

Connecting the sequence dots: shedding light on the genesis of antibodies reported to be designed in silico

Maximiliano Vásquez, Eric Krauland, Laura Walker, Dane Wittrup & Tillman Gerngross

To cite this article: Maximiliano Vásquez, Eric Krauland, Laura Walker, Dane Wittrup & Tillman Gerngross (2019) Connecting the sequence dots: shedding light on the genesis of antibodies reported to be designed in silico, mAbs, 11:5, 803-808, DOI: [10.1080/19420862.2019.1611172](https://doi.org/10.1080/19420862.2019.1611172)

To link to this article: <https://doi.org/10.1080/19420862.2019.1611172>



© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.



View supplementary material [↗](#)



Published online: 20 May 2019.



Submit your article to this journal [↗](#)




Article views: 12993



View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at

Connecting the sequence dots: shedding light on the genesis of antibodies reported to be designed *in silico*

Maximiliano Vásquez ^a, Eric Krauland ^a, Laura Walker^a, Dane Wittrup^{a,b}, and Tillman Gerngross^{a,c}

^aAdimab LLC, Lebanon, NH, USA; ^bMassachusetts Institute of Technology, Cambridge, MA, USA; ^cThayer School of Engineering, Dartmouth College, Hanover, NH, USA

ABSTRACT

Two recent publications out of the same research laboratory report on structure-based *in silico* design of antibodies against viral targets without sequence disclosure. Cross-referencing the published data to patent databases, we established the sequence identity of said computationally designed antibodies. In both cases, the antibodies align with high sequence identity to previously reported antibodies of the same specificity. This clear underlying sequence relationship, which is far closer than the antibody templates reported to seed the computational design, suggests an alternative origin of the computationally designed antibodies. The lack of both reproducible computational algorithms and of output sequences in the initial publications obscures the relationship to previously reported antibodies, and sows doubt as to the genesis narrative described therein.

ARTICLE HISTORY

Received 16 April 2019
Accepted 18 April 2019

KEYWORDS

Structure-based design; *in silico* antibody; peer review; sequence disclosure; virus neutralization

Introduction

Human antibodies are a major modality to treat human disease, and therefore the focus of significant technology development.¹ Historically, two main approaches have been used to discover antibodies against targets of therapeutic interest: 1) *in vivo* technologies based on isolating B cell diversity from animals^{2–5} or humans,^{6–8} following immunization or exposure to infectious agents, respectively, and 2) *in vitro* technologies based on the display of synthetic or semi-synthetic human IgG diversity on the surface of phage or yeast.^{9–12}

More recently, a third approach of designing human antibodies *in silico* against specific epitopes has been fueled by two major trends: 1) ever-increasing structural information of antibodies and their potential targets, as well as 2) access to more powerful computational tools. Notwithstanding the potential impact, the general lack of published successes in this area has highlighted the extreme challenge of designing antibodies *in silico*,^{13,14} although progress in the area of affinity maturation has been made.^{15,16}

As such, two publications^{17,18} originating from the same research lab (based on the first and corresponding author), reporting the *in silico* design of epitope-specific, broadly neutralizing, human antibodies against two infectious disease targets, garnered our attention. Strikingly, in both cases the extraordinary accomplishments were not supported by a detailed description of methods or intermediate results, nor were the end-products of these efforts, namely the amino acid sequences of the designed antibodies, disclosed, making it impossible to independently reproduce the reported functional characterization. To understand how these results could have been achieved, we endeavored to better

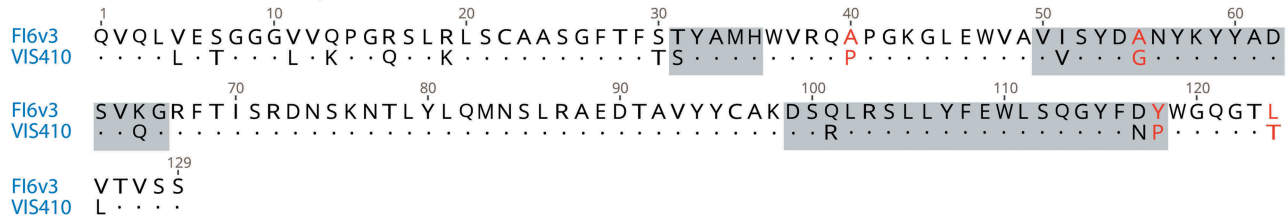
understand the identity and, potentially, the genesis of these antibodies. In this communication, we present evidence that in both cases, previously published antibody sequences and structures are the basis for the *in silico* designed antibodies.

Results

Influenza antibody

VIS410 is described¹⁷ as a broadly neutralizing antibody generated by a process that used “a database of nonredundant combinations of complementary determining region (CDR) canonical structures (antibody scaffolds), (to select) multiple antibody templates satisfying shape complementarity criteria and systematically engineered energetically favorable, hotspot-like interactions between CDR residues and these anchor residues on hemagglutinin (HA).” The authors then present experimental data on binding, neutralization, and protective efficacy in influenza animal models. However, the sequence of VIS410, the output of the design and engineering process, was neither provided in the publication, nor deposited into a public database. Using VIS410 as a search term in the USPTO database readily leads to a US patent application¹⁹ that also designates VIS410 as Ab044. This application further establishes that the variable heavy- and light-chain (VH and VL) sequences correspond to sequence ID numbers 25 and 52, which are shown in Figure 1. Searching the patent database with the VIS410 sequences produces exact matches to an earlier US patent publication from 2013 describing Ab044 with a similar inventorship group.²⁰ As shown in Figure S1, a comparison of tables from the two sources,^{17,20} confirms that VIS410 and Ab044 are in fact the same antibody.

Heavy Chain: 87% identity | 96% similarity



Light Chain: 81% identity | 88% similarity

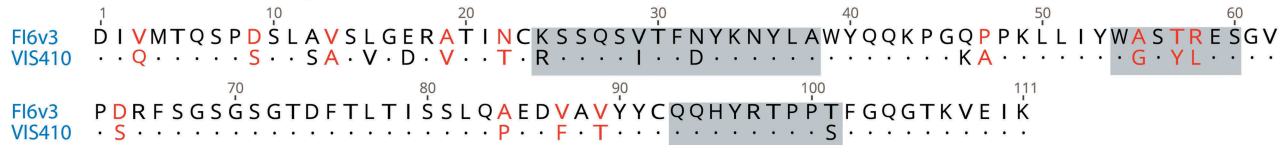


Figure 1. VH and VL alignment of Fl6v3 (PDB file 3ZTJ) and VIS410 (Ab 044) Fl6. Non-conservative substitutions depicted in red font. CDRs are highlighted in gray according to Kabat’s²³ definition.

It is interesting that searching Genbank with the VH sequence, even today (April 2019, nearly four years after the original publication) does not yield a 100% match. However, the search does reveal FI6v3, a broadly neutralizing anti-influenza antibody first described in 2011 by Corti and et al.²¹ An alignment of VIS410 and FI6v3 is shown in Figure 1. The overall percent identity values are 87% and 81%, respectively, for the VH and VL domains. This is achieved with no gaps in the alignment, indicating that all the corresponding CDR lengths are identical; this result is particularly significant for the CDRH3 and for the CDRL1, with the latter showing a relatively rare two amino acid deletion relative to the human germline of origin, as presented in the original publication by Corti et al.²¹ Including conservative substitutions (“positives” as in the default settings for BLAST²²) reveals an even closer relationship with a similarity of 96% and 88%, respectively, for VH and VL. The VIS410 publication includes in supplementary material a list of accession numbers to antibody variable regions as “top ranking templates” used for the design of anti-influenza antibodies.¹⁷ In Figure 2 we present the corresponding VH sequences, retrieved from the NCBI database using those accession numbers, and aligned with both VIS410 and FI6v3. Focusing on the CDRs, the portion of the antibody sequence expected to be most important to determine specificity and govern binding to antigen,^{23–25} the closest template CDRH3 to the output VIS410 CDRH3 is 10% identical, while the FI6v3 CDRH3 to VIS410 CDRH3 is 85% identical. Given the remarkable diversity of CDRH3 sequences, it is hard to conceive how a computational method could have credibly and independently converged near the FI6v3 sequence. Further comparison of the other CDRs

between VIS410 and FI6v3, shows just two non-conservative substitutions in VH CDRs, and three in CDRL2. According to the crystal structure, CDRL2 was deemed non-essential for antigen binding by FI6v3 and therefore an attractive location to introduce sequence alterations that are unlikely to compromise binding.²¹ Based on the structure of FI6v3 in complex with influenza HA, we mapped those differences, as shown in Figure 3. Only one of the non-conservatively substituted amino acid positions appears to be directly involved in antigen contact; this would be position 54 of the VH, which is Ala in FI6v3, but Gly in VIS410. As detailed in the original publication,²¹ the precursor of FI6v3, called FI6, does have Gly at the same position, and this change is among a few others demonstrated to have no impact on functionality. Given the remarkable degree of similarity, it is important to remember that FI6v3, including its sequence and structure, was described almost a year before the submission date of the first patent application describing the discovery of VIS410 (Figure S2)

Zika antibody

More recently, the same research group with new collaborators reported to have “applied computational methods to engineer an antibody, ZAb_FLEP,” with broadly neutralizing activity against Zika virus.¹⁸ As in the earlier¹⁷ publication, no sequence information was provided for ZAb_FLEP. Following the same approach as in the previous section, we initially searched the patent literature with the term “Zab_FLEP,” which proved unproductive. However, searching Zika together with some of the author’s names in a patent

Name (Accession)	Sequence	Differences in CDRs
218681906	QVQLVESGGGVVQPGRSLRLSCAASGFTFSSYGMHWVRQAPGKGLEWVAIVSYDGSNKYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCAKDGKCGGSGCYSGL----LDYHGQGT LVTVSS	24
33318928	EVQLVESGGGLVQPGGSLRLSCAASGFTFSKFWMSWVRQAPGKGLEWVAIVSYDGSNKKYADSVKGRFTISRDAAKNSVYLQVYSLRAEDTAIYYCARLSDNCGVGGNCYFSQITWFDSWGQGT LVTVST	32
185486	QVQLVESGGGVVQPGRSLRLSCAASGFTFSYGMHWVRQCPGKGLEWVAIVSYDGSNKYYADSVKGRFTISRDNKNTLYLQMDSLRTEDTAVYYCAKGVYCSSSSCYSYYHYHMDVWGKGT LVTVSS	26
VIS410	QVQLLETGGGLVKPQSLKLSCAASGFTFSTYAMHWVRQAPGKGLEWVAIVSYDGSNKYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCAKDSRLRSLLYFEWLSQGYFDPYWGQGT LVTVSS	NA
Fl6v3	QVQLVESGGGVVQPGRSLRLSCAASGFTFSTYAMHWVRQAPGKGLEWVAIVSYDANVYKYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCAKDSQLRSLLYFEWLSQGYFDYWGQGT LVTVSS	8

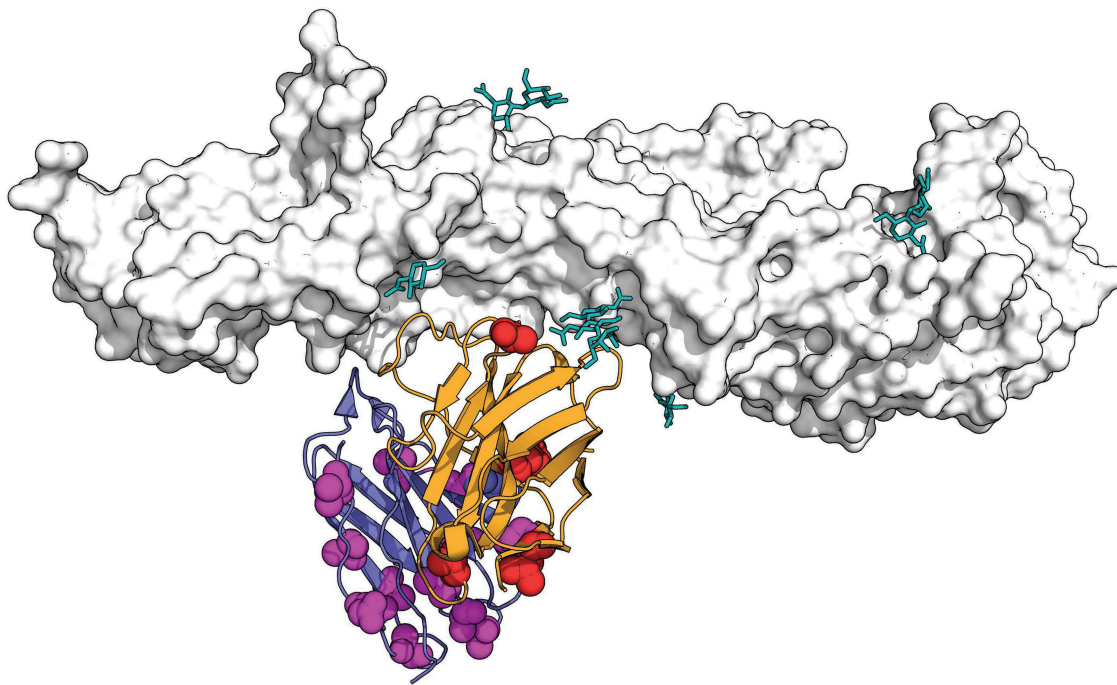


Figure 3. Depiction of non-conservative substitutions (red for the VH and magenta for the VL) in the context of the structure of F16v3 complexed with influenza H1 (gray, glycans shown as teal sticks). VH and VL are shown in orange and purple, respectively. PDB file 3ZTN.

database led us to a published patent application²⁶ titled “Antibodies that bind Zika virus envelope protein and uses thereof”. All six named inventors are also authors of the publication describing ZAb_FLEP.¹⁸

Comparison of figures in the publication and patent application leaves little doubt that ZAb_FLEP corresponds to mAb 8; see Figure S3.^{18,26} The sequences of mAb 8 are presented in the patent application²⁶ (sequence IDs 6 and 15), and are shown in Figure 4.

As in the influenza case, a Genbank search with mAb 8 sequences fails to retrieve an exact match, but results in sequence hits to EDE1 C8,^{27–30} a previously reported dengue/Zika cross-reactive antibody first described in 2015 by Dejnirattisai et al.²⁹ An alignment of EDE1 C8 and mAb 8 is shown in Figure 4.

As in the influenza example, the similarity is remarkable, with 89% and 90% identity, for VH and VL, respectively. This is obtained without any gaps in the alignment, and thus all CDR lengths are identical between ZAb_FLEP and EDE1 C8. Considering conservative substitutions, as before, yields similarities of 95% and 98%, respectively, for VH and VL. Within the CDRs of the VH, there is a single V to A non-conservative substitution (as defined by BLAST default settings). There are only three conservative substitutions in VL CDRs; in fact, one in each CDR, all involving S/T exchanges. The non-conservative substitutions in the context of the known EDE1 C8 complex with Zika protein²⁷ are depicted in Figure 5. Once again, given the remarkable sequence similarity, it is important to emphasize the prior publication of the EDE1 C8 sequences, in the context of both dengue and Zika (Figure S2).^{27,30}

Heavy Chain: 89% identity | 95% similarity

	1	10	20	30	40	50	60
EDE1 C8	EVQLVESGGGLVQPGGSLRLSCSASGFTFSTYSMHWVRQAPGKGLYYSVAITGEGDSAIFYAD						
mAb8L.....A.....S.....W.....S.....Y.....						
	70	80	90	100	110	120	
EDE1 C8	SVKGRFTISRDNKNTLYFEMNSLRPEDTAVYYCVGGYSNFFYYYYTMDVWGQGT T V T V S S						
mAb8LQ.....KV.....A.....A.....M.....						

Light Chain: 90% identity | 98% similarity

	1	10	20	30	40	50	60
EDE1 C8	EIVLTQSPATLSLSPGERATLSCRASQSI STFLAWYQH K P G Q A P R L L I Y D A S T R A T G V P A R						
mAb8	D.....M.....S.....T.....Q.....S.....I.....						
	70	80	90	100	108		
EDE1 C8	FSGRSRSGTDFTLTLSITLPEPDFAVYYCQQRYNWPPYTFGQGTKVEIK						
mAb8TR.....S.....L.....						

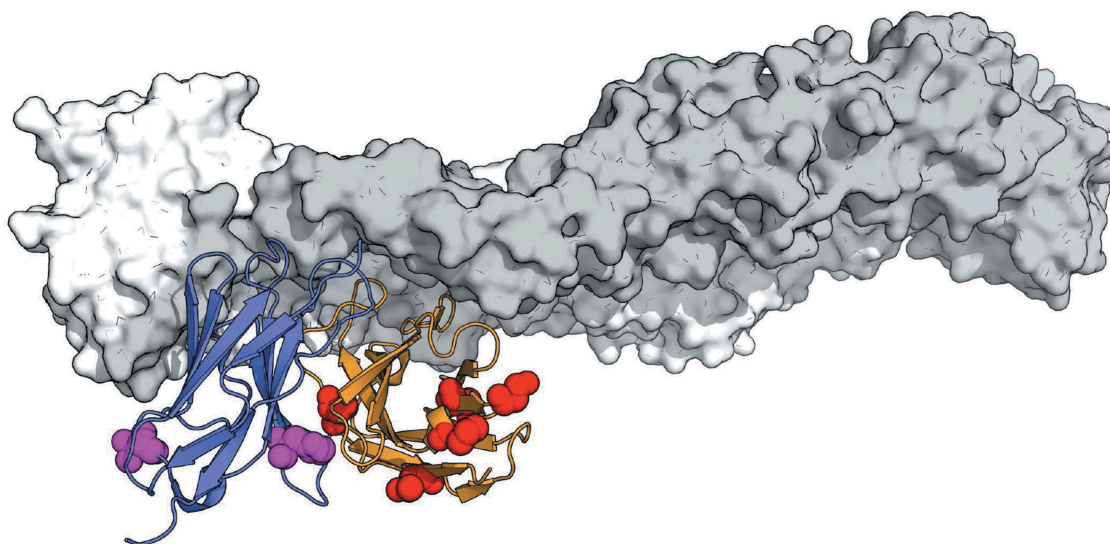


Figure 5. Depiction of non-conservative substitutions (red for the VH and magenta for the VL) in the context of the structure of EDE1 C8 complexed with ZIKV E (gray). VH and VL are shown in orange and purple, respectively. PDB 5LBS, chains AHL.

It is interesting that EDE1 C8 is mentioned, among other antibodies, as a *potential* template for the design of ZAb_FLEP, as indicated in the supplementary material:¹⁸ “Multiple antibody scaffolds (including mouse-derived pan-flavivirus 4G2, anti-EDE1 Dengue mAbs C8, C10 and anti-EDE2 Dengue mAb A11, anti-TDRD3 antibody and anti-HIV antibody PGT124) were used as starting templates for antibody engineering.”¹⁸ The lack of sequence disclosure for ZAb_FLEP and any direct data comparisons to EDE1 C8, however, obscures from readers, as well as peer reviewers, the remarkable similarity of ZAb_FLEP and EDE1 C8. Given the apparent origin of ZAb_FLEP from EDE1 C8, we wonder why a direct comparison between the two was not reported, especially in light of the authors’ statement:¹⁸ “The *in vitro* neutralization potential of ZAb_FLEP **approaches** the potency of select Zika antibodies” (emphasis added).

The narrative in the patent application,²⁶ which is intended to teach the skilled artisan how to practice the invention, only provides sequences from an anti-TDRD3 antibody as input template, and it makes no explicit mention of EDE1 C8 as input. Moreover, sequences identical to EDE1 C8, represented as mAb 3 in the patent document, are said to have been “designed by computing the epitope-paratope connectivity network,” whereby “variable regions and CDRs (are) generated (and) shown in (...) Figure 1(a,b)” (see Figure S4²⁶). However, the designed mAb 3 has a non-traditional two amino acid addition (Arg-Ser) at the VL N-terminus. This sequence matches a non-coding cloning site present in the original EDE1 C8 VL expression vector.²⁹ It is inconceivable that an unsupervised algorithm would produce vestigial vector sequence unrelated to antigen recognition.

Discussion

In this report, we examine two instances in which the same

neutralization breadth and potency.^{17,18} In neither case were the sequences of the designed antibodies disclosed, leading us to question the enforcement of editorial policies regarding reproducibility. Perhaps more concerning is the potential for contamination of the scientific literature with claims by innocent third parties. For example, in a commentary article³¹ about the clinical evaluation of VIS410,³² it is said that “VIS410, however, is not just another HA-stem specific human monoclonal antibody. This human IgG1 monoclonal antibody is the result of man-made design and protein engineering and so is not derived from a natural source.” Clearly, the author of this comment was not in possession of the comparison presented in Figure 1.

We present with a high degree of confidence the actual sequence identity of the designed antibodies, and a more plausible genesis narrative. Comparisons of these sequences to those of previously described human B cell-derived antibodies to the same targets show striking similarities. By contrast, those designed sequences appear very dissimilar from the templates said to have been used to start the design process (Figure 2 and S4); we leave it to the reader to judge the likelihood of these highly homologous sequences being re-discovered coincidentally, or simply derived from existing antibodies targeting the same epitopes as those of the computationally designed antibodies.

In conclusion, the presented fact pattern calls into question the publications’ claimed genesis of VIS410 and ZAb_FLEP. Furthermore, the lack of sequence disclosure exposes a serious weakness in the peer review process in the emerging field of computational antibody design.^{13–16} (It is instructive to compare the level of transparency provided by some¹⁶ to the opaque disclosures in the publications examined here.^{17,18}) Such obfuscation prevents independent confirmation, and is contrary to basic scientific norms. We find it difficult to view these authors’ approach^{17,18} in any light other than an intent

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.