as in the case of Figure 2Ci. The only time this complex should be referred to as "LC:Fd" is if the LC (light chain) has an *N*- or C-terminal appendage, as in Figure 2Cii-iv. In these cases, it does not suffice to simplify "LC:Fd" to "Fab" because this introduces ambiguity as to which chain of the Fab – the LC or the Fd – the appendages are attached. A rule specifically for the term "Fab" arises from this: if "x" is an appendage to the *N*- or C-terminus of the Fd in a Fab, refer to this as "x-Fab" or "Fab-x", respectively; if "y" is an appendage to the *N*- or C-terminus of the light chain in a Fab, refer to this as "y-LC:Fd"/"Fd:y-LC" or "LC-y:Fd"/"Fd:LC-y", respectively.

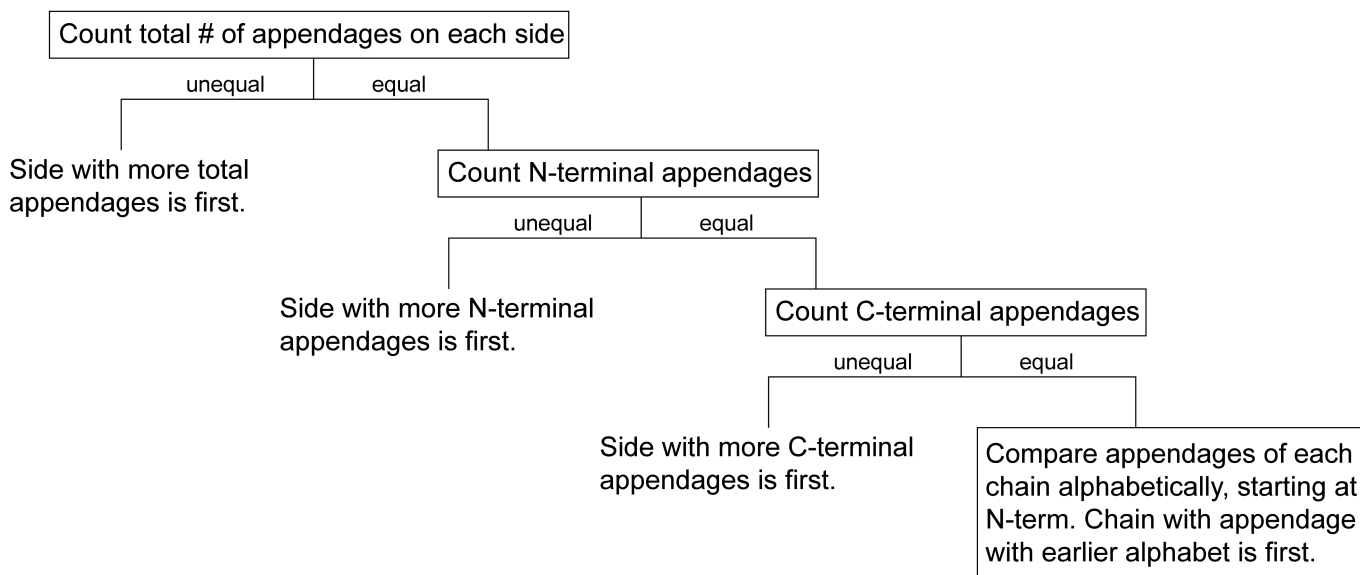Comparing the appendages to the two sides of the dimerization center:

Figure 3. Decision tree for defining order in which chains of the center are described. Since there are multiple ways to visually depict the same asymmetric format, this decision tree ensures that the "left" and "right" sides of an asymmetrical molecule are set using a standard logic to avoid inconsistencies in describing the same format based on different diagrams.

The case of "y" (appendage to the light chain of a Fab) can be described as either the light chain with appendages before ":" and Fd after (e.g., "y-LC:Fd" for N-terminal appendage), or vice-versa (e.g., "Fd:y-LC"). This enables another rule: modules that are most directly appended to the center must be written closest to the asterisk. For example, in Figure 2Civ, we could have written [Fab-VH-CL:VL-CH1*Fab]-heteroFc. However, based on the aforementioned rule, this would imply that the VL-CH1 chain is directly appended to the heteroFc, which is the slightly different format shown in Figure 2Cv. Thus, returning to our generic example, while "y-LC:Fd" and "Fd:y-LC" represent the same interaction between two chains, there are specific use cases for each.

When a colon is present, the interaction between the two chains described is assumed to occur mainly through the most biologically likely pair (which may recapitulate the physical reality of such interactions). This follows on to the previous rule, which often results in the interacting partners of each chain not necessarily being written directly on either side of the colon, e.g., "Fd:y-LC". In this case, we assume that Fd and LC form the main noncovalent interaction between the two chains, despite Fd and y being the modules described closest to the colon.

Finally, the level of detail contained in a VERITAS name is customizable. The format names described thus far are target-agnostic. For example, the format in Figure 2Cii could be: 1) monospecific and tetravalent (all four Fv domains have the same target), 2) bispecific and bivalent for two different targets (e.g., two anti-Target1 scFvs and two anti-Target2 Fabs), 3) bispecific and trivalent for a first target and monovalent for a second target, 4) trispecific and monovalent for a first target, monovalent for a second target, and bivalent for a third target, or 5) tetraspecific and monovalent for all four targets. The format name "scFv-LC:Fd-Fc" encompasses all these possible

molecules. To specify valency and specificity using VERITAS names, the targets of the Fvs are included in parentheses before the corresponding module, with the format name written in the expanded form (i.e., with square brackets) as needed. For example, one could differentiate between (2) and (3) as follows:

2) Bispecific and bivalent for two different targets
   (Target1)scFv-(Target2)LC:Fd-Fc
3) Bispecific and trivalent for one target, monovalent for another target
   [(Target1)scFv-(Target2)LC:Fd*(Target1)scFv-(Target1)LC:Fd]-Fc.

For a VERITAS format like Fab-heteroFc, the two Fab domains likely contain different Fvs. In such a case, when indicating the targets of each Fab in the expanded form, the domains are ordered within square brackets alphabetically, by target or gene name. For example:

[(TargetA)Fab*(TargetB)Fab]-heteroFc

This ensures that the same molecule is not described in two different ways, e.g.,

[(TargetB)Fab*(TargetA)Fab]-heteroFc *(incorrect, targets are not ordered alphabetically)*

The basic scheme is also point mutation-agnostic. The format in Figure 2Bi could use knobs-into-holes (KiH), charge-pair mutations (CPM), or any other strategy to heterodimerize the heteroFc. Again, this granularity can be added as desired using parentheses after the constant module to differentiate these details from information about targets:

[Fab*scFv]-heteroFc(KiH)

[Fab*scFv]-heteroFc(CPM)

The VERITAS scheme presents a framework to generate standardized and understandable names for asymmetric, antibody-based molecule formats and beyond. With this framework, any multimeric format can be broken down into modules. Those modules are then classified as either a multimerization center or appendages. Once the multimer center of a format is defined, VERITAS communicates the relationships between the center and the appendages. The main strength of this scheme is its ability to relay structural information about a format in a concise manner, without the need for diagrams.

This is useful for several reasons. First, this can address the current variability in names given to formats across previous publications. For example, the format with VERITAS name [VL-CH1:VH-CL*Fab]-heteroFc is called "Hetero H, CrossMab" by Labrijn et al.,[18] while the same format is called "CrossMab^{CH1-CL} IgG" by Ma et al.,[19] evidencing an inconsistent use (or disuse) of the descriptor "IgG" (Figure 1). Similarly, the format in Figure 2Bi with VERITAS name [Fab*scFv]-heteroFc has been variously called "scFv-Fab IgG" and "Fab-scFv-Fc", among other names.[18,21–23] VERITAS imposes rules on the usage of descriptors to avoid ambiguity and to ensure that the structure of a format can be easily derived from its name. It is also simple to add new definitions of descriptors as needed, adding to the flexibility and forward-looking extensibility of the scheme.

Second, VERITAS names are unambiguously correlated to format structure, a feature which is lacking in the current, colloquial names used to describe antibody formats. For example, consider the molecule which has been described as (scFv)4-Fc, scFv2-Fc-scFv2, ADAPTIR, or intrabody.[15,18,21,22] Inconveniently, none of these names considered individually provides enough information to derive the structure of the format. (scFv)4-Fc could describe many different molecules, such as one with two scFvs appended to the N-terminus of each Fc chain. scFv2-Fc-scFv2 does not specify whether the two scFvs described before and after the Fc are on the same chain or different chains. ADAPTIR and intrabody do not intrinsically imply anything about the format without further context about the meaning of these names. The VERITAS name for this format is scFv-Fc-scFv. Given the rules of the VERITAS scheme, we can derive the structure. There are no square brackets or asterisks in this name, so the molecule is symmetrical. The center of the molecule is the Fc. Both chains of the Fc have an scFv appended at the N-terminus *and* the C-terminus.

Colloquial nomenclature can lead to errors. For example, in one review article, the format diagrams for "HLE-BiTE" and "bi. DART-Fc" (VERITAS scFv-scFv-scFc and [VH-VL: VL-VH*]-Fc, respectively) are swapped (whereas the diagrams for BiTE and DART alone are correct).[15] It is conceivable that such mistakes could become more prevalent as the MsAb format landscape expands and diversifies, which could be inconvenient or even costly for researchers and the biopharmaceutical industry. In contrast, VERITAS names describe the exact relationship between different modules in a format, including differentiation between different amino acid chains in a multimer, relative orientation of various modules on these chains, and noncovalent interactions between different chains. Therefore, VERITAS decreases the chances for errors arising from nomenclature because the structure of a format can be derived from its name.

We believe this system is future-proof because it is extensible to a wide variety of protein-based multimeric formats (Figure 4a-b). For essentially any antibody-based molecule developed in the future, simply following the rules of the scheme will produce a standardized name for a new format, allowing for it to be easily discussed and compared to other formats. The scheme can also theoretically handle non-antibody-based multimeric proteins by simply adding additional asterisks within the *N-* and *C*-terminal brackets to indicate the appendages to each subunit of a multimer center. In this case, however, if there are noncovalent interactions in the appendages that need to be called out with a colon, the rule about the chain that is directly linked to the center being closest to the asterisk must be given an addendum: the description of the module that is directly linked to the *first chain of the center* should be closest to the *left* of the *first asterisk*; any following noncovalent interaction descriptions should be formatted such that the module directly linked to its respective center chain is closest to the *right* of its *preceding* asterisk (Figure 4Biii).

Antibody-drug conjugates are not specifically handled in this scheme; drug conjugation is a modification that is currently out-of-scope.

Adopting the VERITAS scheme in MsAb research and development can benefit research efforts. Because the orientation of the modules in these formats is immediately apparent, it is easier to list options of formats when interrogating a biological mechanism of interest, facilitating the brainstorming process both for initial programs and for backup molecules for ongoing programs. Integrating VERITAS with tools like BioFonts and BioMaps would also enhance biologics research.[24] VERITAS could act as a human-readable input method with which users can interact with these applications to create diagrams.

Additionally, by updating historical data about molecules and their formats, it is conceivable that these molecule names could be inputs to a machine learning algorithm. With the direct correlation between these text-based names and structure, and the flexibility to include or omit information about valency, specificity, and point mutations, it may be possible to deduce the effect of molecular structure on molecular attributes and functional properties, which could inform future research directions.

Finally, if these descriptive but concise names are adopted into scientific literature, they can improve accessibility of that literature for readers with visual impairments, both by reducing the reliance on diagrams to understand the format of molecules and by serving as the standard alternate text for diagrams of antibody-based formats.

VERITAS was developed based on a view of antibody formats as modular structures and the idea of a "toolbox" of multimerization modules and antigen-binding modules as
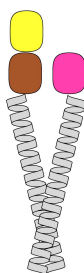
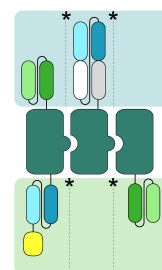**a** **i.** Fab-<u>heteroFc</u>-[scFv-scFv-scFv*scFv-scFv]
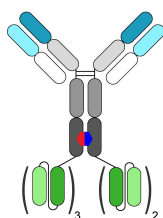
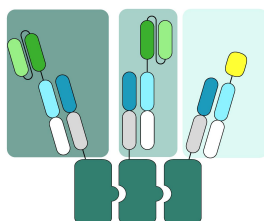**ii.** [protein:protein-Fab*protein-Fd:LC-protein]-<u>heteroFc</u>-[scFv-Fab*]
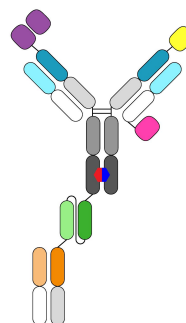
b **i.** [protein-protein*protein]-<u>leucine zipper</u>

ii. [scFv*scFab*]-<u>protein heterotrimer</u>-[scFv-protein**scFv]

iii. ✔ **YES** ✔
[scFv-LC:Fd*Fd:scFv-LC*Fd:protein-LC]-<u>protein heterotrimer</u>

[scFv-LC:Fd*scFv-LC:Fd*Fd:protein-LC]-<u>protein heterotrimer</u>
✘ **NO** ✘

Figure 4. VERITAS is extensible to (A) future antibody-based formats and (B) non-antibody-based multimeric proteins. (Biii) shows an addendum to rule for describing noncovalent interactions in the terminal appendages of a format. In all appendage descriptions (descriptions within square brackets), there is a "most proximal" module which is directly linked to the center. In asymmetric formats with a dimer center, the most proximal modules of each chain are placed directly next to the asterisk. However, in the case of the asymmetric trimer given in this figure, it is not possible to place the most proximal module of the second chain (Fd) directly next to both the first and second asterisks. Thus, the following rule is implemented: for an asymmetric format with more than two chains in its multimerization center, the most proximal module of the first (leftmost) chain is written to the left of the first asterisk. For the second, third, etc. chains, the most proximal module is written to the right of the preceding asterisk.

described by Brinkmann and Kontermann.[11] There were several requirements that the scheme needed to fulfill.

First, the scheme needed to be human-readable. It was decided that information about specificity, valency, and point mutations within constant regions could be out-of-scope for the basic scheme to limit the length of format names. Details about specific molecules could be added back in using parentheses (as described above).

Second, the scheme needed to differentiate between very similar molecules (e.g., IgG with two C-terminal scFvs versus IgG with one C-terminal scFv). Thus, there needed to be specific symbols to represent specific relationships between modules (without requiring special symbols that would prevent usage in computer databases). The dash was chosen to represent modules on the same amino acid chain because this is already commonly used in many existing formats (e.g., HLE-BiTE®, IgG-scFv). Square brackets were chosen to represent N- and C-terminal appendage information to differentiate from parentheses, which were used to add details about targets, and because curly braces were not supported by an internal database. The asterisk was chosen to separate modules on different chains because it is a minimally obtrusive symbol that allows for readability of the scheme (other options considered were pipe |, but this was too similar to the square brackets and more

obtrusive to readability; and carat ^, but this was incompatible with the internal database). Colon was chosen to represent noncovalent bonds for a similar reason. We have underlined the multimerization center in the VERITAS names described in this text for readability, but because the center domain can be determined logically, underlining is omitted from VERITAS format names in applications where plaintext is required.

Third, the scheme should be extensible to all existing and future formats while remaining standardized. To achieve this, the general rules of the scheme were set up: build format names around a main multimerization center, describe *N*- and C-terminal appendages to the chains of that center systematically, and follow specific rules to decide the order in which modules should be described to avoid variation in names for the same format. However, there were no limitations placed on the modules that can be described, so new module descriptions can easily be adopted as needed to describe new formats.

Abbreviations

AbML	Antibody Markup Language
ADAPTIR	Aptevo Therapeutics' bispecific antibody platform
ATTACK	Asymmetric Tandem Trimerbody for T cell Activation and Cancer Killing
BiTE®	Bispecific T cell Engager
CD3	cluster of differentiation 3
CH1	constant domain 1 of heavy chain
CH2	constant domain 2 of heavy chain
CH3	constant domain 3 of heavy chain
CL	constant domain of light chain
CODV	cross-over dual variable
CPM	charge-pair mutation
DART	Dual Affinity Re-Targeting
DVD-Ig	dual variable domain immunoglobulin
EGFR	EGFR – epithelial growth factor receptor
Fab	fragment antigen binding (composed of VH, CH1, VL, and CL)
Fc	fragment crystallizable (composed of a dimer of CH1, CH2, and CH3)
Fv	variable region of antibody (composed of VH and VL)
heteroFc	heterodimeric Fc
Fd	fragment difficult (composed of VH and CH1)
HLE	half-life extended
Ig	immunoglobulin
IgG	immunoglobulin G
MsAb	multispecific antibody
KiH	knobs-into-holes
LC	light chain of antibody (composed of VL and CL)
scFab	single chain Fab
scFv	single chain Fv
VERITAS	Verified Taxonomy for Antibodies
VH	variable domain of heavy chain
VHH	variable domain of heavy chain only antibody
VL	variable domain of light chain

ORCID

Riti Biswas http://orcid.org/0000-0003-3766-1980
Kevin Graham http://orcid.org/0000-0003-3969-4604
Christine E. Tinberg http://orcid.org/0000-0002-6179-0435
Alan J. Russell http://orcid.org/0000-0001-5101-4371

References

1. Kaplon H, Chenoweth A, Crescioli S, Reichert JM. Antibodies to watch in 2022. MAbs. 2022;14(1):2014296. doi:10.1080/19420862.2021.2014296.
2. Reichert JM. Antibodies to watch in 2014. MAbs. 2014;6(1):5–14. doi:10.4161/mabs.27333.
3. Klinger M, Benjamin J, Kischel R, Stienen S, Zugmaier G. Harnessing T cells to fight cancer with BiTE®antibody constructs - past developments and future directions. Immunol Rev. 2016;270(1):193–208. PMID: 26864113. doi:10.1111/imr.12393.
4. Ellerman D. Bispecific T-cell engagers: towards understanding variables influencing the in vitro potency and tumor selectivity and their modulation to enhance their efficacy and safety. Methods. 2019;154:102–17. doi:10.1016/j.ymeth.2018.10.026. PMID: 30395966.
5. Steinmetz A, Vallee F, Beil C, Lange C, Baurin N, Beninga J, Capdevila C, Corvey C, Dupuy A, Ferrari P, et al. CODV-Ig, a universal bispecific tetravalent and multifunctional immunoglobulin format for medical applications. MAbs. 2016;8:867–78. PMID: 26984268. doi:10.1080/19420862.2016.1162932.
6. Holliger P, Prospero T, Winter G. Diabodies: small bivalent and bispecific antibody fragments. Proc Natl Acad Sci U S A. 1993;90(14):6444–48. doi:10.1073/pnas.90.14.6444.
7. Shahied LS, Tang Y, Alpaugh RK, Somer R, Greenspon D, Weiner LM. Bispecific minibodies targeting HER2/neu and CD16 exhibit improved tumor lysis when placed in a divalent tumor antigen binding format. J Biol Chem. 2004;279(52):53907–14. PMID: 15471859. doi:10.1074/jbc.M407888200.
8. Johnson S, Burke S, Huang L, Gorlatov S, Li H, Wang W, Zhang W, Tuaillon N, Rainey J, Barat B, et al. Effector cell recruitment with novel Fv-based dual-affinity re-targeting protein leads to potent tumor cytolysis and in vivo B-cell depletion. J Mol Biol. 2010;399(3):436–49. PMID: 20382161. doi:10.1016/j.jmb.2010.04.001.
9. Schaefer W, Regula JT, Bahner M, Schanzer J, Croasdale R, Durr H, Gassner C, Georges G, Kettenberger H, Imhof-Jung S, et al. Immunoglobulin domain crossover as a generic approach for the production of bispecific IgG antibodies. Proc Natl Acad Sci U S A. 2011;108(27):11187–92. PMID: 21690412. doi:10.1073/pnas.1019002108.
10. Wu C, Ying H, Grinnell C, Bryant S, Miller R, Clabbers A, Bose S, McCarthy D, Zhu RR, Santora L, et al. Simultaneous targeting of multiple disease mediators by a dual-variable-domain immunoglobulin. Nat Biotechnol. 2007;25(11):1290–97. PMID: 17934452. doi:10.1038/nbt1345.
11. Brinkmann U, Kontermann RE. The making of bispecific antibodies. MAbs. 2017;9(2):182–212. PMID: 28071970. doi:10.1080/19420862.2016.1268307.
12. Gunasekaran K, Pentony M, Shen M, Garrett L, Forte C, Woodward A, Ng SB, Born T, Retter M, Manchulenko K, et al. Enhancing antibody Fc heterodimer formation through electrostatic steering effects: applications to bispecific molecules and

monovalent IgG. J Biol Chem. 2010;285(25):19637–46. PMID: 20400508. doi:10.1074/jbc.M110.117382.

13. Ha JH, Kim JE, Kim YS. Immunoglobulin Fc Heterodimer platform technology: from design to applications in Therapeutic antibodies and proteins. Front Immunol. 2016;7:394. PMID: 27766096. doi:10.3389/fimmu.2016.00394.

14. Sweet-Jones J, Ahmad M, Martin ACR. Antibody markup language (AbML) — a notation language for antibody-based drug formats and software for creating and rendering AbML (abYdraw). MAbs. 2022;14(1):2101183. PMID: 35838549. doi:10.1080/19420862.2022.2101183.

15. Elshiaty M, Schindler H, Christopoulos P. Principles and current clinical landscape of multispecific antibodies against cancer. Int J Mol Sci. 2021;22(11):5632. PMID: 34073188. doi:10.3390/ijms22115632.

16. Kontermann RE. Dual targeting strategies with bispecific antibodies. MAbs. 2012;4(2):182–97. PMID: 22453100. doi:10.4161/mabs.4.2.19000.

17. Kontermann RE, Brinkmann U. Bispecific antibodies. Drug Discov Today. 2015;20:838–47. PMID: 25728220. doi:10.1016/j.drudis.2015.02.008.

18. Labrijn AF, Janmaat ML, Reichert JM, Parren P. Bispecific antibodies: a mechanistic review of the pipeline. Nat Rev Drug Discov. 2019;18(8):585–608. PMID: 31175342. doi:10.1038/s41573-019-0028-1.

19. Ma J, Mo Y, Tang M, Shen J, Qi Y, Zhao W, Huang Y, Xu Y, Qian C. Bispecific antibodies: from research to clinical application. Front Immunol. 2021;12:626616. PMID: 34025638. doi:10.3389/fimmu.2021.626616.

20. Fan G, Wang Z, Hao M, Li J. Bispecific antibodies and their applications. J Hematol Oncol. 2015;8(1):130. PMID: 26692321. doi:10.1186/s13045-015-0227-0.

21. Spiess C, Zhai Q, Carter PJ. Alternative molecular formats and therapeutic applications for bispecific antibodies. Mol Immunol. 2015;67 (2):95–106. PMID: 25637431. doi:10.1016/j.molimm.2015.01.003.

22. Moore GL, Bernett MJ, Rashid R, Pong EW, Nguyen DT, Jacinto J, Eivazi A, Nisthal A, Diaz JE, Chu SY, et al. A robust heterodimeric Fc platform engineered for efficient development of bispecific antibodies of multiple formats. Methods. 2019;154:38–50. PMID: 30366098. doi:10.1016/j.ymeth.2018.10.006.

23. Kunz RK, Rojnuckarin A, Schmidt CM, Miranda LP. Development of human-machine language interfaces for the visual analysis of complex biologics and RNA modalities and associated experimental data. AAPS Open. 2023;9(1):9. doi:10.1186/s41120-023-00073-w.

24. Suurs FV, Lub-de Hooge MN, de Vries EGE, de Groot DJA. A review of bispecific antibodies and antibody constructs in oncology and clinical challenges. Pharmacol Ther. 2019;201:103–19. PMID: 31028837. doi:10.1016/j.pharmthera.2019.04.006.