In Vitro Evolution of Antibodies Inspired by *In Vivo* Evolution

Helena Persson[1,2], Ufuk Kirik[3†], Linnea Thörnqvist[3], Lennart Greiff[4,5], Fredrik Levander[3] and Mats Ohlin[3,6]*

[1] Drug Discovery and Development Platform, Science for Life Laboratory, Stockholm, Sweden, [2] School of Engineering Sciences in Chemistry, Biotechnology and Health, Royal Institute of Technology, Stockholm, Sweden, [3] Department of Immunotechnology, Lund University, Lund, Sweden, [4] Department of Clinical Sciences, Lund University, Lund, Sweden, [5] Department of Otorhinolaryngology, Head & Neck Surgery, Skåne University Hospital, Lund, Sweden, [6] Human Antibody Therapeutics, Drug Discovery and Development Platform, Science for Life Laboratory, Lund University, Lund, Sweden

In vitro generation of antibodies often requires variable domain sequence evolution to adapt the protein in terms of affinity, specificity, or developability. Such antibodies, including those that are of interest for clinical development, may have their origins in a diversity of immunoglobulin germline genes. Others and we have previously shown that antibodies of different origins tend to evolve along different, preferred trajectories. Apart from substitutions within the complementary determining regions, evolution may also, in a germline gene-origin-defined manner, be focused to residues in the framework regions, and even to residues within the protein core, in many instances at a substantial distance from the antibody's antigen-binding site. Examples of such germline origin-defined patterns of evolution are described. We propose that germline gene-preferred substitution patterns offer attractive alternatives that should be considered in efforts to evolve antibodies intended for therapeutic use with respect to appropriate affinity, specificity, and product developability. We also hypothesize that such germline gene-origin-defined *in vitro* evolution hold potential to result in products with limited immunogenicity, as similarly evolved antibodies will be parts of conventional, *in vivo*-generated antibody responses and thus are likely to have been seen by the immune system in the past.

Keywords: affinity maturation, antibody, antibody therapeutics, developability, evolution, humanization, immunoglobulin germline gene, somatic hypermutation

THE NEED FOR ANTIBODY EVOLUTION

Biopharmaceutical drugs, in particular monoclonal antibodies, have transformed treatment of disease in recent years. Currently (December 2017), more than 50 antibodies are approved for therapeutic use in the USA and/or the EU, 10 of which received their first approval in 2017 (1). Furthermore, nine antibodies are under regulatory review by the European Medicines Agency or the US Food and Drug Administration, and many more are undergoing evaluation in late stage clinical trials (1, 2). The large revenue associated with sales of many antibodies in clinical use, and the immense clinical benefit of antibodies in the treatment of a diversity of diseases, have stimulated enormous efforts and approaches that aim to develop antibodies for use in the clinic (2). These approaches include, e.g., utilization of well-behaving antibody scaffolds upon which antibodies are built by genetic engineering (3, 4), development of transgenic animals with human immunoglobulin genes

---

Abbreviations: CDR, complementary determining region; FR, framework region; H, heavy; NGS, next-generation sequencing; V, variable.

that allow development of human antibodies in an *in vivo* setting (5, 6), humanization of highly effective antibodies of non-human (typically murine) origin (7), and exploitation of human *in vivo* immune responses for identification of high quality antibodies with unique properties (8–10). Antibodies with multiple specificities [generated either through development of dual-specific binding sites (11) or by design of bispecific (or higher order) constructs (12)] may engage appropriate effector cells to target cells carrying particular antigens, or facilitate antibody binding specifically to cells that carry both antigens engaged by the antibody's specificities. Antibodies can also be further developed by selection of appropriate constant domains (natural or those designed for function) and glycosylation patterns so as to assure that they display appropriate effector functions [e.g., mediated through FcR binding (including antibody-dependent cellular cytotoxicity) or complement activation] and biological half-life (mainly an effect of FcRn binding) suitable for the intended application. Importantly, processes for identification and development of antibodies with appropriate properties in terms of "Chemistry, Manufacturing, and Control" have also been developed (13, 14) to ensure that a potential drug candidate can actually be turned into a viable commercial product.

The biophysical properties of antibodies as well as the affinity and specificity of the binding site can be affected by diversification of the variable (V) domains and subsequent selection of variants that display appropriate properties. Such modulation of antibody properties is often achieved by random mutagenesis, directed mutagenesis primarily of residues in complementarity determining regions (CDRs) (based on the assumption that CDRs carry much of the specificity-determining diversity of human antibodies), DNA shuffling, or chain shuffling, and subsequent selection of variants with improved properties. Selection can be achieved using a diversity of technologies like those relying on display on phage, yeast, or bacteria and the processes can achieve substantial maturation of e.g. the affinity, which may translate into improved performance in a given therapeutic setting. Ribosomal display even carries an inherent ability to evolve genes during the selection procedure (15) thereby advancing a maturation process, as selection from an original library proceeds. Given the potential size of molecular diversity space, it is only possible to assess a very limited fraction of that space in any given experiment. It is thus important to limit diversity to that acceptable for appropriate protein folding, as only such diversity is likely to deliver a functional product and thereby an appropriate outcome of the evolution process. Furthermore, it is envisaged that minimal diversification away from human antibody sequences that commonly occur *in vivo* is preferred in order to minimize the likelihood that the protein will be immunogenic and able to induce an anti-drug immune response during clinical use.

BIG DATA AS A SOURCE OF INSPIRATION FOR DEVELOPMENT OF MOLECULAR EVOLUTION STRATEGIES

Big datasets, in particular those that are the results of next-generation sequencing (NGS), now offer extensive, unprecedented insight into immune antibody diversity as it is represented in human subjects (16). High-throughput single cell approaches to identify paired antibody heavy (H) and light chain sequences used in combination with NGS, and occasionally in combination with mass spectrometry-based proteomics, offer an additional dimension to antibody investigation and discovery that aid our understanding of immune repertoires also at a global level (17, 18). We are confident that this kind of large-scale insight can be translated into a multitude of actionable strategies in the field of improved antibody engineering and evolution. Exploitation of such technologies have indeed already defined routes through which immune responses develop and evolve in response to natural infection and vaccination (19, 20). Such studies have now provided a driving force for novel vaccine development in the field of infectious diseases (21–23).
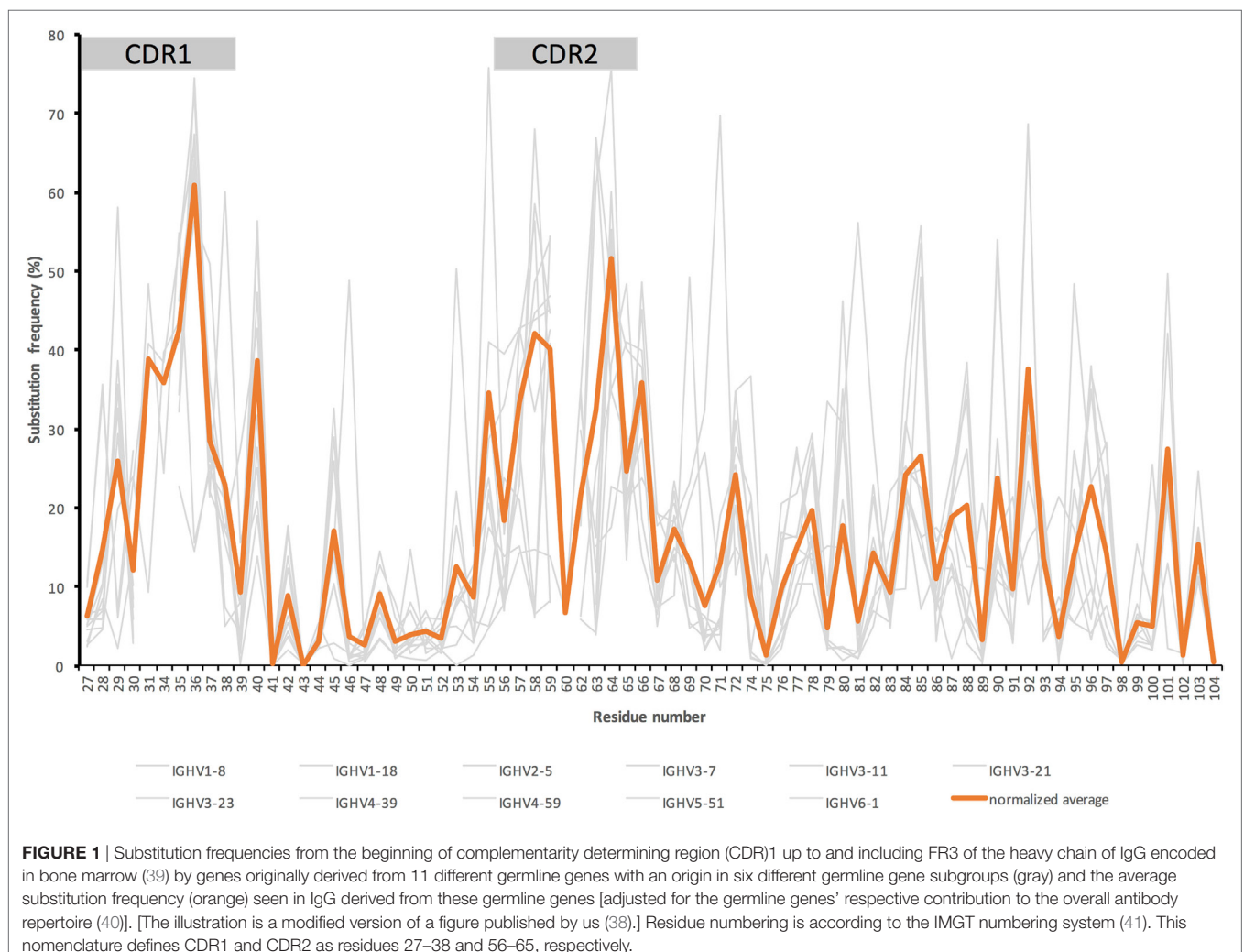
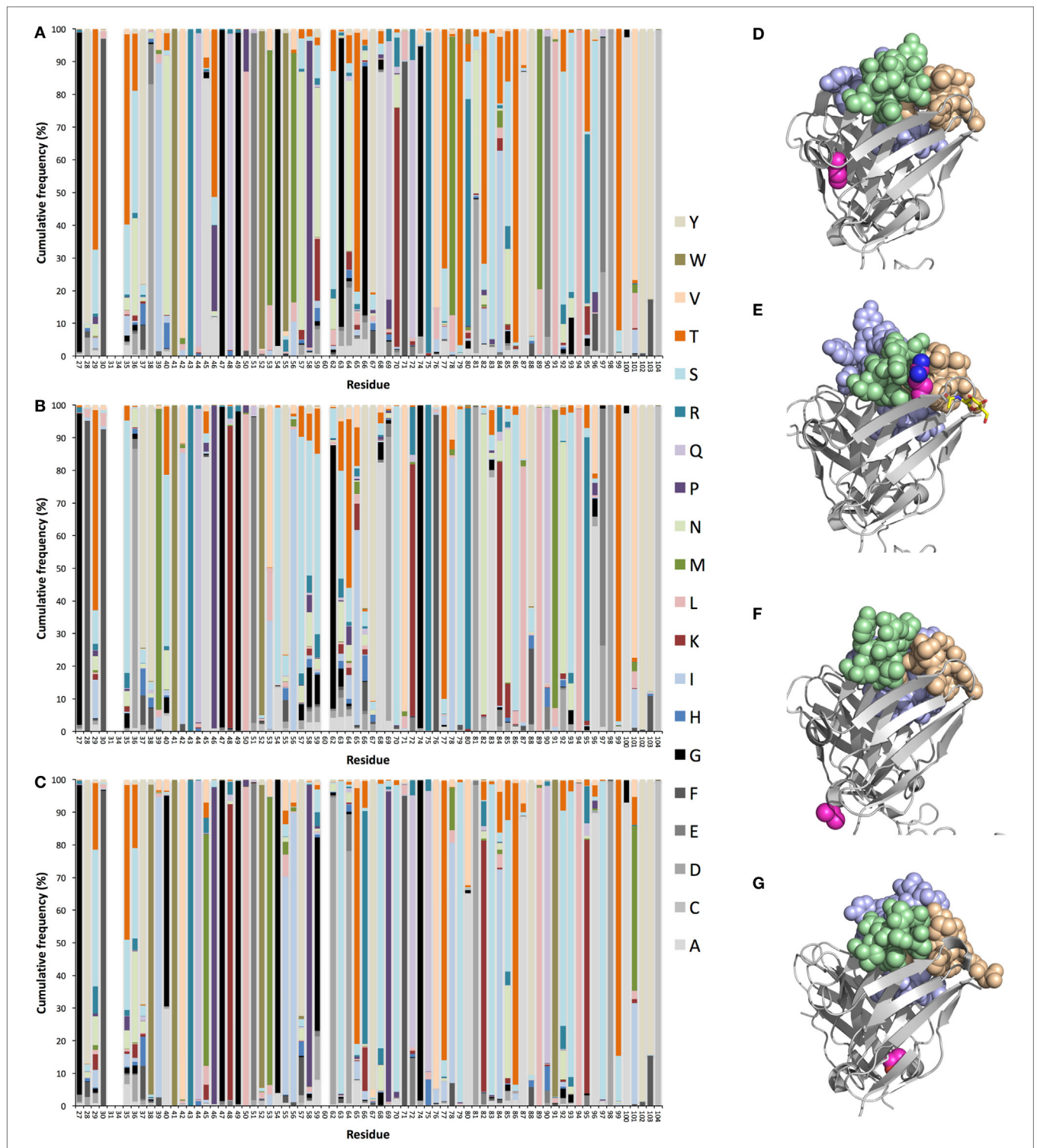## ANTIBODY EVOLUTION BEYOND CONVENTIONAL CDR—GENERAL CONSIDERATIONS

Importantly, large collections of diversified antibody sequences can also be used to inspire new thinking in the field of antibody evolution, in order to develop processes that efficiently evolve antibodies into variants that display appropriate binding properties and high developability. It was early on realized that antibody diversity at a global scale is focused into regions, the so-called CDRs, while intervening framework regions are more similar in sequence between different antibodies (24), in particular among sequences that belong to one and the same immunoglobulin clan (25). Such diversity is generated through variability in immunoglobulin germline-encoded genes, by the gene rearrangement process (focusing additional diversity into the third CDR), and by somatic hypermutation that may target mutational hotspots and substitution-prone codons that are focused to the parts of the genes that encode CDRs (26–29). These CDRs are well recognized to establish multiple direct interactions with antigen, although some interactions may certainly occur outside of these regions as well (25, 30, 31). Indeed, other parts of the immunoglobulin V domains have also been associated with functional antibody evolution. Studies of antibody *in vitro* evolution have suggested that "*mutations leading to higher affinity correspond to residues distant from the binding site*" (32) and that "*affinity maturation of antibodies with affinity in the low nanomolar range occurs most effectively via changes in "vernier" or second-sphere residues rather than contact residues*" (33). Furthermore, studies of antibody evolution *in vivo* suggested that "*There is a clear preference for mutations at the Ag-binding site. However, positions outside this region that also affect binding are often preferred targets for somatic hypermutations*" (34), but also that "*mutations in the contacting residues have an adverse effect on the antigen–antibody interaction*" (35), and that "*FWR mutations in noncontact residues are essential for the binding, breadth, and potency of most broadly neutralizing anti-HIV-1 antibodies*" (36). Although diversity in antibodies, generated at the level of diversity of germline genes themselves and through

the hypermutation process, at a global scale focus onto the CDRs (**Figure 1**), others (37) and we (38) have demonstrated that diversity-generating hypermutation extends far beyond traditional CDRs in a manner that is defined by the germline gene origin of the antibody-encoding gene. Multiple sites, targeted by diversification in products of individual germline genes, can be identified in antibody H chains (**Figures 2A–C**; Figure S1 in Supplementary Material) (37, 38) and light chains (37). It thus appears that antibodies *in vivo* evolve through unique paths. These paths may be regulated by features built into the gene sequence itself (mutational hot- and cold-spots) or into features of their encoded proteins that favors evolution along particular trajectories. Irrespective of the specific reasons for the preferred paths of evolution, it is obvious that the immune system as such will have experienced and tolerated a substantial level of sequence diversity in antibodies, both within the CDRs and well beyond. This opens a window of opportunity for antibody *in vitro* optimization, the products of which, we hypothesize, may be well tolerated from an immunogenicity perspective in a therapeutic setting as it is already exploited by humoral immunity.

ANTIBODY EVOLUTION EVEN BEYOND CONVENTIONAL CDR—SPECIFIC EXAMPLES

The sites available for natural evolution reside throughout the V domain of the H chain. Some germline genes offer opportunities for amino acid substitution in core residues, in H and light chain V domain interphase residues and in surface residues situated close to or far away from the paratope itself (38). A fourth hypervariable loop (CDR4) (42) located within FR3, but in close proximity to CDR1 and CDR2 of the folded protein, a region that may directly contact antigen (30, 43), may be particularly attractive for this purpose in the case of antibodies derived from some germline genes, such as *IGHV1–8*, *IGHV1–18*, and *IGHV5–51* (38). We also recently demonstrated patterns of insertions and deletions of entire codons, implied in the diversification of antibodies 20 years ago (44–46), that specifically target CDRs (including CDR4) of antibodies encoded by some germline genes (38). As such, diversification may be an important contributor to the development of particularly effective immune responses *in vivo* (47) it should be considered for *in vitro* antibody evolution

FIGURE 1 | Substitution frequencies from the beginning of complementarity determining region (CDR)1 up to and including FR3 of the heavy chain of IgG encoded in bone marrow (39) by genes originally derived from 11 different germline genes with an origin in six different germline gene subgroups (gray) and the average substitution frequency (orange) seen in IgG derived from these germline genes [adjusted for the germline genes' respective contribution to the overall antibody repertoire (40)]. [The illustration is a modified version of a figure published by us (38).] Residue numbering is according to the IMGT numbering system (41). This nomenclature defines CDR1 and CDR2 as residues 27–38 and 56–65, respectively.

FIGURE 2 | Diversification of heavy (H) chain variable (V) domain of IgG encoded in bone marrow (39) by genes originally derived from germline genes *IGHV1–8* **(A)**, *IGHV3–11* **(B)**, and *IGHV5–51* **(C)**, as previously described (38). Illustrations of such substitution patterns for an additional eight genes are available in Figure S1 in Supplementary Material. Structures of Fv of antibodies with H chain V domains with an origin in *IGHV1–18* (PDB: 3SDY) **(D)**, *IGHV1–8* (PDB: 3 × 3G) **(E)**, *IGHV3–11* (PDB: 4ZS6) **(F)**, and *IGHV5–51* (PDB: 4BUH) **(G)** illustrating examples of side chains of highly mutated positions [side chain atoms of which are shown in magenta (carbon), dark blue (nitrogen), and red (oxygen)]. The highlighted, highly mutated positions are position 71 of *IGHV1–18* **(D)**, position 80 of *IGHV1–8* **(E)**, position 96 of *IGHV3–11* **(F)**, and position 101 of *IGHV5–51* **(G)**. Atoms of H chain complementarity determining regions (CDRs) are illustrated in light brown (CDR1), light green (CDR2), and light blue (CDR3). All residue numbers are in accordance with the IMGT numbering rules (41). This nomenclature defines CDR1 and CDR2 as residues 27–38 and 56–65, respectively.

purposes, possibly allowing us to reach parts of structure and antigen-interaction space not effectively reached by antibodies in other ways. Many other residues beyond CDRs have also been identified to carry diversity in a germline gene-defined manner. For instance, residue 80, a member of the V domain's upper core (48), considered to be important for the position and conformation of H chain CDR2 (49) and stability of the V domain (50), is extensively targeted by substitution in products of some but not all germline genes (**Figure 2**) (38). Intriguingly, arginine 80 encoded by *IGHV1–8* (**Figures 2A,E**), but not by *IGHV3–7*, *IGHV3–11*, *IGHV3–21*, or *IGHV3–23* (38), is, based on diversification of H chains *in vivo*, a suitable target for diversification. Indeed, multiple other residues within the upper core are targetable by substitution in a germline-defined manner (38), modifications that may fine-tune the paratope (48, 51) and thereby perfect the antigen-binding properties. Indeed, evolution that might affect the structure/flexibility of H chain V domain CDR3 (52), the loop created through the VDJ rearrangement process, provides interesting opportunities in terms of perfection of the paratope. For instance, we recently observed substantial germline-defined differences in the ability to incorporate the 3′-most nucleotide of the IGHV-encoding gene into the rearranged sequence (53, 54). Many IGHV genes' 3′-end has a sequence that it, if fully incorporated into the final product, would encode an acidic residue in position 107 of the ascending strand of CDR3. These side chains display opportunities to establish polar interactions, for instance to the side chain of amino acid 40 (located immediately after CDR1), not available to residues (commonly glycine, alanine, or valine) that are incorporated in cases where the full length of the IGHV gene is not used (53, 54). Importantly, the germline-encoded side chain of residue 40 is in many instances able to participate in polar interactions with a polar residue at position 107 in the ascending strand of H chain CDR3, and it holds capacity to be diversified (Figure S2 in Supplementary Material) (38). We postulate that diversification of residue 40 may be attempted as yet another way to modulate the structure of the binding site itself even beyond the nature of the residues in the paratope itself. Altogether, there are multiple germline origin-defined opportunities for paratope evolution in antibodies that may be exploited to achieve products with sequence similarity to antibodies that typically occur in human subjects.

Antibodies may need optimization for utilization in specific applications for reasons other than specificity and/or affinity. Developability of a candidate molecule is such an aspect that may require molecular optimization to provide a final candidate with properties like the absence of off-target binding, efficiency of manufacture, high stability, and ability to be incorporated in a suitable, formulated product. Indeed, antibodies, even those approved or in late stage clinical development, may differ widely in such respects (14). For instance, some side chains are sensitive to, e.g., oxidation, β-elimination, deamidation, or isomerization, and the protein may be poorly soluble, or sensitive to proteolysis and aggregation, processes that reduce the stability/homogeneity of the protein product. When such problems are identified in a lead antibody that is undergoing clinical development it may be necessary to optimize the protein sequence through molecular evolution (55), potentially through guidance by computational predictions (56), to keep these undesired processes at an acceptable level. It has been suggested that antibodies developed by phage display technology may be more prone to developability issues in relation to product specificity or solubility than antibodies developed in mice (57, 58), suggesting that such *in vitro*-generated antibodies more frequently may require optimization in this respect. We again hypothesize that a preferred solution to the problem may be sought through diversification of the protein in accordance with the mutational profile of antibodies of the same germline gene origin. This may be feasible even when sensitive residues are found far from the antibody's paratope. For instance, methionine 101, situated far from the antibody's binding site, as encoded by germline gene *IGHV5–51*, is surface exposed and might be prone to oxidation. This particular residue is, however, frequently diversified in IgG derived from germline gene *IGHV5–51* (37, 38), in particular to isoleucine, threonine, or valine (**Figures 2C,G**). This fact provides reassurance that it is worthwhile to attempt a designed mutational strategy (incorporating commonly occurring residues in this particular position) to resolve stability problems associated with this particular residue. Other residues, far from the binding site, such as residue 71 of *IGHV1–18* (**Figure 2D**) [in close proximity to the structure pattern-defining residue at position 76 (59)], and residue 96 of *IGHV3–11* (**Figures 2B,F**) are prone to substitution *in vivo*. A range of such sites have been defined by others, in the form of gene-specific substitution profiles (GSSP) (37), and us (38) (**Figure 1**; Figure S1 in Supplementary Material). Interestingly, a past study suggested that substitutions far from the binding site indeed stabilized antibodies that also carried destabilizing mutations in their paratope (60). However, although the authors were confident about that main conclusion of this study, it was retracted based on a lack of confidence in a subset of confirmatory data (61). Formal demonstration of the usefulness of targeting of paratope-distant residues in antibody stabilization is thus not fully in place. Nevertheless, these natural paths of evolution, in our opinion, forms a natural basis for antibody perfection with maximum chance of tolerability and minimal risk of induction of anti-drug antibody responses following treatment as such antibody sequences would have been commonly seen by the immune system in the past. This will be the case as long as the treated subject, given the substantial diversity of the immunoglobulin loci (40, 62, 63), actually encodes antibodies derived from the germline gene/allele in question.

CONCLUSION

In summary, others and we have demonstrated substantial germline origin-unique substitution patterns in antibody V domains. Such patterns, even those beyond the classical CDRs, ought to be exploited in efforts aimed at affinity or specificity maturation, or optimization in terms of product developability. Future in-depth analysis of the outcome in terms of immunogenicity of antibodies perfected in this knowledge-driven manner will be required. However, we envisage that this procedure has true potential to establish antibodies with very limited immunogenicity as subjects treated with them are likely to have seen

highly similarly, molecularly evolved, antibodies in the past as part of their own immune responses.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Regionala etikprövningsnämnden (Lund) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Regionala etikprövningsnämnden (Lund).

AUTHOR CONTRIBUTIONS

HP: conceived the study and manuscript preparation. UK and LT: bioinformatics pipeline development, bioinformatics analysis, and manuscript preparation. FL: initial bioinformatics pipeline development and manuscript preparation. LG: patient management and manuscript preparation. MO: conceived the study, bioinformatics analysis, and main responsibility for manuscript preparation. All authors approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fimmu.2018.01391/full#supplementary-material.

REFERENCES

1. Kaplon H, Reichert JM. Antibodies to watch in 2018. *MAbs* (2018) 4:1–21. doi:10.1080/19420862.2018.1415671

2. Strohl WR. Current progress in innovative engineered antibodies. *Protein Cell* (2018) 9:86–120. doi:10.1007/s13238-017-0457-8

3. Söderlind E, Strandberg L, Jirholt P, Kobayashi N, Alexeiva V, Åberg AM, et al. Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat Biotechnol* (2000) 18:852–6. doi:10.1038/78458

4. Tiller T, Schuster I, Deppe D, Siegers K, Strohner R, Herrmann T, et al. A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. *MAbs* (2013) 5:445–70. doi:10.4161/mabs.24218

5. Bruggemann M, Osborn MJ, Ma B, Hayre J, Avis S, Lundstrom B, et al. Human antibody production in transgenic animals. *Arch Immunol Ther Exp (Warsz)* (2015) 63:101–8. doi:10.1007/s00005-014-0322-x

6. Ching KH, Collarini EJ, Abdiche YN, Bedinger D, Pedersen D, Izquierdo S, et al. Chickens with humanized immunoglobulin genes generate antibodies with high affinity and broad epitope coverage to conserved targets. *MAbs* (2018) 10:71–80. doi:10.1080/19420862.2017.1386825

7. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. Replacing the complementarity-cetermining regions in a human-antibody with those from a mouse. *Nature* (1986) 321:522–5. doi:10.1038/321522a0

8. Meijer PJ, Andersen PS, Haahr Hansen M, Steinaa L, Jensen A, Lantto J, et al. Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *J Mol Biol* (2006) 358:764–72. doi:10.1016/j.jmb.2006.02.040

9. Corti D, Lanzavecchia A. Efficient methods to isolate human monoclonal antibodies from memory B cells and plasma cells. *Microbiol Spectr* (2014) 2:AID–0018–2014. doi:10.1128/microbiolspec.AID-0018-2014

10. Smith SA, Crowe JE. Use of human hybridoma technology to isolate human monoclonal antibodies. *Microbiol Spectr* (2015) 3:AID–0027. doi:10.1128/microbiolspec.AID-0027-2014

11. Bostrom J, Yu SF, Kan D, Appleton BA, Lee CV, Billeci K, et al. Variants of the antibody herceptin that interact with HER2 and VEGF at the antigen binding site. *Science* (2009) 323:1610–4. doi:10.1126/science.1165480

12. Fournier P, Schirrmacher V. Bispecific antibodies and trispecific immunocy-tokines for targeting the immune system against cancer: preparing for the future. *BioDrugs* (2013) 27:35–53. doi:10.1007/s40259-012-0008-z

13. Liu YQ, Caffry I, Wu JM, Geng SB, Jain T, Sun TW, et al. High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. *Mabs* (2014) 6:483–92. doi:10.4161/mabs.27431

14. Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, et al. Biophysical properties of the clinical-stage antibody landscape. *Proc Natl Acad Sci U S A* (2017) 114:944–9. doi:10.1073/pnas.1616408114

15. Schaffitzel C, Hanes J, Jermutus L, Pluckthun A. Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. *J Immunol Methods* (1999) 231:119–35. doi:10.1016/S0022-1759(99)00149-0

16. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* (2014) 32:158–68. doi:10.1038/nbt.2782

17. Dekosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, et al. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* (2015) 21:86–91. doi:10.1038/nm.3743

18. Lavinder JJ, Horton AP, Georgiou G, Ippolito GC. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Curr Opin Chem Biol* (2015) 24:112–20. doi:10.1016/j.cbpa.2014.11.007

19. Bonsignori M, Zhou TQ, Sheng ZZ, Chen L, Gao F, Joyce MG, et al. Maturation pathway from germline to broad HIV-1 neutralizer of a CD4-mimic antibody. *Cell* (2016) 165:449–63. doi:10.1016/j.cell.2016.02.022

20. Lee J, Boutz DR, Chromikova V, Joyce MG, Vollmers C, Leung K, et al. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat Med* (2016) 22:1456–64. doi:10.1038/nm.4224

21. Kwong PD, Mascola JR, Nabel GJ. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat Rev Immunol* (2013) 13:693–701. doi:10.1038/nri3516

22. Burton DR, Hangartner L. Broadly neutralizing antibodies to HIV and their role in vaccine design. *Ann Rev Immunol* (2016) 34:635–59. doi:10.1146/annurev-immunol-041015-055515

23. Crowe JE Jr. Principles of broad and potent antiviral human antibodies: insights for vaccine design. *Cell Host Microbe* (2017) 22:193–206. doi:10.1016/j.chom.2017.07.013

24. Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for

antibody complementarity. *J Exp Med* (1970) 132:211–50. doi:10.1084/jem.132.2.211

25. Kirkham PM, Schroeder HW Jr. Antibody structure and the evolution of immunoglobulin V gene segments. *Semin Immunol* (1994) 6:347–60. doi:10.1006/smim.1994.1045

26. Chang B, Casali P. The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement. *Immunol Today* (1994) 15:367–73. doi:10.1016/0167-5699(94)90175-9

27. Dorner T, Foster SJ, Farner NL, Lipsky PE. Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. *Eur J Immunol* (1998) 28:3384–96. doi:10.1002/(SICI)1521-4141(199810)28:10<3384::AID-IMMU3384>3.0.CO;2-T

28. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G: C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol* (2004) 172:3382–4. doi:10.4049/jimmunol.172.6.3382

29. Saini J, Hershberg U. B cell variable genes have evolved their codon usage to focus the targeted patterns of somatic mutation on the complementarity determining regions. *Mol Immunol* (2015) 65:157–67. doi:10.1016/j.molimm.2015.01.001

30. Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr. Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *EMBO J* (1992) 11:603–9.

31. Kunik V, Peters B, Ofran Y. Structural consensus among antibodies defines the antigen binding site. *PLoS Comput Biol* (2012) 8:e1002388. doi:10.1371/journal.pcbi.1002388

32. Daugherty PS, Chen G, Iverson BL, Georgiou G. Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc Natl Acad Sci U S A* (2000) 97:2029–34. doi:10.1073/pnas.030527597

33. Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A* (2000) 97:10701–5. doi:10.1073/pnas.170297297

34. Burkovitz A, Sela-Culang I, Ofran Y. Large-scale analysis of somatic hypermutations in antibodies reveals which structural regions, positions and amino acids are modified to improve affinity. *FEBS J* (2014) 281:306–19. doi:10.1111/febs.12597

35. Ramirez-Benitez MC, Almagro JC. Analysis of antibodies of known structure suggests a lack of correspondence between the residues in contact with the antigen and those modified by somatic hypermutation. *Proteins* (2001) 45:199–206. doi:10.1002/prot.1140

36. Klein F, Diskin R, Scheid JF, Gaebler C, Mouquet H, Georgiev IS, et al. Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* (2013) 153:126–38. doi:10.1016/j.cell.2013.03.018

37. Sheng ZZ, Schramm CA, Kong R, Mullikin JC, Mascola JR, Kwong PD, et al. Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Front Immunol* (2017) 8:537. doi:10.3389/fimmu.2017.00537

38. Kirik U, Persson H, Levander F, Greiff L, Ohlin M. Antibody heavy chain variable domains of different germline gene origins diversify through different paths. *Front Immunol* (2017) 8:1433. doi:10.3389/fimmu.2017.01433

39. Levin M, Levander F, Palmason R, Greiff L, Ohlin M. Antibody-encoding repertoires of bone marrow and peripheral blood-a focus on IgE. *J Allergy Clin Immunol* (2017) 139:1026–30. doi:10.1016/j.jaci.2016.06.040

40. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* (2010) 184:6986–92. doi:10.4049/jimmunol.1000445

41. Lefranc MP. IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb Protoc* (2011) 2011:633–42. doi:10.1101/pdb.ip85

42. Bond CJ, Wiesmann C, Marsters JC, Sidhu SS. A structure-based database of antibody variable domain diversity. *J Mol Biol* (2005) 348:699–709. doi:10.1016/j.jmb.2005.02.063

43. Dreyfus C, Laursen NS, Kwaks T, Zuijdgeest D, Khayat R, Ekiert DC, et al. Highly conserved protective epitopes on influenza B viruses. *Science* (2012) 337:1343–8. doi:10.1126/science.1222908

44. Goossens T, Klein U, Kuppers R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc Natl Acad Sci U S A* (1998) 95:2463–8. doi:10.1073/pnas.95.5.2463

45. Ohlin M, Borrebaeck CA. Insertions and deletions in hypervariable loops of antibody heavy chains contribute to molecular diversity. *Mol Immunol* (1998) 35:233–8. doi:10.1016/S0161-5890(98)00030-3

46. Wilson PC, De Bouteiller O, Liu YJ, Potter K, Banchereau J, Capra JD, et al. Somatic hypermutation introduces insertions and deletions into immunoglobulin genes. *J Exp Med* (1998) 187:59–70. doi:10.1084/jem.187.1.59

47. Kepler TB, Liao HX, Alam SM, Bhaskarabhatla R, Zhang R, Yandava C, et al. Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe* (2014) 16:304–13. doi:10.1016/j.chom.2014.08.006

48. Ewert S, Honegger A, Pluckthun A. Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. *Methods* (2004) 34:184–99. doi:10.1016/j.ymeth.2004.04.007

49. Tramontano A, Chothia C, Lesk AM. Framework residue-71 is a major determinant of the position and conformation of the 2nd hypervariable region in the Vh domains of immunoglobulins. *J Mol Biol* (1990) 215:175–82. doi:10.1016/S0022-2836(05)80102-0

50. Krauss J, Arndt MA, Zhu Z, Newton DL, Vu BK, Choudhry V, et al. Impact of antibody framework residue VH-71 on the stability of a humanised anti-MUC1 scFv and derived immunoenzyme. *Br J Cancer* (2004) 90:1863–70. doi:10.1038/sj.bjc.6601759

51. Foote J, Winter G. Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol* (1992) 224:487–99. doi:10.1016/0022-2836(92)91010-M

52. Li T, Tracka MB, Uddin S, Casas-Finet J, Jacobs DJ, Livesay DR. Rigidity emerges during antibody evolution in three distinct antibody systems: evidence from QSFR analysis of Fab fragments. *PLoS Comput Biol* (2015) 11:e1004327. doi:10.1371/journal.pcbi.1004327

53. Thörnqvist L, Ohlin M. Data on the nucleotide composition of the first codons encoding the complementary determining region 3 (CDR3) in immunoglobulin heavy chains. *Data Brief* (2018) 19:337–52. doi:10.1016/j.dib.2018.04.125

54. Thörnqvist L, Ohlin M. The functional 3'-end of immunoglobulin heavy chain variable (IGHV) genes. *Mol Immunol* (2018) 96:61–8. doi:10.1016/j.molimm.2018.02.013

55. Lowe D, Dudgeon K, Rouet R, Schofield P, Jermutus L, Christ D. Aggregation, stability, and formulation of human antibody therapeutics. *Adv Protein Chem Struct Biol* (2011) 84:41–61. doi:10.1016/B978-0-12-386483-3.00004-5

56. Seeliger D, Schulz P, Litzenburger T, Spitz J, Hoerer S, Blech M, et al. Boosting antibody developability through rational sequence optimization. *MAbs* (2015) 7:505–15. doi:10.1080/19420862.2015.1017695

57. Casson LP, Manser T. Random mutagenesis of two complementarity determining region amino acids yields an unexpectedly high frequency of antibodies with increased affinity for both cognate antigen and autoantigen. *J Exp Med* (1995) 182:743–50. doi:10.1084/jem.182.3.743

58. Spencer S, Bethea D, Raju TS, Giles-Komar J, Feng Y. Solubility evaluation of murine hybridoma antibodies. *MAbs* (2012) 4:319–25. doi:10.4161/mabs.19869

59. Saul FA, Poljak RJ. Structural patterns at residue positions 9, 18, 67 and 82 in the VH framework regions of human and murine immunoglobulins. *J Mol Biol* (1993) 230:15–20. doi:10.1006/jmbi.1993.1121

60. Wang F, Sen S, Zhang Y, Ahmad I, Zhu X, Wilson IA, et al. Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. *Proc Natl Acad Sci U S A* (2013) 110:4261–6. doi:10.1073/pnas.1301810110

61. Wang F, Zhang Y, Ahmad I, Zhu X, Wilson IA, Smider VV, et al. Retraction for Wang et al., Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. *Proc Natl Acad Sci U S A* (2017) 114:E7855. doi:10.1073/pnas.1712805114

62. Kidd MJ, Chen Z, Wang Y, Jackson KJ, Zhang L, Boyd SD, et al. The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *J Immunol* (2012) 188:1333–40. doi:10.4049/jimmunol.1102097

63. Watson CT, Breden F. The immunoglobulin heavy chain locus: genetic varia-
    tion, missing data, and implications for human disease. *Genes Immun* (2012)
    13:363–73. doi:10.1038/gene.2012.12

Conflict of Interest Statement: The authors declare that the research was con-
ducted in the absence of any commercial or financial relationships that could be
construed as a potential conflict of interest.