Phage Display Systems for Antibody Engineering and Evaluation

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Health

at

The University of Waikato

by

Kyrin Rene Hanning

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2025

Abstract

Antibodies are versatile effector molecules of the adaptive immune system, distinguished by their capacity to recognise and engage diverse antigens with high specificity and affinity. The widespread adoption of antibodies, and their various formats, in diagnostics and therapeutics has been made possible through the co-operation of experimental and computational molecular engineering techniques. Functional screening of variant antibody libraries has proven essential for successful antibody development. Several protein display technologies have been deployed for this important screening phase, including phage display. Whilst one of the older approaches, phage display still offers many advantages over alternatives and synergises well with other advances in molecular biology techniques. This thesis demonstrates how phage display, when paired with modern techniques in novel ways, remains an extremely flexible and powerful tool for engineering and evaluating antibodies.

Acknowledgements

I am massively grateful to my chief supervisor, Dr. William Kelton, for their exceptional mentorship and unwavering support throughout my PhD journey. It's crazy to think I not only ended up in the medical research field but undertook it in Hamilton of all places. Sincere thanks also go to my secondary supervisors, Dr. Joanna Hicks and Prof. Vickery Arcus, for their support. A special thanks to Associate Professor Tina Summerfield at the University of Otago for first introducing me to the world of research all those years ago. I am also deeply appreciative of the University of Waikato Doctoral Scholarship that made my study here possible.

None of this would have been possible without the inspiration and support I received from C2 lab members, past and present; Emma, Stacy, Annmaree, Liz, Keely, Jolyn, Kevin, Patrick, Marina, Chloe, Greg, and Kirsten, to name a few. Thank you all so much. An especially enormous thanks to Dr. Judith Burrows, Dr. Geetanjali Rai, and Dr. Olivia Patty for all your work to keep the lab functional and sane.

Finally, I would like to acknowledge Mum, Sara, Cory, Mike, Dani, Seth, Harriet, Breanne, Theo, Serena, Matilda, and Elisa. To quote Dr. Hanning I: “you guys are a regular party.” I hugely appreciate all your support during this wild ride.

Table of Contents

Chapter One

Introduction

Bacteriophage (phage) display has made invaluable contributions to protein research for nearly forty years (Ledsgaard et al., 2022). The technique comprises of displaying a protein of interest as a phage coat fusion protein and packaging the encoding gene within the phage progeny, presenting an elegant and powerful approach to observing genotype-phenotype relationships. Advances in experimental and computational molecular biology methodologies are continually refining phage display as a powerful high-throughput strategy to screen protein variants for target interactions. The approach has long been utilised by antibody development pipelines, and hundreds of phage display-derived antibodies have been described, including over 14 clinically approved therapeutic antibodies (Alfaleh et al., 2020). While phage display is most often used in this capacity to identify antibody variants with improved antigen affinity, it is also a practical methodology to directly develop antibody specificity and affinity for immune receptors (Chen et al., 2021). Phage display is also a convenient tool to survey the permissible mutational space of antigens, providing enhanced targets for antibody candidate screening (Dingens et al., 2017; Harvey et al., 2021; Starr et al., 2020). Deep sequencing technologies are central to the use of phage display in modern antibody engineering applications, as the capture of immense quantities of variant sequences improves the scope and precision of variant screening analyses. Numerous library design strategies have been developed to better capitalise from the available sequencing depth. This includes specifying the target amino acids for mutagenesis, the composition of those substitutions, and controlling the number of simultaneous mutations for each variant within the library. As a result, a diverse range of approaches to phage display variant library construction have been developed, ranging from random mutagenesis to fully synthetically synthesised libraries with precise mutations (Kille et al., 2013; Li et al., 2018). Despite the numerous options for library construction, there is a lack of site-specific approaches to library assembly that facilitate simultaneous mutations across large spans of sequence whilst remaining cost-effective. The importance of realising library design has been further emphasised by recent advances with machine learning-based analyses. Such techniques are enabling the analysis of more complex library compositions and extrapolating from variant sequence space screened by protein display (Taft et al., 2022).

This thesis aims to utilise high throughput phage display with new tools such as next generation sequencing and machine learning to facilitate antibody engineering and discovery. More specifically, the work integrates Oxford Nanopore sequencing with phage-based epitope mapping, demonstrates a novel library construction approach for phage displayed nanobody affinity maturation, and further applies this method to create libraries for complex multi-receptor engineering challenges within the antibody Fc domain.

This thesis comprises as in-depth literature review followed by three chapters addressing each of the research objectives and concludes with a discussion of future perspectives. The literature review chapter is adapted from a peer-reviewed Review Article published in *Trends in Pharmacological Sciences*:

Hanning, K. R., Minot, M., Warrender, A. K., Kelton, W., and Reddy, S. T. (2022). Deep mutational scanning for therapeutic antibody engineering. *Trends in Pharmacological Sciences,* 43 (2). https://doi.org/10.1016/j.tips.2021.11.010

The subsequent research-focused chapters are presented as a combination of published manuscripts, submitted manuscripts, or manuscripts ready for submission, as outlined in the Research Objectives section.

Research Objectives

To blend new tools in molecular biology with phage display for antibody engineering and characterisation this thesis explores the following objectives

Objective one:

Explore the viability of integrating nanopore sequencing with phage display for the purposes of linear epitope mapping, utilising a vaccine candidate example.

The approach and example are described in the following publication and is presented in Chapter Three as a final peer-reviewed manuscript:

Browne-Cole, K.*, **Hanning, K.*,** Beijerling, K., Rousseau, M., Loh, J. and Kelton, W. (2025). A rapid approach to for linear epitope vaccine profiling reveals unexpected epitope tag immunogencity. *Scientific Reports,* 15 (1). https://doi.org/ 10.1038/s41598-025-92928-3

*Denotes shared first authorship. Individual contributions are outlined at the start of Chapter Three.

Objective two:

Develop a highly modular, site-specific yet cost-effective approach to multiplex variant library synthesis for the purposes of deep mutational scanning and learning of antibody sequences. The research on this objective is presented as a complete manuscript that is ready for submission.

Hanning, K.R., Walker, E.J., Beijerling, K., Irvine., E.B., Steel, J.J., and Kelton, W. (2025). Simple high-throughput encoding of deep mutational scanning libraries by oligo-based golden gate assembly. *bioRxiv,* https://doi.org/10.1101/2025.07.16.665225

Objective three:

Use multiplex library assembly methods to create libraries of IgA Fc domains for future screening against multiple overlapping Fc receptors to tune affinity profiles. The research on this objective is presented in the future directions section.

Acknowledgement of Funds

This research was funded by a Waikato Medical Research Fund (grant; #320) and a Maurice Wilkins Centre (grant: #4026). Financial support for the duration of my PhD was provided by a University of Waikato Doctoral Scholarship.

References

Alfaleh, M. A., Alsaab, H. O., Mahmoud, A. B., Alkayyal, A. A., Jones, M. L., Mahler, S. M., & Hashem, A. M. (2020). Phage display derived monoclonal antibodies: From bench to bedside. *Frontiers in Immunology*, *11*, 1986. https://doi.org/10.3389/fimmu.2020.01986

Chen, D., Zhao, Y., Li, M., Shang, H., Li, N., Li, F., Wang, W., Wang, Y., Jin, R., Liu, S., Li, X., Gao, S., Tian, Y., Li, R., Li, H., Zhang, Y., Du, M., Cao, Y., Zhang, Y., … Zhang, H. (2021). A general Fc engineering platform for the next generation of antibody therapeutics. *Theranostics*, *11*(4), 1901–1917. https://doi.org/10.7150/thno.51299

Dingens, A. S., Haddox, H. K., Overbaugh, J., & Bloom, J. D. (2017). Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. *Cell Host & Microbe*, *21*(6), 777-787.e4. https://doi.org/10.1016/j.chom.2017.05.003

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*(7), 409–424. https://doi.org/10.1038/s41579-021-00573-0

Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z.-G., Opperman, D. J., Reetz, M. T., & Acevedo, J. P. (2013). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synthetic Biology*, *2*(2), 83–92. https://doi.org/10.1021/sb300037w

Ledsgaard, L., Ljungars, A., Rimbault, C., Sørensen, C. V., Tulika, T., Wade, J., Wouters, Y., McCafferty, J., & Laustsen, A. H. (2022). Advances in antibody phage display technology. *Drug Discovery Today*, *27*(8), 2151–2169. https://doi.org/10.1016/j.drudis.2022.05.002

Li, A., Sun, Z., & Reetz, M. T. (2018). Solid-phase gene synthesis for mutant library construction: The future of directed evolution? *Chembiochem: A European Journal of Chemical Biology*, *19*(19), 2023–2032. https://doi.org/10.1002/cbic.201800339

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J. D. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, *182*(5), 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012

Taft, J. M., Weber, C. R., Gao, B., Ehling, R. A., Han, J., Frei, L., Metcalfe, S. W., Overath, M. D., Yermanos, A., Kelton, W., & Reddy, S. T. (2022). Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain. *Cell*, *185*(21), 4008-4022.e14. https://doi.org/10.1016/j.cell.2022.08.024

Chapter Two
Literature Review

Preface

A compilation of literature demonstrating deep mutational scanning (DMS) as a methodology to engineer therapeutically relevant antibodies was published as a peer-reviewed Review Article in *Trends in Pharmacological Sciences*. The article discusses the techniques and technologies that comprise DMS, a mutagenesis-based approach that is well-suited for characterising residues associated with protein interplays. Monoclonal antibody (mAb) development can benefit from DMS workflows in several ways. Examples are provided demonstrating the use of DMS to generate mAbs with higher affinity, specificity, and/or stability (i.e., synthetic affinity maturation), and by utilising DMS of target antigens to provide insights into the real-world viability of therapeutic candidates via simulated epitope escape and mapping. The pairing of DMS data with machine learning analyses has benefited *in silico* protein design, DMS experimental design and analyses; enabling the rapid generation and characterisation of highly diverse mAb candidates.

The format of the published article has been modified for this thesis to allow for the inclusion of additional text to provide the most up-to-date review of the literature. Text originally formatted as boxes are represented here as additional figures. All text added after the initial publication is presented in blue to differentiate it from the already published text.

Hanning, K. R., Minot, M., Warrender, A. K., Kelton, W., and Reddy, S. T. (2022). Deep mutational scanning for therapeutic antibody engineering. *Trends in Pharmacological Sciences,* 43 (2). https://doi.org/10.1016/j.tips.2021.11.010

Author contributions

K.R.H. and W.K. compiled and wrote this manuscript. M.M. and A.K.W. contributed to sections relating their expertise: Machine learning and antibody stability, respectively. Responsibility for the overall scope was shared between W.K and S.T.R.

Deep mutational scanning for therapeutic antibody engineering

Kyrin R. Hanning[a], Mason Minot[b], Annmaree K. Warrender[a], William Kelton[a], and Sai T. Reddy[b]

[a]Te Huataki Waiora School of Health, University of Waikato, Hamilton, New Zealand
[b]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH) Zurich, Basel, Switzerland

Abstract

The biophysical and functional properties of monoclonal antibody (mAb) drug candidates are often improved by protein engineering methods to increase the probability of clinical efficacy. One emerging method is deep mutational scanning (DMS) which combines the power of exhaustive protein mutagenesis and functional screening with deep sequencing and bioinformatics. The application of DMS has yielded significant improvements to affinity, specificity, and stability for several preclinical antibodies alongside novel applications such as introducing multi-specific binding properties. DMS has also been applied directly on target antigens to precisely map antibody-binding epitopes and notably to profile the mutational escape potential of viral targets (e.g., SARS-CoV-2 variants). Finally, DMS combined with machine learning is enabling advances in computational screening and engineering of therapeutic antibodies.

Protein engineering with deep mutational scanning

Antibodies have become the most successful class of biologic drugs for treating indications ranging from autoimmune and neurodegenerative diseases to cancer and infectious disease. The unmistakable influence of these molecules on modern medicine is highlighted by the recent milestone approval of the 100th **monoclonal antibody** (**mAb**; see Glossary; Box 1) drug by the FDA (Kaplon & Reichert, 2021; Mullard, 2021). Despite their success, there are still many unmet medical needs requiring the development of next-generation antibodies with higher affinity, greater specificity, improved stability, and enhanced immune functionality (Fig. 2.1) (Carter & Lazar, 2018). New experimental and computational protein engineering approaches have emerged to enable the development of such enhanced mAb candidates (Fowler et al., 2010; Norman et al., 2020; Schoeder et al., 2021). One notable example is **deep mutational scanning (DMS)**, which represents an integrated experimental-computational method for comprehensively mapping protein sequence to function in high-throughput (Fig. 2.2).

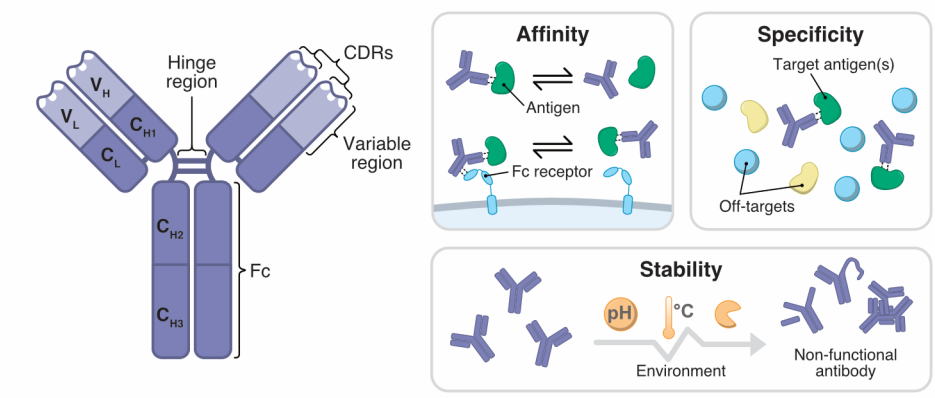


Figure 2.1. Antibodies are created by the tetrameric assembly of four polypeptide chains. Antigen binding is governed by three complementarity determining regions (CDRs) within each of the variable light (VL) and heavy (VH) domains. A flexible hinge region links these domains to the fragment crystallisable (Fc) region that is responsible for receptor-mediated immune responses. Antibody engineering has focused on every part of the molecule to enhance its biophysical properties. Modifying the CDRs can improve the binding affinity or specificity of an antibody for a cognate antigen, and changes to the Fc domain may be immunomodulatory. Antibody stability is influenced by mutations in all domains (VL, VH, CL, CH1, CH2, CH3 for IgG1) as well as in the hinge region. Ideal therapeutic antibodies possess high affinities for specific antigen targets, elicit an appropriate immune response, and have high stability.

Central to DMS is the generation of DNA mutagenesis libraries that are designed to encode amino acid mutations across selected regions of proteins, for example the

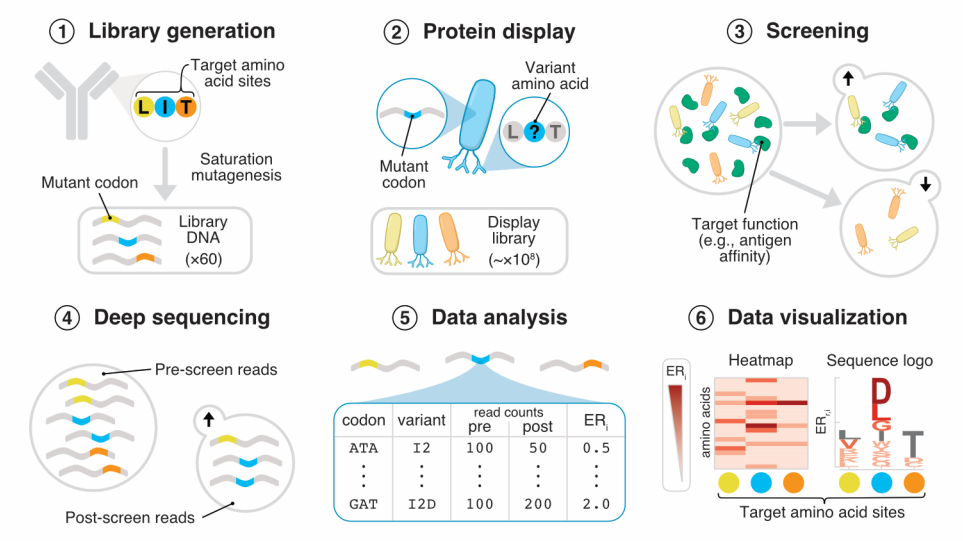


Figure 2.1. A typical deep mutational scanning (DMS) workflow for monoclonal antibody (mAb) engineering. Amino acid sites (yellow, blue, orange) within a targeted region of a protein are selected for mutagenesis. ① The DNA codons of these target sites are systematically mutated such that each position contains codons from all 20 amino acids. ② Variant sequences are pooled into a library (library DNA) and expressed using a protein display system (e.g., phage display; Box 3). ③ The displayed library is then screened for a desired function (e.g., improved antigen affinity) via an appropriate selection method (e.g., phage panning) that enables the collection of screened variants. ④ Deep sequencing of the display library pre- and post-screening is paired with ⑤ bioinformatic analysis to determine changes in the frequency of variant codons, where enriched amino acid substitutions positively contribute to function. ⑥ In single-site DMS experiments, calculated enrichment ratios are often visualized with heatmaps or sequence logos. The information generated by DMS identities sites within the protein that are important for function as well as the amino acid residues that are best-tolerated at those sites. These data can directly feed into the further development of other biophysical properties or inform computational analyses, such as machine learning, to restrict the sequence space when generating multi-site or combinatorial libraries for additional screening. Abbreviation: $ER_{r,i}$, amino acid enrichment ratio.

complementarity determining regions (CDRs) of antibodies. Such libraries are often based on site saturation mutagenesis in which where each position (codon) of the selected region is mutated to encode each of the 20 possible amino acids. Library design typically involves single-site tiling: the systematic targeting of single positions while all other positions are maintained as the original sequence, thus resulting in a pool of variants that all differ from the original sequence by a single amino acid mutation. However, multi-site libraries where several positions simultaneously undergo saturation mutagenesis are becoming increasingly accessible (Box 2). Following library screening and the selection of variants (e.g., for antigen binding), deep sequencing is performed to determine variant read counts in the pre- and post-selection populations. Shifts in observed mutation frequencies are then correlated with positive or negative response to the selection pressure applied. Despite only sampling a small fraction of

possible protein sequence space, DMS provides valuable information on key positions and residues essential for a desired function. This information can be used to guide the design of more complex, combinatorial mutagenesis libraries, which can then be screened to identify variants with enhanced properties. Functional screening of mAb libraries by DMS requires that the protein sequence (phenotype) is linked to the underlying genetic information that codes for the variant (genotype); this is often achieved by using a protein display system (Box 3). Combination of this genotype-to-phenotype linkage with deep sequencing is a critical aspect of DMS because it enables the massively parallel screening of variant libraries (sometimes exceeding 1×10^8 variants). Before DMS, site saturation mutagenesis approaches frequently required variant partitioning into microplate wells for mAb engineering (Burks et al., 1997; G. Chen et al., 1999). We review here modern DMS workflows for mAb engineering that combine advances in antibody display platforms, oligonucleotide synthesis, deep sequencing, and, more recently, the application of machine-learning algorithms. The resulting systems enable the rapid generation and screening of mAb candidate libraries with unprecedented efficiency.

Box 1. Glossary

Complementarity-determining regions (CDRs): flexible loop regions of diversity in mAb variable domains that are largely responsible for antigen binding. Both the variable region heavy (VH) and light (VL) chains contain three CDR loops, and most diversity is concentrated in CDRH3/CDRL3.

Deep learning: a subfield of machine learning with a focus on mimicking human brain activity by training artificial neural networks. Layers of nodes are used in place of neurons, and threshold and weighting values control the transfer of data between the individual nodes. As the model sees more data, these values adjust to improve accuracy.

Deep mutational scanning (DMS): a high-throughput protein engineering technique for detailed sequence-function analysis. Protein residues are systematically mutated within a sequence to create a tiled library in which each position targeted is saturated with each of the possible 20 amino acids. Deep sequencing is used to track variant enrichment when library screening is performed for a desired property (e.g., antibody binding to an antigen). Libraries for deep mutational scanning are commonly single-site in design (i.e., only one residue is targeted for saturation mutagenesis at a time), although saturation of several sites with multi-site library designs is also possible.

Dissociation constant (Kd): an equilibrium constant that measures the strength of the binding interaction between a ligand and a partner (e.g., antibody and antigen). Low Kd values indicate high affinity and high Kd values indicate low affinity.

Epistatic interaction: describes the phenotypic in influence one mutation exerts on another mutation or mutations within a gene or protein. Some mutations may themselves be phenotypically silent but enhance the effects of subsequent mutations.

Epitope: the region of an antigen that binds to an antibody. Epitopes may be either linear (e.g., a chain of sequential amino acids) or conformational (e.g., amino acids encoded on different chains but in close proximity when folded).

Fluorescence-activated cell sorting (FACS): a laboratory technology based on ow cytometry that is used for sorting individual cells at high-throughput. Cells are labelled with fluorescent molecules, passed through a vibrating nozzle to generate single charged droplets, and sorted according to relative signal by laser excitation.

Monoclonal antibody (mAb): a laboratory-made antibody of single specificity with wide use as therapeutic drugs in cancer, inflammation, and infectious disease.

Box 2. Design of libraries for DMS experiments

Advances in oligonucleotide synthesis are enabling unique strategies for the rational design of mutagenesis libraries such as those used for DMS. Traditional single-site DMS is often performed using tiled saturation mutagenesis libraries covering a select region within a protein (**Figure I**). For example, in a 10 amino acid region of a mAb (e.g., a CDR), each position is systematically mutated to one of the 20 possible amino acids while the original sequence is fixed at the other nine positions. The resulting library will have a total diversity of only 200 (10 positions × 20 amino acids). More recently, methods to construct libraries for multi-site DMS have emerged which can capture the influence of epistatic interactions that are missed by libraries designed for single-site DMS. Several positions can be simultaneously saturated such that a triple-site library targeting the same 10 amino acid CDR region described previously will now have a total diversity of 960,000 (120 combinations of positions × 20^3 amino acids). Multi-site DMS libraries are distinct from unsaturated combinatorial libraries commonly used in protein engineering, where full diversity coverage cannot be guaranteed.

In DMS workflows, the final library size is ultimately dictated by the number of codons required to saturate each site. Degenerate oligonucleotides have been most commonly used to generate systematic diversity in proteins. The use of the NNK triplet, rather than NNN, enables encoding of 32 individual codons (rather than 64 codons) that cover all 20 amino acids and a single stop codon. Narrowing codon usage to minimise the amount of redundancy in the library allows for an increased protein sequence space to be investigated for a given library size. This becomes particularly important for multi-site DMS approaches that investigate pairwise or higher-order interaction linkage. Although more expensive to generate for each position, the use of three oligonucleotides with select degenerate codons (NDT/VHG/TGG) further reduces the total diversity to 22 codons and eliminates the potential for stop codons (Kille et al., 2013).

Massively parallel synthesis methods have further improved oligonucleotide cost efficiencies and now allow the generation of oligonucleotide variant pools from which site-saturated variant libraries can be directly assembled (Li et al., 2018). This approach offers precise codon control and complete non-redundancy by using 20 oligonucleotides encoding each of the 20 amino acids at each position in the library. These services are offered commercially by companies such as Twist Bioscience and GenScript.

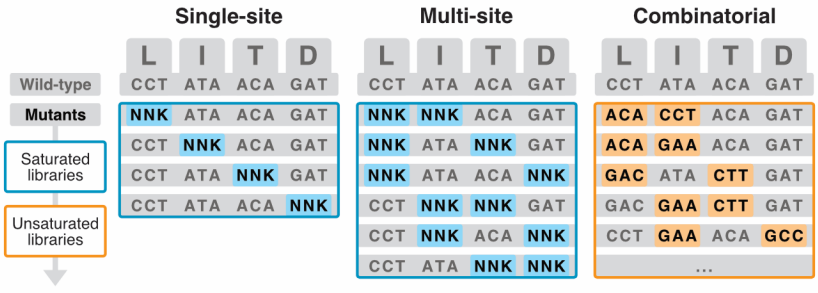


Figure I. Single-site DMS libraries saturate diversity at individual positions and are often created via the use of degenerate codons (N = A, C, G, or T; K = G or T). Higher diversity is generated by the simultaneous saturation of more than one position in libraries used for multi-site DMS. These libraries are distinct from the unsaturated combinatorial libraries that are commonly used in protein engineering, which may be generated by approaches such as error-prone PCR.

Box 3. Genotype-phenotype linkage by common mAb display systems

Ribosomal/mRNA display

In vitro transcription of mRNA from DNA lacking a stop codon produces transcripts that stall in the ribosome at the end of translation (Hanes et al., 1998). The result is an mAb fragment tethered via the ribosome to the coding sequence, allowing panning (screening for binding) against an immobilised antigen. Libraries in excess of 10^{12} variants can be generated, depending on the stability of the fragment and its ability to fold correctly.

Bacteriophage Display

Antibodies are anchored to the pIII coat protein of M13 bacteriophages using a phagemid system of production that decouples protein display from the phage replication cycle. Phagemids are specialised plasmids that encode the pIII fusion alongside replication origins capable of both double-stranded maintenance in bacteria and single-stranded derivatisation for packaging into phage virions. The addition of a helper phage supplies the remaining structural components required for packaging. Libraries of mAb fragments(Vaughan et al., 1996), or even full-length antibodies (Mazor et al., 2010), are panned against target antigens. More than 14 phage display-derived antibodies have been clinically approved (Alfaleh et al., 2020).

Bacterial Display

Directing mAb expression to the oxidising periplasmic environment of Gram-negative bacteria enables correct disulfide bond formation for assembly. For screening purposes, the subsequent export of these fragments to the outer membrane via anchoring on proteins such as OmpA or OmpF proteins enables antigen access (T. Chen et al., 2019; Daugherty et al., 1998). Alternatively, these fragments can remain sequestered in the periplasm and accessed via enzymatic digestion of the outer membrane/peptidoglycan layers (B. R. Harvey et al., 2004; Mazor et al., 2008). In both cases, high-throughput FACS is used to screen libraries of 10^{11} variants.

Yeast Display

The need to reduce assembly-induced library bias in bacterial platforms led to the development of eukaryotic display systems in *Saccharomyces cerevisiae* (Boder & Wittrup, 1997). Most commonly, mAb fragments are directly fused to one of subunits of a-agglutinin, Aga2p. This complex assembles via disulfide bonding to the membrane-anchored Aga1p subunit enabling screening by FACS. Library sizes of ~10^9 clones can be produced, but these are more than an order of magnitude lower than the possible diversity accessible by phage or bacterial display (Almagro et al., 2019).

Mammalian Display

Using mammalian cells for mAb display allows for higher levels of correct assembly, particularly for full-length antibodies. Although small library size is the primary obstacle to this technique (~10^5-10^7), the preservation of native glycosylation patterns that contribute to mAb effector function is an important benefit of the approach. Membrane anchoring is frequently achieved using transmembrane domain fusions that allow screening by FACS.

Library construction and diversity monitoring in DMS experiments

Approaches for constructing mAb libraries designed for DMS differ widely between protein display systems, although solid-phase oligonucleotide synthesis technologies have greatly improved the ease of these processes (Box 2). Mutagenesis libraries for *in vitro* (mRNA) and microbial display approaches (phage, bacteria and yeast) have generally relied on amplification-based manipulation of plasmid sequences with mutagenic oligonucleotide primers (Firnberg & Ostermeier, 2012; Wrenbeck et al., 2016). The popular and high-efficiency PFunkel technique relies on the extraction of uracil-containing single-stranded (ss) DNA PCR templates from bacteriophages. Following mutagenesis by primer-mediated amplification, unmodified uracil-rich templates undergo selective enzymatic degradation to generate a transformable library. Inspired by the PFunkel method, Wrenbeck *et al*. (2016) specifically sought to remove laborious ssDNA production steps by developing a simple and rapid one-pot approach for DMS-specific library creation based on nicking mutagenesis. Exonucleases are used in place of bacteriophages to create ssDNA templates that are then amplified by pooled mutagenic oligonucleotides. Where less diversity is required, other approaches have removed the requirement for ssDNA templates entirely. Oligonucleotide pools have been directly amplified by PCR to generate DMS diversity for phage display, and these molecules can be cloned by regular restriction enzyme approaches (Garrett et al., 2020a). The recently described deep insertion deletion (indel) missense programmable library engineering (DIMPLE) pipeline combines oligonucleotide pool design and synthesis, PCR, and type-IIS restriction cloning approaches to generate variant libraries that included structured indels (Macdonald et al., 2023). These mutational types are typically absent from DMS experiments owing to technical limitations, but their inclusion could prove impactful for antibody engineering applications.

Owing to the difficulty of long-term plasmid maintenance (stable transfection) in mammalian cells, the development of large libraries for DMS has been more challenging. With some exceptions (Forsyth et al., 2013), mAb libraries require stable integration into the host genome, often via viral-mediated (Beerli et al., 2008; D. Chen et al., 2021) or transposon-mediated gene delivery (Waldmeier et al., 2016). Targeted genomic integration has also been performed using CRISPR/Cas9 and homology-directed repair. Both yeast display (Oh et al., 2020), and mammalian display systems (Mason et al., 2018; Parthiban et al., 2019) have exploited CRISPR/Cas9 to generate mAb libraries at precisely defined locations using diversity derived from oligonucleotide pools. Although not yet applied to the generation of mAb libraries, the recent description of recombinase-based landing-pad systems package within

lentiviruses presents another avenue for the targeted integration of diversity over large stretches of sequence space (Matreyek et al., 2020). Similarly, recently developed high-throughput retron library recombineering methods (Schubert et al., 2021) show promise as a flexible, *in vivo* variant library production approach that can be utilised across display systems (Crawford et al., 2025; Lopez et al., 2022).

Following library generation, deep sequencing is used for quantitative characterisation and quality control (i.e., to quantify the frequency of mutations). Deep sequencing of the mAb library is then repeated again following a phenotypic selection or screening (e.g., binding to target antigen) (Fowler et al., 2014). Bioinformatic analysis of pre- and post-selection libraries enables quantification of the relative change in amino acid frequencies at each site in the library where over-represented residues indicate a positive contribution to the desired phenotype. Multiple deep sequencing technologies are available, depending on the read length required and tolerated error rate, with PacBio, Oxford nanopore, and Illumina being most common and are reviewed in detail elsewhere (Rouet et al., 2018). Molecular barcodes incorporated during variant library synthesis and/or sequencing preparation can improve the utility and use of sequencing strategies, including those that combine sequencing technologies (Garrett et al., 2020b).

Tools for the analysis and visualisation of DMS data

The use of deep sequencing to obtain rich sequence-function datasets requires bioinformatic processing and analysis tools for simple data interpretation. Sequences are preprocessed to remove low quality reads, counted, and the amino acid enrichment ratio (ER_v) is calculated at each position in a sequence for each amino acid variant (v). Typically, this is achieved by taking the ratio of observed amino acid counts, after (N_v^{after}) and before (N_v^{before}) selection relative to the original residue count (wt) (Eq. 1), where a value higher than 1 indicates enrichment. The enrichment ratio may further be log-transformed or rescaled to sum to one to give amino acid preference at each position.

$$ER_{r,i} = \frac{N_{r,i}^{after} / N_{r,wt}^{after}}{N_{r,i}^{before} / N_{r,wt}^{before}} \qquad (Equation\ 1)$$

However, enrichment ratios are highly susceptible to bias, particularly when variants are present at low frequency (Bloom, 2015). To account for this bias, several analysis tools have

developed statistical frameworks for error estimation in enrichment ratios, relying on replicate data to minimise other sources of systematic error in DMS experiments (e.g., amplification errors by polymerases or oligonucleotide synthesis errors) (Table 1).

Data visualisation of the results of these tools, as well as quality metrics, are automatically generated. Amino acid sequence logos describing positive differential selection or position-function heat maps are frequently used to show the relative influence of each mutation on phenotype (Fig. 2.2). The recently described dms-view package extends this visualization by allowing interactive mapping of DMS sites of interest onto 3D protein structures (Hilton et al., 2020). Together these tools have made DMS approaches accessible to researchers with limited bioinformatics experience.

Table 1. Pipeline tools for the analysis of DMS experiments

Tool	Error estimation	Key features of approach	Refs
dms_tools	Likelihood-based inference of mutation counts (Bayesian)	Increases enrichment score accuracy when individual variant sequencing read counts are below 1000 Works very well with shotgun sequencing data	(Bloom, 2015)
Enrich2	Random-effects model applied to replicate data	Good at noisy variant removal and detecting variants with small effects Standard error calculation enables statistical evaluation of variant significance Supports the analysis of time-series data from DMS experiments	(Rubin et al., 2017)
DimSUM	Replicate-specific error estimation generated by sharing information across all assayed variants	Excellent accuracy with over-dispersion in datasets (higher variability than expected) Benchmarked performance is generally better than Enrich2 Requires fewer replicate experiments than previous approaches	(Faure et al., 2020)

Implementing DMS to engineer mAb affinity, specificity and stability

mAbs derived from immune repertoires, or via synthetic library screening, frequently require additional protein engineering steps to optimise their therapeutic properties. One such property is the affinity of a mAb for its cognate antigen, typically expressed as a **dissociation constant (K_d)**, and is a key determinant of functional potency. Early-stage mAb candidates often do not possess the affinities required of therapeutics, and affinity maturation is therefore an essential

step in their development. The residues most directly responsible for mAb affinity and specificity lie within the CDR loops of the antibody variable domains. Generating CDR-targeted libraries for DMS has emerged as an efficient and rapid approach to identify mAb variants with improved affinities.

By performing single-site DMS on all six CDR loops of the variable region heavy (VH) and light (VL) chains of a mAb targeting TNF-α, Fujino *et al.* (2012) identified a suite of mutations enriched following selection for antigen binding. This DMS experiment informed the rational design of an unsaturated 12-site combinatorial mutagenesis library, which was subsequently screened by ribosome display to isolate variants with substantially enhanced affinity relative to the original mAb (up to 2110-fold). These variants possessed between five and seven amino acid mutations across their CDRs, several of which were situated at the periphery of the known antigen-binding interface. A similar single-site DMS approach was used to identify affinity enhancing mutations within the CDRs of an analogue of cetuximab, a clinically approved mAb targeting EGFR (Forsyth et al., 2013). Broad agreement was observed with the results from previous affinity-maturation experiments on the same mAb family. The top 25% of variants from the DMS workflow, grouped by enrichment ratio, captured the majority of these prior mutations despite differences in library design and screening. Binding kinetic analysis of 16 variants containing single mutations demonstrated promising affinity improvements of between 1.3- and 5.3-fold but revealed poor correlation with calculated enrichment ratios meaning detailed variant ranking by DMS alone was inaccurate. Significant effort has since focused on implementing more robust statistical analyses that correct enrichment ratios for errors (Table 1). Alternatively, Adams *et al.* (2016) developed a novel experimental framework (tite-seq) to improve the accuracy of affinity estimation during DMS. Instead of relying on single enrichment ratios as a proxy for affinity, full titration curves are created by screening and sequencing yeast display mAb fragment libraries against various concentrations of antigen to directly estimate K_d values at high-throughput. At each concentration, **fluorescence activated cell sorting (FACS)** is used to sort the library into four bins based on affinity. DMS was performed on the CDR regions of a mAb targeting the fluorescein molecule, and screening was performed with 11 different antigen concentrations. Comparison of tite-seq affinities estimated from the bulk experiment compared favourably to those obtained from select single variant FACS analysis (Pearson correlation coefficients ranging from 0.82 to 0.89).

There has been early recognition in the field that single-site DMS is unable to resolve **epistatic interactions** or combinatorial mutational effects on protein function, namely where

one mutation exerts a phenotypic effect contingent on another mutation or mutations. New oligonucleotide synthesis and library design techniques (Box 3) have enabled the construction of multi-site DMS libraries to tackle this issue. For example, D. Chen *et al.* (2021) created multi-site libraries for DMS within the IgG1 Fc domain to specifically screen for enhanced affinity to select Fc receptors responsible for driving *in vivo* effector functions. Multiple variants with significantly enhanced antibody-dependent cell cytotoxicity profiles were directly isolated following lentiviral-mediated mammalian display. A similar approach has also been used to improve the affinity of the unique dual-specific mAb 5A12, which targets two structurally unrelated antigens (VEGF and Ang2) (Adams et al., 2016). Originating from an anti-VEGF mAb, the process of raising additional Ang2 specificity compromised the affinities for both targets in 5A12. The authors were able to quickly rectify this by DMS with multi-site libraries containing mutations in each of the CDRs of either the VL or VH. Synergistic mutations that improved affinities for VEGF or Ang2 were first identified and then combined to generate a panel of improved variants with sub-nanomolar monovalent affinity, a first for dual specific antibodies. The most promising candidate, 5A12.1, exhibited a 25-fold affinity improvement for both VEGF and Ang2 antigens but also developed unanticipated off-target specificity for Ang1. Koenig *et al*., (2015) again used a multi-site DMS approach targeting the CDRs of 5A12.1 to engineer reduced Ang1 affinity while retaining affinity for VEGF and Ang2. The authors created a high-resolution map of residues responsible for binding each of the three antigens and identified five variants with improved specificity. A more recent example by Banach *et al.* (2022) followed up a single-site DMS experiment with DNA-shuffling and multi-site mutagenesis to improve the affinity of a promising anti-malarial mAb, CIS43. The top performing variant contained three substitutions, notably two of which were considered naturally rare (≤0.13% in human repertoires each), and provided six-fold improved protection in a mouse model.

Optimising mAb affinity and specificity by protein engineering sometimes comes at the expense of stability (Koenig, Sanowar, et al., 2017), which potentially reduces therapeutic efficacy. Rapid workflows employing DMS are increasingly being used in mAb development pipelines to identify important mutations, frequently distal to the antigen binding site, that compensate for destabilising effects.

Amino acid sites important for thermostability were identified in a high-affinity engineered mAb against VEGF by DMS of the entire protein sequence (Koenig, Lee, et al., 2017). DMS revealed residues in both the hydrophobic core and at the interface of the VH/VL

chains that were strongly intolerant to mutations. One mutation in particular, located 25 Å away from the VEGF binding site, exhibited the greatest improvement in thermostability while also increasing the VEGF-binding affinity by three-fold. The observation of long-range mutational effects was mirrored in separate work that undertook DMS of an anti-lysozyme mAb. Several affinity enhancing mutations were identified at the VH/VL interface that likely stabilised the assembly (Warszawski et al., 2019). In both studies, variants enriched when selected for stability were broadly similar to those selected for affinity, implying a strong relationship between stable structure and sustained antigen binding.

DMS stability screens have also focused on the Fc region of mAbs. For example, Traxlmayr *et al.*, (2012) determined the stability landscape of the CH2 and CH3 domains of human IgG1 by subjecting yeast-displayed fragments to thermal denaturation. Correctly refolded variants were enriched by FACS following baiting with structurally specific ligands such as an Fc receptor. Losses in library diversity were most severe in residues of the CH3-CH3 homodimer interface, highlighting the essential contribution this interaction makes to mAb stability. Although particularly striking in this example, the vast majority of substitutions made in DMS experiments are either neutral or deleterious to mAb function. Nonetheless, defining neutrally tolerated mutations (enrichment ratio ~1) is especially relevant when considering other clinically desirable properties, such as the requirement for low immunogenicity. The cataloging of neutral substitutions provides a narrowed sequence space that can be screened for secondary attribute enhancement without the loss of primary function.

The trade-off between affinity and stability can be further mitigated by the use of multi-site libraries for DMS to evaluate epistatic linkages during screening. Aggregation propensity, solubility, and immunogenicity properties of the clinical antibody bococizumab were optimised by DMS using a multi-site library targeting a single CDR (Koenig, Lee, et al., 2017). High solubility and low aggregation propensity were selected for by isolating high-expressing variants displayed on mammalian cells. A total of 27 amino acid combinations were enriched with complementary decreases in immunogenic responses as assessed by *in vitro* helper T cell proliferation assays. The most promising variant had an immunogenic response equivalent to other clinically approved antibodies such as trastuzumab.

Predicting mAb epitope binding and escape by DMS

The identification of mAb antigen binding **epitopes** is critical to the understanding of therapeutic mechanisms and for the design of superior vaccine immunogens (Jardine et al.,

2016). Screening a vast array of potential antigenic epitopes can be a hugely laborious task, and many approaches have only coarse sequence resolution (Garrett et al., 2020b). DMS, on the other hand, provides high-resolution, position-specific epitope mapping by screening site-saturation libraries of antigen targets for binding to candidate mAbs. In addition to the precise identification of an epitope, DMS also provides insights into mutations that could drive antigen escape, and this has particular relevance for antigens from continuously evolving pathogens (e.g., influenza and SARS-CoV-2).

Initial examples of epitope mapping by DMS sought to validate the approach, and included model mAb panels against cellular prion protein (Doolan & Colby, 2015), and *Staphylococcus aureus* alpha toxin (Van Blarcom et al., 2015). These techniques confirmed both previously identified linear and conformational epitope sequences as well as revealing several undiscovered residues. Epitope mapping of homotrimeric tumour necrosis factor against the clinically available infliximab, and *Bordetella pertussis* toxin against the promising mAb Hu1B7, further demonstrated DMS as a rapid methodology to achieve high resolution for more cryptic epitopes (Kowalsky et al., 2015). DMS continues to be useful for fine conformational epitope prediction of clinical antibodies (Medina-Cucurella et al., 2018), but some antigens with complex structure or multi-domain assemblies can be difficult to correctly display for DMS, hence alternative epitope mapping technologies remain prevalent in the field (e.g., synthetic peptide mapping) (Forsström et al., 2014). Where DMS has instead excelled is in the comprehensive prediction of epitopes able to escape from mAb binding.

Antigens under the strong selective pressure of therapeutic or host-generated antibodies frequently mutate to evade neutralisation (Eguia et al., 2021). The identification of antigen amino acid substitutions that facilitate escape is therefore crucial to predicting the limitations of a therapeutic mAb *in vivo*, and consequently aids the development of mAb cocktails targeting heterogeneous epitopes. Using DMS for this purpose has recently seen a surge of intense focus, especially with regard to viral antigens such as the Zika virus envelope protein (Sourisseau et al., 2019), H3N2 virus hemagglutinin (Doud et al., 2017; Guthmiller et al., 2022; Lee et al., 2018; Wu et al., 2017, 2020), HIV envelope glycoprotein (Dingens et al., 2017), and the SARS-CoV-2 spike receptor binding domain (RBD) (Harvey et al., 2021).

In the context of SARS-CoV-2, infectivity is principally determined by RBD interaction with the host cell-surface protein angiotensin-converting enzyme 2 (ACE2). Starr *et al.* (2020) recently performed DMS with single-site saturation mutagenesis libraries of the RBD displayed on the surface of yeast to quantify the effect substitutions have on binding to ACE2.

Importantly, DMS identified several mutations such as N501Y that were eventually found in circulating variants of SARS-CoV-2 (e.g., alpha variant). The RBD is also a primary targeting site of neutralising antibodies, either mAbs used for therapeutic treatment or polyclonal antibodies generated in response to viral infection or vaccination. DMS has also been performed on the RBD to identify the impact of single-position substitutions on binding to neutralising therapeutic antibodies, as well as serum from convalescent or vaccinated individuals (Starr, Czudnochowski, Liu, et al., 2021a; Starr, Greaney, Addetia, et al., 2021; Starr, Greaney, Dingens, et al., 2021; Tortorici et al., 2021; Greaney et al., 2021; Dong et al., 2021; Tsai et al., 2021). Of concern was the observation of single-mutation epitope escape from all current clinically trialled therapeutic mAbs, although DMS revealed several preclinical mAb candidates with more promise in resisting RBD escape, including one with particularly broad pan-sarbecovirus RBD specificity (Starr, Czudnochowski, Liu, et al., 2021b).

Together these experiments, combined with the sequencing of emerging strains, provide strong rationale for the use of clinical mAb cocktails. For example, DMS was performed to map prospective RBD escape variants for the previously clinically approved mAb cocktail consisting of REGN10933 and REGN10987; this revealed that there was a single mutation variant that escaped both antibodies (Starr, Czudnochowski, Liu, et al., 2021b). Similarly, a recent study by Schiepers *et al.* (2023) demonstrated supressed efficacy of serum antibody responses upon sequential exposure to SARS-CoV-2 antigen variants. This phenomenon, often known as primary addiction, is particularly pronounced with shorter antigen mutation distances and has highlights the importance of understanding antigen escape for vaccine design. One limitation to these SARS-CoV-2 studies is that they have almost entirely relied on single-site DMS libraries, however, it is apparent that emerging variants of SARS-CoV-2 (e.g., beta, gamma, delta and omicron variants) often have multiple mutations in their RBD that may play a role in their increased transmissibility and/or partial resistance to neutralising antibodies and vaccine-induced immunity. Moreover, evolutionary pathways frequently accumulate bystander mutations that have no phenotypic effect themselves but enhance the influence of subsequent mutations. The difficulty of fully surveying and interpreting combinatorial mutations represented in such emerging variants remains an important challenge for the field and has ultimately led to the retraction of multiple promising mAbs.

DMS and computationally guided protein engineering

A move towards studying higher-order interactions by DMS requires analysis beyond simple

heatmaps and sequence logos. As oligonucleotide synthesis technologies continue to improve the diversity of library construction, interpreting epistatic interactions and those within larger stretches of protein sequence space becomes intractable. Machine-learning techniques are capable of extrapolating from the inherently limited fraction of protein sequence space sampled by DMS (Fig. 2.3). Predicting the potential influence of unseen mutations provides guidance on sites where further experimental screening should be focused. It should also be noted that outside of a DMS framework, there has been a rapid rise in machine learning approaches being applied to many aspects of mAb development such as immunogenicity reduction (Marks et al., 2021) or affinity enhancement (Pittala & Bailey-Kellogg, 2020; Wang et al., 2020).

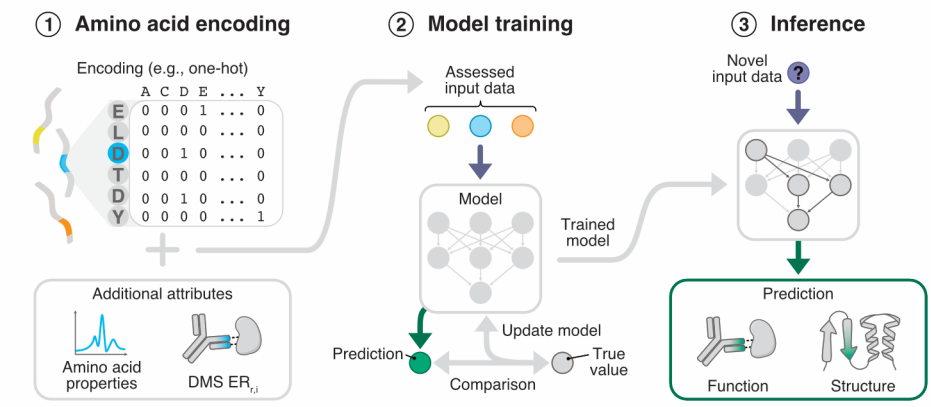


Figure 2.3. Machine learning extrapolates from deep mutational scanning (DMS)-derived data to predict the influence of protein mutations. ① The outputs from DMS experiments are encoded in a machine-readable format. In this case, one-hot encoding is used to create an input matrix for each variant sequence, and this is combined with attribute data to create the final vector. ② The resulting data are used to iteratively train machine-learning algorithms where the model parameters are constantly refined with new data (e.g., deep learning models). ③ Once trained, the model can predict combinations of positions likely to have a beneficial effect on the property screened for or aid in structural predictions. These predictions can massively reduce the total protein sequence space required to be searched to develop enhanced mAb variants. Abbreviation: $ER_{r,i}$, amino acid enrichment ratio.

Machine learning broadly describes algorithms capable of learning patterns within high-dimensional data sets. Unlike traditional mechanistic modelling, these algorithms do not require user-specified rules or approximations of physical laws governing the behaviour of systems; instead, the algorithms learn patterns directly from training data. Generalisable techniques have emerged to augment DMS protein-engineering workflows that use a variety of traditional machine-learning models (e.g., support vector machines, logistic regression) (Reeb et al., 2020; Sarfati et al., n.d.; Tareen et al., 2020) as well as more complex **deep learning** approaches (e.g., convolution neural networks, recurrent neural networks, generative adversarial networks, language models) (Graves et al., 2020). As a subset of machine learning,

deep learning uses neural network architectures to more efficiently handle large-scale datasets and requires less input data curation than other machine learning approaches. These data hungry algorithms have risen to prominence in protein engineering alongside advances in computer hardware and cloud computing availability. The data provided for training deep learning models can either be labelled (supervised; e.g., sequences binned as high or low affinity) or unlabelled (unsupervised; sequences only). In the case of unsupervised learning, the model itself identifies patterns which can then be used to group input data. While the choice of algorithm used to analyse DMS experiments for mAb engineering is often context-dependent, a recent focus has been the design of libraries with precisely defined diversity.

Removing non-productive sequences from a library pool biases selections towards a desired function saving time and resource in mAb engineering workflows. DMS experiments have been essential as benchmarks or as training data in several new computational library minimisation strategies. Due to the exponential rate of advancement in the field, we concede this snapshot is non-exhaustive.

To improve library design for initial candidate discovery, Shin *et al.* (2021) drew inspiration from text-to-speech algorithms to fit an unsupervised model to the native sequences of 33 different protein families. The trained model was then used to predict the influence of point mutations on variant fitness for proteins within each family and benchmarked against observations from experimental datasets obtained via DMS. The approach was found to be highly generalisable with state-of-the-art mutation effect prediction performance and was subsequently fit to more than a million natural camelid nanobody sequences. A key innovation of the approach is the removal of need for sequence alignment for input, enabling diverse CDR lengths to be sampled. The model was then used to predict sequences for a highly focused nanobody library that broadly covered the naïve repertoire sequence space (~180,000 CDRH3 sequences). Upon yeast display, variant expression levels within the nanobody library were significantly better than for synthetic libraries that are orders of magnitude larger in size, and weak binders for human serum albumin could be isolated directly (9.8 μM).

Affinity-maturation techniques for candidate development have also benefited from DMS and deep learning. Mason *et al.* (2021), performed CRISPR/Cas9-mediated DMS on the CDRH3 region of the full-length IgG trastuzumab using a mammalian display platform. The initial DMS data was used to constrain a subsequent combinatorial library with a theoretical diversity of ~700 million sequences. Screening and sequencing of only ~40,000 combinatorial variants (0.0054% of theoretical library diversity) enabled supervised training of an ensemble

of neural networks that classified antigen binding for the remaining unseen sequences *in silico*. Candidates were computationally filtered for developability (viscosity, clearance, solubility, and immunogenicity) and experimentally expressed to determine affinity. The workflow generated multiple candidates on par, or better, than the native trastuzumab across the metrics tested.

Targeted CDR diversification was also used in a series of phage libraries created by randomising and varying the length of CDRH3 regions of multiple clinical mAbs (ranibizumab, trastuzumab, and bevacizumab) (Liu et al., 2020). The experimental data was used to train an ensemble of neural networks to predict antigen binding. Termed Ens-Grad, the model is then used to eliminate non-specific sequences with affinity for undesired targets. In addition, the authors apply their method to design a computationally optimised phage library which they then demonstrate yields more diverse sequences in fewer panning rounds than traditional screening.

Concluding remarks and future perspectives

DMS has emerged as an efficient approach to engineer next generation therapeutic antibodies. Strategic implementation of DMS, especially when paired with machine learning, has the potential to improve mAb development timelines and candidate success rates by rapid early stage optimization of key properties, including affinity, specificity, stability, and escapability. Technological advancements in oligonucleotide synthesis and machine learning have overcome many limitations to DMS, although several questions remain unresolved (see outstanding questions). The rich data generated by DMS approaches will not only continue to inform mAb engineering efforts but will also be important for training *in silico* protein design algorithms. Ultimately, we expect therapeutic mAb design will be reliant on extremely robust *in silico* prediction tools, reducing the need for intensive wet-lab screening. Recent advances in protein structure prediction have presented potentially new avenues to mAb design (Akpinaroglu et al., 2021; Jumper et al., 2021; Meier et al., 2021; Rives et al., 2021; Ruffolo et al., 2021) but the prediction of antigen binding (docking) from these models remains a frontier. Data generated from DMS experiments will continue to be indispensable in generating such predictive models.

References

Adams, R. M., Mora, T., Walczak, A. M., & Kinney, J. B. (2016). Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife*, *5*, e23156. https://doi.org/10.7554/eLife.23156

Akpinaroglu, D., Ruffolo, J. A., Mahajan, S. P., & Gray, J. J. (2021). Improved antibody structure prediction by deep learning of side chain conformations. *bioRxiv*, 2021.09.22.461349. https://doi.org/10.1101/2021.09.22.461349

Alfaleh, M. A., Alsaab, H. O., Mahmoud, A. B., Alkayyal, A. A., Jones, M. L., Mahler, S. M., & Hashem, A. M. (2020). Phage display derived monoclonal antibodies: From bench to bedside. *Frontiers in Immunology*, *11*, 1986. https://doi.org/10.3389/fimmu.2020.01986

Almagro, J. C., Pedraza-Escalona, M., Arrieta, H. I., & Pérez-Tapia, S. M. (2019). Phage display libraries for antibody therapeutic discovery and development. *Antibodies*, *8*(3), Article 3. https://doi.org/10.3390/antib8030044

Banach, B. B., Tripathi, P., Da Silva Pereira, L., Gorman, J., Nguyen, T. D., Dillon, M., Fahad, A. S., Kiyuka, P. K., Madan, B., Wolfe, J. R., Bonilla, B., Flynn, B., Francica, J. R., Hurlburt, N. K., Kisalu, N. K., Liu, T., Ou, L., Rawi, R., Schön, A., … DeKosky, B. J. (2022). Highly protective antimalarial antibodies via precision library generation and yeast display screening. *Journal of Experimental Medicine*, *219*(8), e20220323. https://doi.org/10.1084/jem.20220323

Beerli, R. R., Bauer, M., Buser, R. B., Gwerder, M., Muntwiler, S., Maurer, P., Saudan, P., & Bachmann, M. F. (2008). Isolation of human monoclonal antibodies by mammalian cell display. *Proceedings of the National Academy of Sciences*, *105*(38), 14336–14341. https://doi.org/10.1073/pnas.0805942105

Bloom, J. D. (2015). Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics*, *16*(1), 168. https://doi.org/10.1186/s12859-015-0590-4

Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology*, *15*(6), 553–557. https://doi.org/10.1038/nbt0697-553

Burks, E. A., Chen, G., Georgiou, G., & Iverson, B. L. (1997). In vitro scanning saturation mutagenesis of an antibody binding pocket. *Proceedings of the National Academy of Sciences*, *94*(2), 412–417. https://doi.org/10.1073/pnas.94.2.412

Carter, P. J., & Lazar, G. A. (2018). Next generation antibody drugs: Pursuit of the "high-hanging fruit." *Nature Reviews Drug Discovery*, *17*(3), 197–223. https://doi.org/10.1038/nrd.2017.227

Chen, D., Zhao, Y., Li, M., Shang, H., Li, N., Li, F., Wang, W., Wang, Y., Jin, R., Liu, S., Li, X., Gao, S., Tian, Y., Li, R., Li, H., Zhang, Y., Du, M., Cao, Y., Zhang, Y., … Zhang, H. (2021). A general Fc engineering platform for the next generation of antibody therapeutics. *Theranostics*, *11*(4), 1901–1917. https://doi.org/10.7150/thno.51299

Chen, G., Dubrawsky, I., Mendez, P., Georgiou, G., & Iverson, B. L. (1999). In vitro scanning saturation mutagenesis of all the specificity determining residues in an antibody binding site. *Protein Engineering, Design and Selection*, *12*(4), 349–356. https://doi.org/10.1093/protein/12.4.349

Chen, T., Wang, K., Chi, X., Zhou, L., Li, J., Liu, L., Zheng, Q., Wang, Y., Yu, H., Gu, Y., Zhang, J., Li, S., & Xia, N. (2019). Construction of a bacterial surface display system based on outer membrane protein F. *Microbial Cell Factories*, *18*(1), 70. https://doi.org/10.1186/s12934-019-1120-2

Crawford, K. D., Khan, A. G., Lopez, S. C., Goodarzi, H., & Shipman, S. L. (2025). High throughput variant libraries and machine learning yield design rules for retron gene editors. *Nucleic Acids Research*, *53*(2), gkae1199. https://doi.org/10.1093/nar/gkae1199

Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L., & Georgiou, G. (1998). Antibody affinity maturation using bacterial surface display. *Protein Engineering Design and Selection*, *11*(9), 825–832. https://doi.org/10.1093/protein/11.9.825

Dingens, A. S., Haddox, H. K., Overbaugh, J., & Bloom, J. D. (2017). Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. *Cell Host & Microbe*, *21*(6), 777-787.e4. https://doi.org/10.1016/j.chom.2017.05.003

Dong, J., Zost, S. J., Greaney, A. J., Starr, T. N., Dingens, A. S., Chen, E. C., Chen, R. E., Case, J. B., Sutton, R. E., Gilchuk, P., Rodriguez, J., Armstrong, E., Gainza, C., Nargi, R. S., Binshtein, E., Xie, X., Zhang, X., Shi, P.-Y., Logue, J., … Crowe, J. E. (2021). Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. *Nature Microbiology*, *6*(10), 1233–1244. https://doi.org/10.1038/s41564-021-00972-2

Doolan, K. M., & Colby, D. W. (2015). Conformation-Dependent Epitopes Recognized by Prion Protein Antibodies Probed Using Mutational Scanning and Deep Sequencing. *Journal of Molecular Biology*, *427*(2), 328–340. https://doi.org/10.1016/j.jmb.2014.10.024

Doud, M. B., Hensley, S. E., & Bloom, J. D. (2017). Complete mapping of viral escape from neutralizing antibodies. *PLOS Pathogens*, *13*(3), e1006271. https://doi.org/10.1371/journal.ppat.1006271

Eguia, R. T., Crawford, K. H. D., Stevens-Ayers, T., Kelnhofer-Millevolte, L., Greninger, A. L., Englund, J. A., Boeckh, M. J., & Bloom, J. D. (2021). A human coronavirus evolves antigenically to escape antibody immunity. *PLOS Pathogens*, *17*(4), e1009453. https://doi.org/10.1371/journal.ppat.1009453

Faure, A. J., Schmiedel, J. M., Baeza-Centurion, P., & Lehner, B. (2020). DiMSum: An error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. *Genome Biology*, *21*(1), 207. https://doi.org/10.1186/s13059-020-02091-3

Firnberg, E., & Ostermeier, M. (2012). PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLOS ONE*, *7*(12), e52031. https://doi.org/10.1371/journal.pone.0052031

Forsström, B., Axnäs, B. B., Stengele, K.-P., Bühler, J., Albert, T. J., Richmond, T. A., Hu, F. J., Nilsson, P., Hudson, E. P., Rockberg, J., & Uhlen, M. (2014). Proteome-wide epitope mapping of antibodies using ultra-dense peptide arrays. *Molecular & Cellular Proteomics*, *13*(6), 1585–1597. https://doi.org/10.1074/mcp.M113.033308

Forsyth, C. M., Juan, V., Akamatsu, Y., DuBridge, R. B., Doan, M., Ivanov, A. V., Ma, Z., Polakoff, D., Razo, J., Wilson, K., & Powers, D. B. (2013). Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. *mAbs*, *5*(4), 523–532. https://doi.org/10.4161/mabs.24979

Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D., & Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. *Nature Methods*, *7*(9), 741–746. https://doi.org/10.1038/nmeth.1492

Fowler, D. M., Stephany, J. J., & Fields, S. (2014). Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nature Protocols*, *9*(9), 2267–2284. https://doi.org/10.1038/nprot.2014.153

Fujino, Y., Fujita, R., Wada, K., Fujishige, K., Kanamori, T., Hunt, L., Shimizu, Y., & Ueda, T. (2012). Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. *Biochemical and Biophysical Research Communications*, *428*(3), 395–400. https://doi.org/10.1016/j.bbrc.2012.10.066

Garrett, M. E., Itell, H. L., Crawford, K. H. D., Basom, R., Bloom, J. D., & Overbaugh, J. (2020a). Phage-DMS: A comprehensive method for fine mapping of antibody epitopes. *iScience*, *23*(10), 101622. https://doi.org/10.1016/j.isci.2020.101622

Garrett, M. E., Itell, H. L., Crawford, K. H. D., Basom, R., Bloom, J. D., & Overbaugh, J. (2020b). Phage-DMS: A Comprehensive Method for Fine Mapping of Antibody Epitopes. *iScience*, *23*(10), 101622. https://doi.org/10.1016/j.isci.2020.101622

Graves, J., Byerly, J., Priego, E., Makkapati, N., Parish, S. V., Medellin, B., & Berrondo, M. (2020). A review of deep learning methods for antibodies. *Antibodies*, *9*(2), 12. https://doi.org/10.3390/antib9020012

Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., Loes, A. N., Hilton, S. K., Huddleston, J., Eguia, R., Crawford, K. H. D., Dingens, A. S., Nargi, R. S., Sutton, R. E., Suryadevara, N., Rothlauf, P. W., Liu, Z., Whelan, S. P. J., Carnahan, R. H., Crowe, J. E., & Bloom, J. D. (2021). Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host & Microbe*, *29*(1), 44-57.e9. https://doi.org/10.1016/j.chom.2020.11.007

Guthmiller, J. J., Han, J., Utset, H. A., Li, L., Lan, L. Y.-L., Henry, C., Stamper, C. T., McMahon, M., O'Dell, G., Fernández-Quintero, M. L., Freyn, A. W., Amanat, F., Stovicek, O., Gentles, L., Richey, S. T., de la Peña, A. T., Rosado, V., Dugan, H. L., Zheng, N.-Y., … Wilson, P. C. (2022). Broadly neutralizing antibodies target a haemagglutinin anchor epitope. *Nature*, *602*(7896), 314–320. https://doi.org/10.1038/s41586-021-04356-8

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., & Plückthun, A. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proceedings of the National Academy of Sciences*, *95*(24), 14130–14135. https://doi.org/10.1073/pnas.95.24.14130

Harvey, B. R., Georgiou, G., Hayhurst, A., Jeong, K. J., Iverson, B. L., & Rogers, G. K. (2004). Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. *Proceedings of the National Academy of Sciences*, *101*(25), 9193–9198. https://doi.org/10.1073/pnas.0400187101

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*(7), 409–424. https://doi.org/10.1038/s41579-021-00573-0

Hilton, S., Huddleston$, J., Black, A., North, K., Dingens, A., Bedford, T., & Bloom, J. (2020). dms-view: Interactive visualization tool for deep mutational scanning data. *Journal of Open Source Software*, *5*(52), 2353. https://doi.org/10.21105/joss.02353

Jardine, J. G., Kulp, D. W., Havenar-Daughton, C., Sarkar, A., Briney, B., Sok, D., Sesterhenn, F., Ereño-Orbea, J., Kalyuzhniy, O., Deresa, I., Hu, X., Spencer, S., Jones, M., Georgeson, E., Adachi, Y., Kubitz, M., deCamp, A. C., Julien, J.-P., Wilson, I. A., … Schief, W. R. (2016). HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science*, *351*(6280), 1458–1463. https://doi.org/10.1126/science.aad9195

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaplon, H., & Reichert, J. M. (2021). Antibodies to watch in 2021. *mAbs*, *13*(1), 1860476. https://doi.org/10.1080/19420862.2020.1860476

Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z.-G., Opperman, D. J., Reetz, M. T., & Acevedo, J. P. (2013). Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synthetic Biology*, *2*(2), 83–92. https://doi.org/10.1021/sb300037w

Koenig, P., Lee, C. V., Sanowar, S., Wu, P., Stinson, J., Harris, S. F., & Fuh, G. (2015). Deep sequencing-guided design of a high affinity dual specificity antibody to target two angiogenic factors in neovascular age-related macular degeneration. *Journal of Biological Chemistry*, *290*(36), 21773–21786.

Koenig, P., Lee, C. V., Walters, B. T., Janakiraman, V., Stinson, J., Patapoff, T. W., & Fuh, G. (2017). Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proceedings of the National Academy of Sciences*, *114*(4), E486–E495. https://doi.org/10.1073/pnas.1613231114

Koenig, P., Sanowar, S., Lee, C. V., & Fuh, G. (2017). Tuning the specificity of a Two-in-One Fab against three angiogenic antigens by fully utilizing the information of deep mutational scanning. *mAbs*, *9*(6), 959–967. https://doi.org/10.1080/19420862.2017.1337618

Kowalsky, C. A., Faber, M. S., Nath, A., Dann, H. E., Kelly, V. W., Liu, L., Shanker, P., Wagner, E. K., Maynard, J. A., Chan, C., & Whitehead, T. A. (2015). Rapid Fine Conformational Epitope Mapping Using Comprehensive Mutagenesis and Deep Sequencing *. *Journal of Biological Chemistry*, *290*(44), 26457–26470. https://doi.org/10.1074/jbc.M115.676635

Lee, J. M., Huddleston, J., Doud, M. B., Hooper, K. A., Wu, N. C., Bedford, T., & Bloom, J. D. (2018). Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proceedings of the National Academy of Sciences*, *115*(35), E8276–E8285. https://doi.org/10.1073/pnas.1806133115

Li, A., Sun, Z., & Reetz, M. T. (2018). Solid-phase gene synthesis for mutant library construction: The future of directed evolution? *Chembiochem: A European Journal of Chemical Biology*, *19*(19), 2023–2032. https://doi.org/10.1002/cbic.201800339

Liu, G., Zeng, H., Mueller, J., Carter, B., Wang, Z., Schilz, J., Horny, G., Birnbaum, M. E., Ewert, S., & Gifford, D. K. (2020). Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics*, *36*(7), 2126–2133. https://doi.org/10.1093/bioinformatics/btz895

Lopez, S. C., Crawford, K. D., Lear, S. K., Bhattarai-Kline, S., & Shipman, S. L. (2022). Precise genome editing across kingdoms of life using retron-derived DNA. *Nature Chemical Biology*, *18*(2), 199–206. https://doi.org/10.1038/s41589-021-00927-y

Macdonald, C. B., Nedrud, D., Grimes, P. R., Trinidad, D., Fraser, J. S., & Coyote-Maestas, W. (2023). DIMPLE: Deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. *Genome Biology*, *24*(1), 36. https://doi.org/10.1186/s13059-023-02880-6

Marks, C., Hummer, A. M., Chin, M., & Deane, C. M. (2021). Humanization of antibodies using a machine learning approach on large-scale repertoire data. *Bioinformatics*, *btab434*. https://doi.org/10.1093/bioinformatics/btab434

Mason, D. M., Friedensohn, S., Weber, C. R., Jordi, C., Wagner, B., Meng, S. M., Ehling, R. A., Bonati, L., Dahinden, J., Gainza, P., Correia, B. E., & Reddy, S. T. (2021). Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature Biomedical Engineering*, *5*(6), 600–612. https://doi.org/10.1038/s41551-021-00699-9

Mason, D. M., Weber, C. R., Parola, C., Meng, S. M., Greiff, V., Kelton, W. J., & Reddy, S. T. (2018). High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis. *Nucleic Acids Research*, *46*(14), 7436–7449. https://doi.org/10.1093/nar/gky550

Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N., & Fowler, D. M. (2020). An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Research*, *48*(1), e1. https://doi.org/10.1093/nar/gkz910

Mazor, Y., Van Blarcom, T., Carroll, S., & Georgiou, G. (2010). Selection of full-length IgGs by tandem display on filamentous phage particles and Escherichia coli fluorescence-activated cell sorting screening. *The FEBS Journal*, *277*(10), 2291–2303. https://doi.org/10.1111/j.1742-4658.2010.07645.x

Mazor, Y., Van Blarcom, T., Iverson, B. L., & Georgiou, G. (2008). E-clonal antibodies: Selection of full-length IgG antibodies using bacterial periplasmic display. *Nature Protocols*, *3*(11), 1766–1777. https://doi.org/10.1038/nprot.2008.176

Medina-Cucurella, A. V., Zhu, Y., Bowen, S. J., Bergeron, L. M., & Whitehead, T. A. (2018). Pro region engineering of nerve growth factor by deep mutational scanning enables a yeast platform for conformational epitope mapping of anti-NGF monoclonal antibodies. *Biotechnology and Bioengineering*, *115*(8), 1925–1937. https://doi.org/10.1002/bit.26706

Meier, J., Rao, R., Verkuil, R., Liu, J., Sercu, T., & Rives, A. (2021). Language models enable zero-shot prediction of the effects of mutations on protein function. *bioRxiv*. https://doi.org/10.1101/2021.07.09.450648

Mullard, A. (2021). FDA approves 100th monoclonal antibody product. *Nature Reviews Drug Discovery*, *20*(7), 491–495. https://doi.org/10.1038/d41573-021-00079-7

Norman, R. A., Ambrosetti, F., Bonvin, A. M. J. J., Colwell, L. J., Kelm, S., Kumar, S., & Krawczyk, K. (2020). Computational approaches to therapeutic antibody design: Established methods and emerging trends. *Briefings in Bioinformatics*, *21*(5), 1549–1567. https://doi.org/10.1093/bib/bbz095

Oh, E. J., Liu, R., Liang, L., Freed, E. F., Eckert, C. A., & Gill, R. T. (2020). Multiplex evolution of antibody fragments utilizing a yeast surface display platform. *ACS Synthetic Biology*, *9*(8), 2197–2202. https://doi.org/10.1021/acssynbio.0c00159

Parthiban, K., Perera, R. L., Sattar, M., Huang, Y., Mayle, S., Masters, E., Griffiths, D., Surade, S., Leah, R., Dyson, M. R., & McCafferty, J. (2019). A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. *mAbs*, *11*(5), 884–898. https://doi.org/10.1080/19420862.2019.1618673

Pittala, S., & Bailey-Kellogg, C. (2020). Learning context-aware structural representations to predict antigen and antibody binding interfaces. *Bioinformatics*, *36*(13), 3996–4003. https://doi.org/10.1093/bioinformatics/btaa263

Reeb, J., Wirth, T., & Rost, B. (2020). Variant effect predictions capture some aspects of deep mutational scanning experiments. *BMC Bioinformatics*, *21*(1), 107. https://doi.org/10.1186/s12859-020-3439-4

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, *118*(15), e2016239118. https://doi.org/10.1073/pnas.2016239118

Rouet, R., Jackson, K. J. L., Langley, D. B., & Christ, D. (2018). Next-generation sequencing of antibody display repertoires. *Frontiers in Immunology*, *9*, 118. https://doi.org/10.3389/fimmu.2018.00118

Rubin, A. F., Gelman, H., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P., & Fowler, D. M. (2017). A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, *18*(1), 150. https://doi.org/10.1186/s13059-017-1272-5

Ruffolo, J. A., Sulam, J., & Gray, J. J. (2021). Antibody structure prediction using interpretable deep learning. *bioRxiv*. https://doi.org/10.1101/2021.05.27.445982

Sarfati, H., Naftaly, S., Papo, N., & Keasar, C. (n.d.). Predicting mutant outcome by combining deep mutational scanning and machine learning. *Proteins: Structure, Function, and Bioinformatics*. https://doi.org/10.1002/prot.26184

Schiepers, A., van 't Wout, M. F. L., Greaney, A. J., Zang, T., Muramatsu, H., Lin, P. J. C., Tam, Y. K., Mesin, L., Starr, T. N., Bieniasz, P. D., Pardi, N., Bloom, J. D., & Victora, G. D. (2023). Molecular fate-mapping of serum antibody responses to repeat immunization. *Nature*, *615*(7952), 482–489. https://doi.org/10.1038/s41586-023-05715-3

Schoeder, C. T., Schmitz, S., Adolf-Bryfogle, J., Sevy, A. M., Finn, J. A., Sauer, M. F., Bozhanova, N. G., Mueller, B. K., Sangha, A. K., Bonet, J., Sheehan, J. H., Kuenze, G., Marlow, B., Smith, S. T., Woods, H., Bender, B. J., Martina, C. E., del Alamo, D., Kodali, P., … Moretti, R. (2021). Modeling immunity with Rosetta: Methods for antibody and antigen design. *Biochemistry*, *60*(11), 825–846. https://doi.org/10.1021/acs.biochem.0c00912

Schubert, M. G., Goodman, D. B., Wannier, T. M., Kaur, D., Farzadfard, F., Lu, T. K., Shipman, S. L., & Church, G. M. (2021). High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proceedings of the National Academy of Sciences*, *118*(18), e2018181118. https://doi.org/10.1073/pnas.2018181118

Shin, J.-E., Riesselman, A. J., Kollasch, A. W., McMahon, C., Simon, E., Sander, C., Manglik, A., Kruse, A. C., & Marks, D. S. (2021). Protein design and variant prediction using autoregressive generative models. *Nature Communications*, *12*(1), 2403. https://doi.org/10.1038/s41467-021-22732-w

Sourisseau, M., Lawrence, D. J. P., Schwarz, M. C., Storrs, C. H., Veit, E. C., Bloom, J. D., & Evans, M. J. (2019). Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape. *Journal of Virology*, *93*(23). https://doi.org/10.1128/JVI.01291-19

Starr, T. N., Czudnochowski, N., Liu, Z., Zatta, F., Park, Y.-J., Addetia, A., Pinto, D., Beltramello, M., Hernandez, P., Greaney, A. J., Marzi, R., Glass, W. G., Zhang, I., Dingens, A. S., Bowen, J. E., Tortorici, M. A., Walls, A. C., Wojcechowskyj, J. A., De Marco, A., … Snell, G. (2021a). SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature*, *597*, 97–102. https://doi.org/10.1038/s41586-021-03807-6

Starr, T. N., Czudnochowski, N., Liu, Z., Zatta, F., Park, Y.-J., Addetia, A., Pinto, D., Beltramello, M., Hernandez, P., Greaney, A. J., Marzi, R., Glass, W. G., Zhang, I., Dingens, A. S., Bowen, J. E., Tortorici, M. A., Walls, A. C., Wojcechowskyj, J. A., De Marco, A., … Snell, G. (2021b). SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature*, *597*, 97–102. https://doi.org/10.1038/s41586-021-03807-6

Starr, T. N., Greaney, A. J., Addetia, A., Hannon, W. W., Choudhary, M. C., Dingens, A. S., Li, J. Z., & Bloom, J. D. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science*, *371*(6531), 850–854. https://doi.org/10.1126/science.abf9302

Starr, T. N., Greaney, A. J., Dingens, A. S., & Bloom, J. D. (2021). Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Reports Medicine*, *2*(4), 100255. https://doi.org/10.1016/j.xcrm.2021.100255

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J. D. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, *182*(5), 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012

Tareen, A., Ireland, W. T., Posfai, A., McCandlish, D. M., & Kinney, J. B. (2020). MAVE-NN: Quantitative Modeling of Genotype-Phenotype Maps as Information Bottlenecks. *bioRxiv*, 2020.07.14.201475. https://doi.org/10.1101/2020.07.14.201475

Tortorici, M. A., Czudnochowski, N., Starr, T. N., Marzi, R., Walls, A. C., Zatta, F., Bowen, J. E., Jaconi, S., Di Iulio, J., Wang, Z., De Marco, A., Zepeda, S. K., Pinto, D., Liu, Z., Beltramello, M., Bartha, I., Housley, M. P., Lempp, F. A., Rosen, L. E., … Pizzuto, M. S. (2021). Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature*, *597*, 103–108. https://doi.org/10.1038/s41586-021-03817-4

Traxlmayr, M. W., Hasenhindl, C., Hackl, M., Stadlmayr, G., Rybka, J. D., Borth, N., Grillari, J., Rüker, F., & Obinger, C. (2012). Construction of a stability landscape of the CH3 domain of human IgG1 by combining directed evolution with high throughput sequencing. *Journal of Molecular Biology*, *423*(3), 397–412. https://doi.org/10.1016/j.jmb.2012.07.017

Tsai, K.-C., Lee, Y.-C., & Tseng, T.-S. (2021). Comprehensive deep mutational scanning reveals the immune-escaping hotspots of SARS-CoV-2 receptor-binding domain targeting neutralizing antibodies. *Frontiers in Microbiology*, *12*, 1812. https://doi.org/10.3389/fmicb.2021.698365

Van Blarcom, T., Rossi, A., Foletti, D., Sundar, P., Pitts, S., Bee, C., Melton Witt, J., Melton, Z., Hasa-Moreno, A., Shaughnessy, L., Telman, D., Zhao, L., Cheung, W. L., Berka, J., Zhai, W., Strop, P., Chaparro-Riggers, J., Shelton, D. L., Pons, J., & Rajpal, A. (2015). Precise and efficient antibody epitope determination through library design, yeast display and next-Generation sequencing. *Journal of Molecular Biology*, *427*(6, Part B), 1513–1534. https://doi.org/10.1016/j.jmb.2014.09.020

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J., & Johnson, K. S. (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnology*, *14*(3), 309–314. https://doi.org/10.1038/nbt0396-309

Waldmeier, L., Hellmann, I., Gutknecht, C. K., Wolter, F. I., Cook, S. C., Reddy, S. T., Grawunder, U., & Beerli, R. R. (2016). Transpo-mAb display: Transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. *mAbs*, *8*(4), 726–740. https://doi.org/10.1080/19420862.2016.1160990

Wang, M., Cang, Z., & Wei, G.-W. (2020). A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation. *Nature Machine Intelligence*, *2*(2), 116–123. https://doi.org/10.1038/s42256-020-0149-6

Warszawski, S., Katz, A. B., Lipsh, R., Khmelnitsky, L., Nissan, G. B., Javitt, G., Dym, O., Unger, T., Knop, O., Albeck, S., Diskin, R., Fass, D., Sharon, M., & Fleishman, S. J. (2019). Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLOS Computational Biology*, *15*(8), e1007207. https://doi.org/10.1371/journal.pcbi.1007207

Wrenbeck, E. E., Klesmith, J. R., Stapleton, J. A., Adeniran, A., Tyo, K. E. J., & Whitehead, T. A. (2016). Plasmid-based one-pot saturation mutagenesis. *Nature Methods*, *13*(11), 928–930. https://doi.org/10.1038/nmeth.4029

Wu, N. C., Thompson, A. J., Lee, J. M., Su, W., Arlian, B. M., Xie, J., Lerner, R. A., Yen, H.-L., Bloom, J. D., & Wilson, I. A. (2020). Different genetic barriers for resistance to HA stem antibodies in influenza H3 and H1 viruses. *Science*, *368*(6497), 1335–1340. https://doi.org/10.1126/science.aaz5143

Wu, N. C., Xie, J., Zheng, T., Nycholat, C. M., Grande, G., Paulson, J. C., Lerner, R. A., & Wilson, I. A. (2017). Diversity of functionally permissive sequences in the receptor-binding site of influenza hemagglutinin. *Cell Host & Microbe*, *21*(6), 742-753.e8. https://doi.org/10.1016/j.chom.2017.05.011

Chapter Three
A Rapid Approach for Linear Epitope Vaccine Profiling Reveals Unexpected Epitope Tag Immunogenicity

Preface

Antigen epitope profiling is critical to the development of efficacious vaccine and mAb candidates. As alluded to in the previous chapter; DMS can provide a compelling in-depth map of epitope binding and escape patterns, but achieving conformational profiles is contingent on the complexity of both the antigen and variant library. Peptide tiling, wherein the antigen is displayed by a library of short overlapping peptides, is an effective alternative strategy for identifying linear epitopes via protein display. By incorporating oligo pool-derived library construction, phage display and Nanopore sequencing, we developed a rapidly deployable solution to linear epitope profiling. This chapter demonstrates the use of this method in identifying immunogenicity of a histidine tag present within a Group A *Streptococcus* vaccine candidate (TeeVax3).

The research is presented in this chapter as a published, peer-reviewed Research Article. The format has been modified to fit with the theme of this thesis. Supplementary material associated with this manuscript is presented in Appendix A.

Browne-Cole, K.*, **Hanning, K. R.*,** Beijerling, K., Rousseau, M., Loh, J. and Kelton, W. (2025). A rapid approach to for linear epitope vaccine profiling reveals unexpected epitope tag immunogencity. *Scientific Reports,* 15 (1). https://doi.org/ 10.1038/s41598-025-92928-3

Author contributions

K.BC. and K.R.H. contributed equally to this work. K.R.H., J.L., and W.K. conceived of and designed the study. K.BC., K.R.H., K.B., and M.R. performed laboratory experiments and created reagents used in the study. K.R.H. and K.BC. performed the Nanopore sequencing and

bioinformatic analysis. K.R.H., K.BC., and W.K. wrote the manuscript. All authors provided feedback and revisions for the final version of the manuscript.

A rapid approach for linear epitope vaccine profiling reveals unexpected epitope tag immunogenicity

Kirsten Browne-Cole[1§], Kyrin R. Hanning[2§], Kevin Beijerling[2], Meghan Rousseau[1], Jacelyn Loh[3,4], William Kelton[1,2*]

[1]Te Aka Mātuatua School of Science, University of Waikato, Hamilton, New Zealand
[2]School of Pharmacy and Biomedical Science, University of Waikato, Hamilton, New Zealand
[3]Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand
[4]Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand

Abstract

Antibody epitope profiling is essential for assessing the robustness of vaccine-induced immune responses, particularly while in development. Despite advancements in computational tools, high throughput experimental epitope validation remains an important step. Here, we describe a readily accessible method for rapid linear epitope profiling using phage-displayed oligo pools in combination with Nanopore deep sequencing. We applied this approach to TeeVax3, a Group A Streptococcus vaccine candidate, to investigate the antibody response generated in a pre-clinical rabbit model and assess antigen immunogenicity. Surprisingly, we found a strong bias in antibody binding response towards the N-terminal epitope tag used for purification. These tags are widely reported to have low immunogenicity and are frequently left uncleaved in pre-clinical studies. We further confirmed that the observed immune response against the epitope tag dominated even the conformational binding response and, using synthetic peptides, narrowed the epitope down to a set of 10 residues inclusive of the Histidine residues. Our findings highlight the importance of epitope-tag removal in pre-clinical studies and demonstrate the utility of rapid nanopore sequencing for early-stage vaccine evaluation.

Introduction

Vaccine protection is strongly correlated with the quality and longevity of B and T cell responses induced against a given pathogen (Larocca et al., 2016; Liu Yihao et al., 2022; McNamara et al., 2020). While T cell responses are driven by MHC loading of linear peptide epitopes present in vaccine antigens, antibodies can engage targets directly by binding to either linear epitopes or conformational epitopes formed when antigenic moieties are brought into close proximity by three-dimensional folding. Understanding the epitope distribution and binding affinity of the antibody response provides invaluable information for vaccine development efforts, especially where responses towards specific neutralizing epitopes are desired (Cao et al., 2022; Sevvana & Kuhn, 2020). Alongside predictive computational tools (Cia et al., 2023), several high-throughput wet lab approaches are widely used to profile the antigenic epitopes driving antibody responses (De Leon et al., 2024; Hu & Irving, 2023), particularly for applications in virology. These approaches can be adapted more widely to vaccine profiling (Mohan et al., 2018).

Crystallography remains the highest accuracy technique for antibody epitope mapping, followed by more recently developed Cryo-EM approaches (Antanasijevic et al., 2022). Unfortunately, throughput is a challenge for these methods, and capturing a comprehensive snapshot of an antibody response against an antigen is very resource intensive. Instead, protein display systems have gained popularity as millions of individual epitope sequences can be physically linked to an underlying genotype that can be recovered by next-generation sequencing. Antigens can be expressed as full-length proteins or tiled as short peptides. The VirScan approach used phage display to present more than 1 x 108 unique linear peptides from human viruses for capture by sera-derived antibodies from a global cohort of individuals (Shrock et al., 2022; Xu et al., 2015). In addition to antigen tiling, peptide sequences can also be systematically mutated using techniques like deep mutational scanning to provide greater precision when defining antibody epitopes (Garrett et al., 2020; Hanning et al., 2022). The use of next-generation sequencing is particularly advantageous as it allows the detection of rare binding events at low frequencies, important when defining the binding ranges of polyclonal antibodies.

Here, we used a peptide phage display approach to investigate antibody responses elicited by TeeVax3 (Loh et al., 2021), a component of the multivalent vaccine candidate under

development for Group A Streptococcus (Strep A) infection. Despite longstanding development campaigns by academic and industry groups (Fan et al., 2024), no vaccines have yet been approved for this pathogen, in part because of large strain diversity and the risk of autoimmune induction by surface antigens (e.g. M protein). By instead concatenating sets of five to seven T antigens from diverse strains into three recombinant proteins (e.g. TeeVax3: T6M-T2M-T25M-T23M-T4N), protective immunity was induced in small animal models of Strep A infection (Loh et al., 2021). Further work characterizing the antibody response in mice and rabbits used phage-displayed Fab immune libraries to isolate and crystallize high-affinity Fab domains against the T antigen (Raynes et al., 2023). High-resolution epitope mapping of antibody repertoires is key to understanding the drivers of vaccine efficacy. Surprisingly, there appeared to be immunodominant conformational epitopes towards the N-terminus of some T antigens with unusual temperature-dependent accessibility at 37 °C.

To better understand antibody biases and binding distribution on the TeeVax3 antigen in high throughput, we created a bacteriophage-based method for the rapid profiling of linear epitopes using Nanopore sequencing. Applying our method to sera from TeeVax3 immunized rabbits revealed a strong bias towards the N-terminal TEV/6-His tag sequence that dominated even the conformation antibody binding response. While epitope tags are routinely removed from vaccines, including most but not all TeeVax3 animal experiments (Loh et al., 2021; Raynes et al., 2023), conventional wisdom suggests these peptides should possess low immunogenicity (López-Laguna et al., 2022; Zhao et al., 2013). In many cases, these tags are left on vaccine antigens, especially through the early stages of in vivo profiling in animal models (Otsyula et al., 2013; Rodrigues et al., 2020; Zang et al., 2021). As well as providing an updated set of tools for linear epitope profiling, our finding provides a cautionary tale for those developing antigen-based vaccines requiring exogenous peptide tags for purification or antigen tracking.

Methods

Antigen purification

The TeeVax3 antigen was expressed in *Escherichia coli* by adapting previously described methods (Loh et al., 2021). Briefly, pROEXhtb vectors encoding TeeVax3 were transformed

into BL21 (DE3) *E. coli* cells and cultures in exponential growth phase induced with 0.1 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG). After overnight growth at 18 °C for 18 hours, cell pellets were resuspended in Lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 20 mM imidazole), sonicated using a Qsonica Q700 with a 417-A microtip (Total time: three mins, pulse-on time: one sec; pulse-off time: one sec), and centrifuged at 14,000 g for 30 min at 4 °C to recover the supernatant. His tagged antigens were purified by Ni-NTA chromatography using 5 mL HisTrap HP columns (Cytiva) on a BioRad NGC FPLC. Non-specifically bound proteins were removed with 50 mM Tris-Cl pH 8, 150 mM NaCl, 20 mM imidazole and the antigen eluted with 50 mM Tris-Cl pH 8, 150 mM NaCl, 1 M imidazole. Amicon Ultra-15 10 kDa centrifugal filters (Merck Millipore) were used to concentrate the protein before size exclusion polishing with a S75 16/60 column (GE Healthcare) using 50 mM Tris pH 8, 150 mM NaCl. Samples were buffer exchanged into 1× Phosphate Buffered Saline (PBS) for storage and analysed by 12% SDS-PAGE gel electrophoresis. For His tag removal, TeeVax3 antigen was buffer exchanged into 20 mM Tris pH 7.5 and digested overnight at 4 °C with TEV Protease (NEB). Undigested antigen and digested His tags were captured by Ni-NTA chromatography as described above, taking the flowthrough fraction as the untagged antigen.

Animal experiments

All animal experiments were approved by The University of Auckland Animal Ethics Committee (Approval #R1663) and were performed in the Vernon Jansen Unit at The University of Auckland. All methods and procedures were performed according to the regulations and standards of the University of Auckland. Experimental methods are reported in accordance with the ARRIVE guidelines. TeeVax3 immunized rabbit serum was obtained from a New Zealand white rabbit immunized subcutaneously with 100 μg of recombinant protein emulsified 1:1 with incomplete Freund's adjuvant and boosted at 2 and 4 weeks. Control serum was obtained in parallel from an unimmunized individual. All animals were deeply anaesthetized at day 42 with 90 mg/kg of intravenous sodium pentobarbitone, and antiserum was collected by terminal blood draw.

Antibody purification

Immunoprecipitation of TeeVax3-specific antibodies was performed by coupling TeeVax3 antigen to AminoLink Plus Coupling Resin (ThermoFisher). Eight mg of TeeVax3 antigen in 0.1 M sodium phosphate, 0.15 M NaCl, pH 7.2 coupling buffer were incubated with 2 mL of

resin at 4°C overnight with gentle rotation. The following day, 5 mL of pH 7.2 PBS and 100 µL of Cyanoborohydride Solution were added and the mixture rotated for four hours at room temperature. Two mL of conjugated resin was added to a gravity flow column and sera diluted five-fold in PBS was passed through the column. The bound fraction was washed with PBS, eluted with 0.1 M glycine pH 3, and neutralized with 10% the solution volume of 1M Tris pH 9.0. Amicon Ultra-15 10 kDa centrifugal filters (Merck Millipore) were used for buffer exchange into PBS prior to purity analysis by SDS PAGE. Negative control antibodies from unimmunized rabbits were purified by Protein G chromatography. Sera from these animals was again diluted more than five-fold in PBS and passed through a 1 mL Protein G column (Cytiva). After washing with >10 column volumes of PBS, the bound antibody pool was eluted with Glycine HCl pH 3.0, neutralized with 1M Tris pH 9.0 and buffer exchanged into PBS following the method used for anti-TeeVax3 antibodies.

Antigen tiling

The tile library was created computationally by splitting the 941 amino acid sequence of the Teevax3 antigen (Supplementary Table S1) into 308 segments each 20 amino acids long. These peptides were offset by three amino acids to ensure complete coverage of potential binding sites. Bioinformatics software (Geneious Prime) was then used to back translate the tiles into nucleotide sequences, followed by the addition of adaptors containing a coding sequence, peptide transport signal and restriction sites to both 5' and 3' ends (5' adaptor CGAGGGCCCAGCCGGCCATGGCCGAGGGT, 3' adaptor TCGGCCTCGGGGGCCA). The resulting 308 tiles of 105 nucleotides each were then ordered as a single oligo pool (Twist Bioscience). In parallel, we designed control oligo nucleotides flanked by the 5' and 3' adapters encoding the PEP-1 cyclic peptide (TGCGGTCCGATCCCAGTGCTGGATGAAAACGGCCTGTTTGCTCCGGGTCCGTGC) and 6-His epitopes (CATCACCATCACCATCACGACTACAAAGACGATGACGACAAG).

Phage construction and peptide expression

Oligos were duplexed by a single cycle of PCR using OneTaq® Quick-Load® 2X Master Mix with Standard Buffer (NEB). A reverse primer (5'-TGGCCCCCGAGGCC-3') was included at 0.2 µM in the reaction alongside 12 ng of oligo template and the following conditions were used: denaturation at 94 °C for one min, one cycle of elongation at 60 °C for one min, followed by final extension for five min at 68 °C. The double-stranded oligo pool was digested with SfiI restriction enzyme for two hours at 50 °C. Digested fragments were cleaned using a QIAquick

nucleotide removal kit (Qiagen) and DNA concentration was quantified using a DeNovix double-stranded Ultra sensitivity assay. The library was then ligated into SfiI digested pAK200 vector and bound to Monarch columns (NEB) after the addition of five volumes of 5 M Guanidine Hydrochloride containing 30% isopropanol. The sample was then washed with 10 mM Tris-HCl pH 7.5, 80% ethanol and eluted with ultra-pure water. Ligated plasmids were transformed into electrocompetent SS320 cells and plated to create the final library. Library glycerol stocks were grown at 37 °C from OD_{600} 0.1 to 0.4 - 0.6 in LB media supplemented with 20 µg/mL chloramphenicol. M13K07 helper phage were added at an MOI of 20x and the culture incubated without shaking for one hour to allow infection. Kanamycin antibiotic was added (35 ug/mL) and peptide expression was induced with 1 mM IPTG before overnight growth to produce phage. Culture supernatants were filtered through 0.22 µm filters and phage were precipitated by adding 20% of the total volume of 20% PEG/2.5 M NaCl and incubating on ice for at least 30 min. Phage were centrifuged at 16000 g, resuspended in 1x PBS and spun again twice more at 16,000 g changing the tube between spins. Titers were quantified by spectrophotometry (DeNovix DS11) using the formula $virions \bullet mL^{-1} = \frac{(A_{269}-A_{320}) \times 6 \times 10^{16}}{(phage\ ssDNA\ bases)}$ (Smith & Scott, 1993).

Control peptide phage ELISA analyses

ELISA plates were coated with 50 µL volumes of Sodium Carbonate buffer pH 9.6 (34 mM NaHCO3, 16 mM Na2CO3) containing 4 µg/mL 2C7 (produced in house) or 2 µg/mL anti-His (Thermo Fisher Scientific MA1-21315) overnight at 4 °C. For ELISAs to measure background antibody binding, the plates were coated with 4 µg/mL of polyclonal antibodies isolated from rabbit serum alongside 2 µg/mL anti-His as a positive control. The following day the plates were blocked with 360 µL of PBSMT (PBS containing 2% skim milk powder and 0.05% Tween20) for 30 min at room temperature. Control phage expressing PEP-1 or 6-His peptides were serially diluted in PBSM (PBS containing 2% skim milk powder) and incubated for 30 min at room temperature. The plates were washed three times with PBST (PBS containing 0.05% Tween20) and 50 µL of anti-M13 HRP diluted 1:20,000 (Sino Biological MM05T) was added in PBSM. After 30 min of incubation the plates were washed a further three times before the addition of 50 µL 1-Step™ Ultra TMB-ELISA (ThermoFisher) for less than ten min. The reaction was quenched with 50 µL 1 M H2SO4 and the plates read at 450 nm on a Spectramax M4. All ELISAs were repeated in triplicate.

Phage panning

Phage panning was carried out in ELISA 8 well strips (Corning). The wells were coated overnight at 4°C with 100 µL of purified anti-TeeVax3 antibodies at 4 µg/mL in 50 mM Na2CO3 buffer at pH 9.5 or incubated with Na2CO3 buffer alone. The next day, the solutions were removed, and the plates were blocked with 300 µL PBSMT. After 1 hour of shaking at room temperature, the strips were washed 3× with PBST. 1×10^{11} phage per well were then added for 1 hour to the strips coated with buffer alone for non-specific phage binding depletion before transfer to the wells coated with antigen. The wells were washed again after one hour, although the number of washes was increased by three each round to increase selection pressure. Phages were eluted with a ten min incubation using 100 µL of 0.1 M glycine/HCl pH 2.7, transferred to a 2 mL tube previously blocked with PBS containing 2% skim milk powder, and neutralized with 1 M Tris pH 8.0 at a 1:5 ratio (vol Tris/vol phage). 20 mL SS320 E. coli cells were grown in 2xYT media and 2% glucose medium to an OD_{600} 0.4-0.6 at 37°C without shaking, and incubated with the eluted phages for 30 min at 30 °C. The shaking speed was then increased to 250 rpm for another 30 min. The phage library was recovered by plating infected cells on three 145 mm round LB agar plates containing 2% glucose and 20 µg/mL chloramphenicol and overnight growth at 37°C.

Nanopore sequencing and analysis

Library plasmid stocks from each screening phase were directly purified from corresponding E. coli glycerol stocks using via QIAprep Spin Miniprep Kit (Qiagen, Germany). Linear, 723 bp DNA fragments representing the tile library CDS were gel excised and purified with a QIAquick Gel Extraction Kit (Qiagen) following restriction digest of each library plasmid stock with BamHI-HF and XbaI (NEB, USA). Each purified library fragment was then prepared according to the SQK-NBD114.24 Oxford Nanopore Technologies (ONT, UK) ligation sequencing amplicons protocol and kit (200 fmol input), and subsequently sequenced using a R10.4.1 flow cell (ONT). Sequencing was stopped once the estimated raw read count reached ≥100x coverage per tile per sample. Base-calling and demultiplexing of raw read data was performed using Dorado (v0.5.3; model dna_r10.4.1_e8.2_400bps_sup@4.3.0). Read data were filtered for quality control using fastp (v0.23.4) (Chen, 2023) and binned into tile counts via local alignment against the 308 segments using minimap2 (v2.24) (Li, 2021) and samtools (v1.19.2) (Danecek et al., 2021).

ELISA analysis for TeeVax3 antibody binding

ELISA analysis was used to compare antibody binding to TeeVax3 with and without the His epitope tag. Antigens were coated overnight at 4 °C on ELISA plates (Corning 3590) at 4 µg/ml in Coating Buffer (50 mM NaHCO3, pH 9.5), washed the next day three times with PBST, and blocked with PBSMT. Blocked plates were washed a further three times with PBS Tween20. Serial dilutions of TeeVax3-specific antibodies isolated from rabbit sera were added to the plate, and a goat anti-rabbit IgG HRP antibody (Abcam) was used at 1:50,000 for detection of binding. Control samples confirming His tag removal from the TeeVax3 antigen used a HRP conjugated anti-polyHistidine antibody [HIS-1, Abcam] diluted 1:20,000 in PBSM. 50 µl 1-Step™ Ultra TMB-ELISA was used for detection of HRP signal in each well and the reaction quenched after 10-15 min with an equal volume of 1 M H2SO4 before reading at 450 nm. All ELISA steps were undertaken at room temperature with incubation periods of 1 hr unless otherwise indicated. Fine epitope mapping was performed by ordering synthetic peptides with N-terminal biotin residues (Genscript: 17-mer YHHHHHHDYDIPTTENL, 14-mer YHHHHHHDYDIPTT, 13-mer HHHHHHDYDIPT, 10-mer HHHHHHDYDI, 7-mer YHHHHHH). ELISA analysis followed the protocol above except the ELISA plates were coated with biotinylated peptides that had been precomplexed with neutravidin at a 4:1 molar ratio for 30 min at room temperature in 50 mM NaHCO3, pH 9.5. ELISA data was fit by non-linear regression of a reparametrized 5 parameter logistic model in JupyterLab (Liao & Liu, 2009). All ELISAs were repeated in triplicate. Pairwise t-tests were used to compare data points at each antibody concentration and a Bonferroni correction was applied to account for multiple comparisons.

Results

Construction of tiled TeeVax peptide libraries for phage display

To profile TeeVax3 linear epitopes in high throughput, we developed a phage-based peptide display approach to make use of rapid Nanopore sequencing (Fig. 3.1a). We first cloned short control peptides encoding the 6-His tag and a cyclic peptide mimetic of lipooligosaccharide derived from the pathogen *Neisseria gonorrhea* (PEP-1) (Ngampasutadol et al., 2006) into pAK200 phage display vectors (Krebber et al., 1997). These peptides were selected as high-affinity antibodies targeting them are readily available. ELISA analysis confirmed control phage binding to both anti-His (HIS.H8) anpakd PEP-1 (mAb 2C7) (Gulati et al., 1996)

monoclonal antibodies above background helper phage binding (Fig. 3.1b). Having demonstrated the suitability of this system for the accessible display of short peptides, we next created a library of peptides from the 940 amino acid TeeVax3 antigen. Bioinformatic tiling was used to generate 308 sequences, each 20 amino acids in length and offset by three amino acids, except for a final tile with a two amino acid offset. To confirm the suitability of nanopore sequencing for tile identification, we further calculated the Levenstein distances between all tiles in the library (Fig. S1). We found a minimum edit distance of 16 within the library, corresponding to a required sequencing error rate of 26.3% for misidentification during binning. All tile sequences were back translated, adapter sequences added to the 5' and 3' ends, and the resulting library ordered as an oligo pool (Twist Biosciences). A single cycle of PCR was used to create the second strand of DNA for restriction enzyme cloning into pAK200 and phage production.

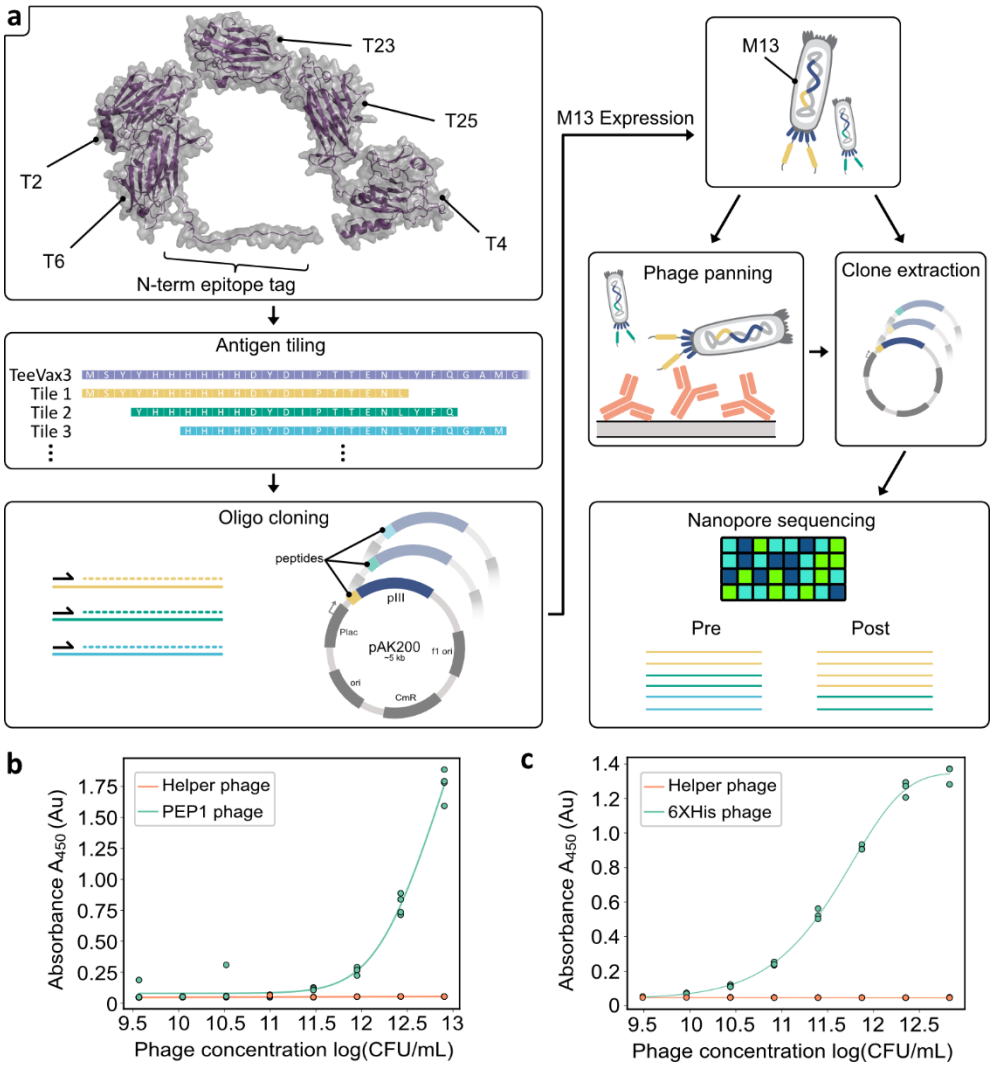


Figure 3.1. Linear antibody epitope profiling using phage panning and Nanopore sequencing. (a) Overall experimental scheme. Target antigen sequences are tiled into discrete peptides and encoded as an oligonucleotide pool. After a single cycle of PCR to create the complementary strands, these sequences are cloned into pAK200 plasmids and expressed on the pIII domain of M13K07 bacteriophage. Peptide tile abundances before and after panning against the vaccine-specific antibody pool are determined by Nanopore sequencing and the ratio of these frequencies used to determine epitope enrichment. ELISA analysis demonstrated specific binding of control phage expressing model PEP-1 cyclic peptides to the 2C7 monoclonal antibody (n = 4) **(b)** and phage displaying a 6-His epitope tag peptide to the HIS.H8 monoclonal antibody (n = 3) **(c)**. Helper phage with no peptide displayed were used as the negative control. Representative plots from independent replicates are shown with each point representing an individual datapoint.

Characterization of TeeVax3 peptide libraries by Nanopore sequencing

Prior to phage epitope profiling, we first confirmed the cloned library coverage using Nanopore sequencing. Plasmid DNA was isolated and cloned amplicons were removed by restriction enzyme digest to eliminate PCR amplification bias. We obtained ≥1.4 million reads from the

library using a Q score cutoff of 15 (Fig. S2a) with a median fragment length of 720 bp corresponding to the expected size for the digested fragment (Fig. S2b). Reads were then binned into tiles via Minimap2. We again observed that 100% of the expected tiles were present, although individual tile frequencies within this pool ranged from 0.07% to 0.57% (Fig. 3.2a). Bacteriophage were then produced and immediately used to reinfect cells to gather the fraction of the library amenable to expression by phage. Clonal dropout at this step would lead to gaps in epitope coverage; therefore, nanopore sequencing was again used to re-evaluate tile frequencies. We again sequenced to a depth of ≥1.4 million reads and saw 100% retention of tile coverage in the phage library. Here, we observed minor differences in tile frequency and a general broadening of the tile abundance distribution when compared to Round 0 (Pre-expression plasmid sequence pool). Individual tile frequencies ranged from 0.036% to 1.82% indicating some expression biases but this was not concentrated to any one region of the antigen (Fig. 3.2b).

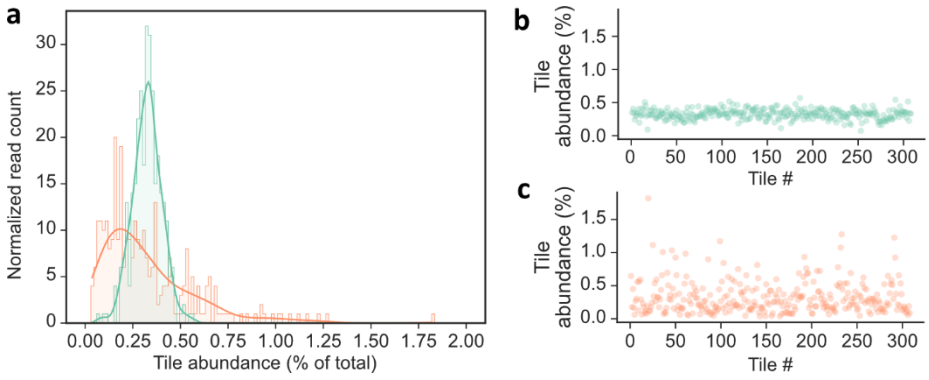


Figure 3.2. Pre-panning peptide library evaluation by Nanopore Sequencing. (a) Distribution of individual tile abundances before (green) and after (orange) expression on phage, normalized to the total read count. Solid lines represent smoothed distributions generated by kernel density estimations of the underlying histogram plots. To determine whether expression bias is distributed evenly throughout the antigen, individual tile abundances were plotted against tile location within the antigen (Tile #) for (b) Round 0 pre-expression (green) and **(c)** Round 0 reinfected post-expression phage (orange).

Linear epitope profiling by multi-round phage display reveals significant binding to the N-terminal epitope tag

To enrich antigen-specific clones from the antibody repertoire of rabbits immunized with TeeVax3, we first purified a TeeVax3-specific pool of antibodies by immunoprecipitation. High-purity TeeVax3 antigen (Fig. 3.3a, Figure S3a) was conjugated to agarose beads to enable the isolation of both linear and conformational TeeVax3 epitope-binding clones. This process yielded approximately 1.3 mg of TeeVax3-specific antibodies per mL of immunized rabbit sera. SDS PAGE confirmed a high purity of the resulting antibody fraction with multiple bands

representing heavy and light chains from multiple antibody subclasses (Fig. 3.3b, Fig. S3b). Antibodies were also isolated from unimmunized sera by Protein G chromatography. Together, we used these polyclonal antibody pools to evaluate background binding to the TeeVax3 library phage (Fig. 3.3c). Polyclonal antibodies enriched for binding to TeeVax3 had 4.7-fold higher signal than the polyclonal pool from unimmunized individuals. However, the unimmunized pool did show low but consistent binding above background (uncoated control) of 1.3-fold. We

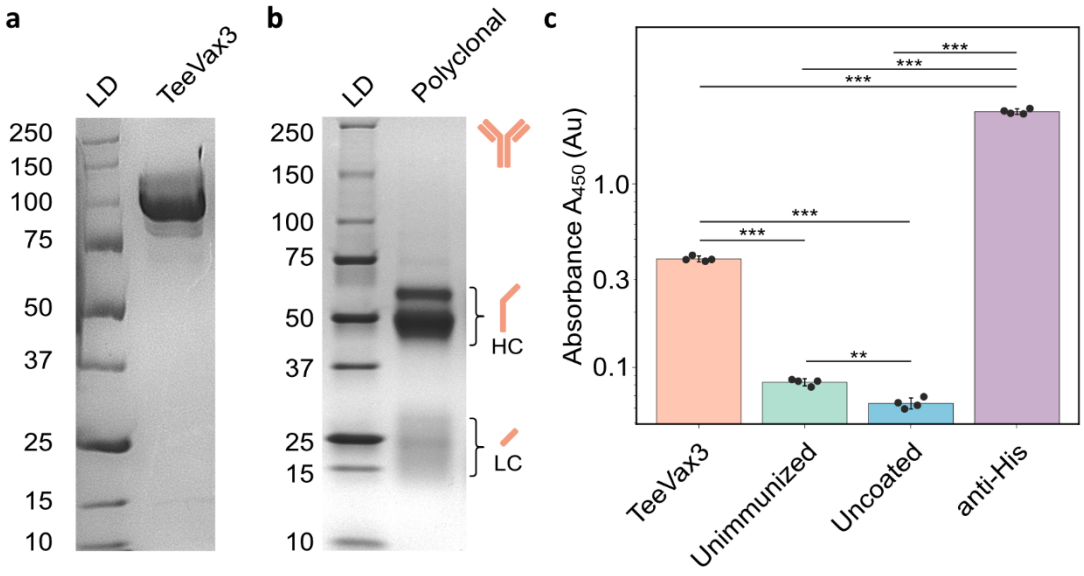


Figure 3.3. Expression of TeeVax3 antigen and isolation of vaccine-specific antibodies. (a) TeeVax3 was expressed in E. coli, purified via Ni-NTA affinity chromatography and run on SDS-PAGE. **(b)** A custom chromatographic resin was created by conjugating TeeVax3 antigen to aldehyde-activated agarose beads and used to purify polyclonal antibodies from TeeVax3-immunized rabbits. Separation by reducing SDS-PAGE gel shows the various heavy chain and light chain sizes expected from a diversity of clonal sequences and antibody subclasses. **(c)** ELISA analysis of background binding of polyclonal antibodies to phage libraries. The plates were coated with polyclonal anti-TeeVax3, antibodies isolated from unimmunized serum by protein G chromatography, and anti-His (HIS.H8) as a positive control. Background signal was evaluated by measuring binding to plates blocked with PBS containing 2% w/v skim milk powder. Data shown is a representative plot with individual datapoints as each well in the plate (n=4). Pairwise t-tests with a Bonferroni correction applied for multiple comparisons were used to assess statistical significance (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

then used the TeeVax3-specificantibodies directly for phage panning of the TeeVax3 linear epitope library. We performed three rounds of panning and used Nanopore Sequencing to track the clonal enrichment of each tile. Enrichment scores were determined by calculating the ratio of observed tile frequencies from the pre-panned phage library and post-panned libraries.

Over 64,000 high-quality sequences were obtained from phages after each round of panning, representing a greater than 200-fold coverage of the theoretical library diversity. We observed

a rapid bias in representation towards tiles from the N-terminus of the TeeVax3 antigen corresponding to the 6-His epitope tag and TEV cleavage sequences (Fig. 3.4a). By round two of panning, the phage library sequence pool was dominated by 81.1% of tile 1 (MSYYHHHHHHDYDIPTTENL) and 9.5% of tile 2 (YHHHHHHDYDIPTTENLYFQ). Little further enrichment was observed in the third round of panning. We next explored the likelihood these sequences were the result of expression bias by comparing the enrichment of these tiles between the Round 0 pre-expression plasmid library and Round 0 reinfected post-expression frequencies. We found that while the proportion of tile 1 increased from 0.347% to 0.645% (1.9-fold), indicating a positive expression bias, tile 2 dropped in abundance from 0.404% to 0.176% (0.44-fold) indicating expression of this sequence was relatively disfavored. In each case, the >10-fold enrichment of these sequences over the first round of panning far exceeds expression bias contributions. More broadly, we observed a decrease in the median tile frequency from 0.328% to 0.260% across all the tiles.

To further investigate clonal bias outside of the N-terminal sequences, we expanded the analysis between the Round 0 reinfected library and Round 1 of panning to include tiles with enrichment in the 95th percentile exclusive of tiles 1-3. These enriched peptides were mapped onto a three-dimensional representative model of the TeeVax3 antigen, generated using AlphaFold2 (Jumper et al., 2021) in the absence of crystal structure data (Fig. S4). This process revealed a lack of antibody binding preference for the T2M domain, T25M domain, and the inter-domain spacer regions, except for the linker between the T6M and T2M domains. Instead, we observed some preference for T6M, T23M, and increasingly T4N binding where a high proportion of overlapping tiles were mapped (Fig. 3.4b).

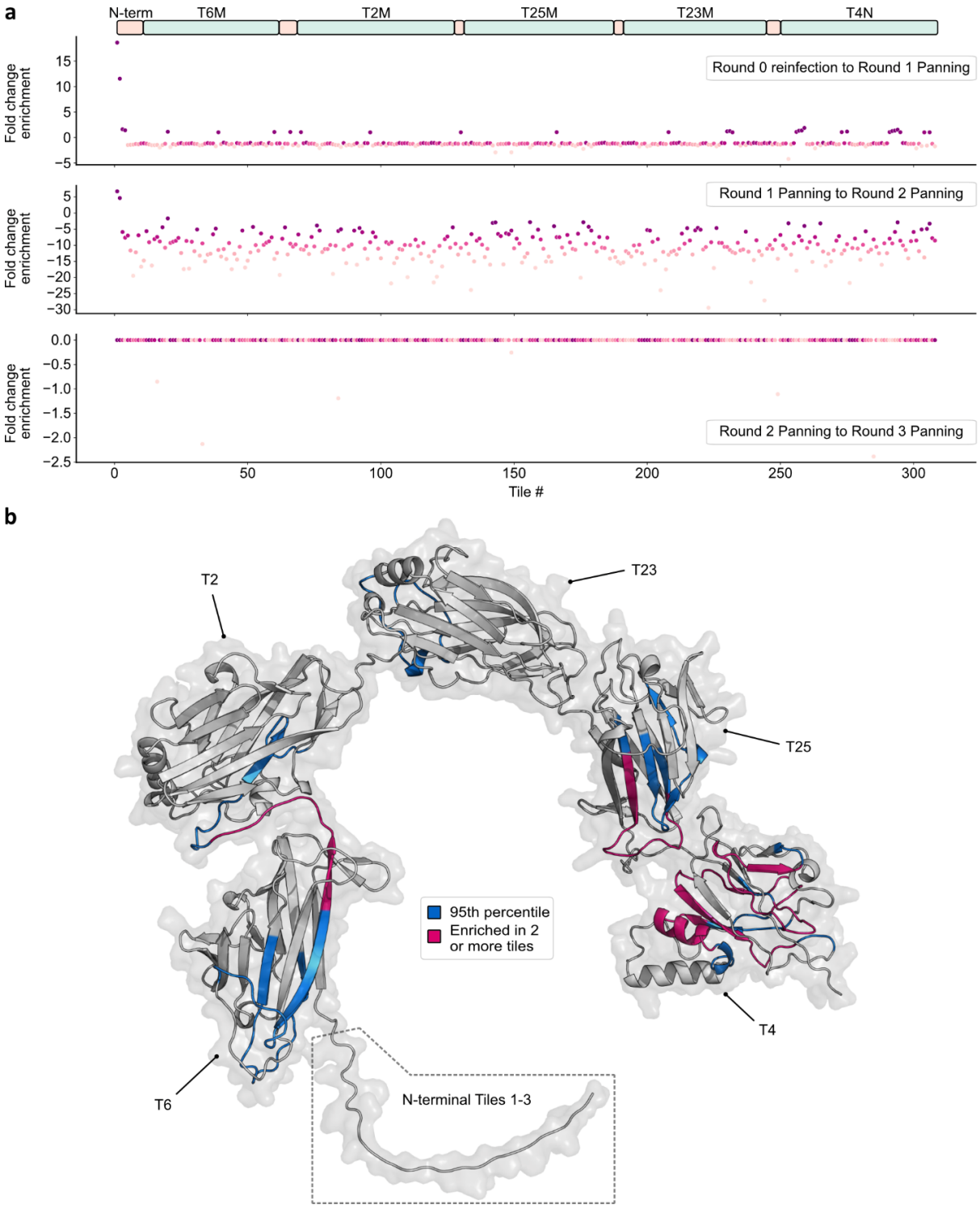


Figure 3.4. Tile enrichment during panning against TeeVax3-specific antibody pool. (a) The TeeVax3 peptide pool was expressed on bacteriophage and panned against immobilized anti-TeeVax3 polyclonal antibodies. Fold enrichment values for each peptide tile were computed by dividing the post-panning frequency by the corresponding pre-panning frequency following Nanopore sequencing. Each data point represents an individual peptide tile sequence. **(b)** After excluding high abundance N-terminal tiles 1-3 as indicated in the dashed box, tiles above the 95th percentile were mapped to a model of TeeVax3. Blue indicates tiles above the 95th percentile and magenta shows sequences represented in two or more tiles.

High-resolution epitope profiling defines a high-affinity peptide within the epitope tag

There remained the possibility that the unexpected binding we observed for the epitope tag could be a small fraction of the total antibody pool with most of the response being dominated by conformational epitope binders rather than linear epitope binders. To measure the relative abundance of linear epitope tag binders to all TeeVax3 antigen binders, against both conformational and linear epitopes, we cleaved the His tag from TeeVax3 using TEV protease and confirmed the absence of the His/TEV tag by ELISA using anti-His HRP for detection (Fig. 3.5a). ELISA analysis was then used to measure the relative anti-TeeVax3 antibody signal from purified sera samples with and without the His/TEV tag. We observed a decrease in antibody EC50 (6.3-fold) upon removal of the epitope tag indicating the binders to Tiles 1 and 2 dominated the pool of antibodies against the antigen even with conformational binding clones present (Fig. 3.5b).

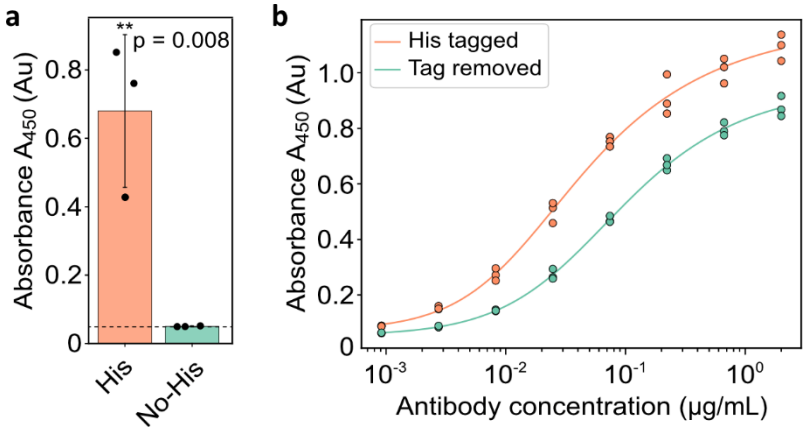


Figure 3.5. Antibody response with and without the N-terminal epitope tag sequence. (a) TeeVax3 antigens were digested with TEV protease and the presence of residual Histidine epitope tag was assayed by single-well ELISA relative to the undigested antigen. The dashed line indicates background binding in the assay. Data are presented as the mean ± SD (n = 3) and a pairwise t-test was performed to calculate the p-value (p = 0.008). **(b)** Representative ELISA plot showing polyclonal antibody responses towards TeeVax3 antigen with and without the epitope tag removed (n = 3). Each point represents an individual datapoint.

The unexpected immunogenicity of the epitope tag led us to profile the specific amino acids driving the antibody response with greater precision. We ordered a series of synthetic biotinylated peptides of decreasing size beginning with the 17 amino acids overlapping in Tiles 1 and 2 that provided the greatest enrichment by phage panning. Subsequent peptides reduced the size of this epitope with the smallest fragment tested consisting of a 7 amino acid sequence encompassing the 6-His tag (Fig. 3.6a). The peptides were captured with neutravidin and TeeVax3-specific antibodies were assayed for binding. We found the highest relative signal

with the longest 17 amino acid epitope and a trend of similar binding for each of the 14-mer, 13-mer, and 10-mer tiles (Fig. 3.6b).

A substantial drop in signal was observed for the shortest tile covering the 6-His tag with an additional Tyrosine motif. This dataset suggests the strongly immunogenic motif we observe encompasses both residues from the Histidine epitope tag and the TEV linker regions.

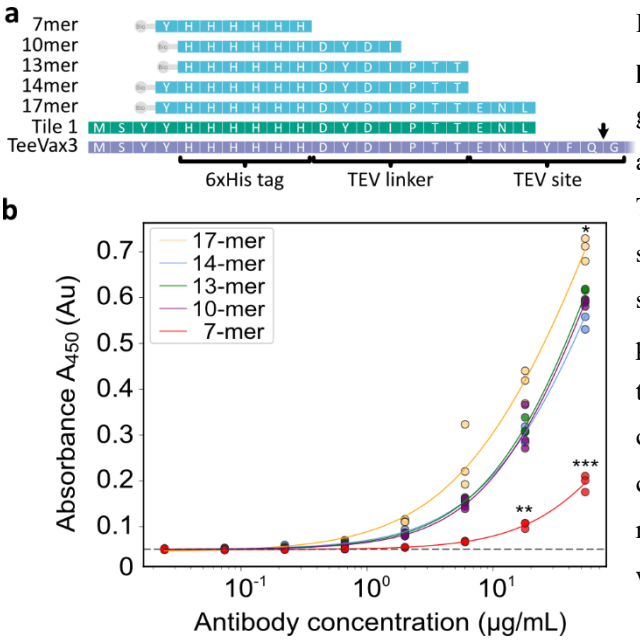


Figure 3.6 (left). Fine epitope profiling with synthetic peptides. (a) Peptide series with N-terminal biotin fusions generated from the Tile 1 TeeVax3 sequence. **(b)** ELISA analysis measuring the binding capacity of polyclonal TeeVax3-specific antibodies against biotinylated peptide series. The dashed line represents the average background signal obtained from negative wells (n = 4) where the polyclonal TeeVex3 antibodies were omitted. Pairwise t-tests were used for pairwise comparison at each concentration with Bonferroni correction for multiple comparisons (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). A representative plot from independent replicates is shown with each point representing an individual datapoint.

Discussion

High-throughput epitope mapping tools provide a rapid snapshot of antibody binding bias against vaccine or pathogen antigens. In this study, we explored the use of Nanopore sequencing to rapidly profile linear epitope antibody binders derived from rabbit sera against the TeeVax3 candidate vaccine for Group A Streptococcus. This sequencing approach can have higher error rates than other next-generation sequencing approaches, such as Illumina, but we reasoned that the unique offset nature of the peptide tiles enables identification regardless of sequencing error. To develop the method, an oligo pool was ordered and used to create a library of 20-mer peptides derived from the TeeVax3 antigen. We then used a multi-round phage panning approach against enriched TeeVax3-specific polyclonal antibodies to monitor the enrichment of the expressed tiled peptides. After three rounds of panning, the resulting libraries were dominated by peptide tiles stemming from the N-terminus. We were surprised to find a strong enrichment of peptide sequences encoding the 6 His/TEV purification tag and therefore

evaluated the relative binding signal of antibodies against the TeeVax3 antigen with and without epitope tag cleavage. This analysis confirmed the polyclonal pool was dominated by linear binders to sequences from the epitope tag and confirmed the unexpected immunogenicity observed by phage panning. To further understand the scope of the immunogenic epitope, we ordered a series of biotinylated peptides of decreasing size and observed binding even for the 6-His tag with one additional residue. Our findings indicate the residues displaying the strongest immunogenicity lie within a 10 amino acid sequence covering the 6-His epitope tag and the N-terminus of the TEV linker sequence.

Prior studies have used sequencing approaches with very low error rates, such as Illumina (Paull et al., 2021; Qi et al., 2021) for linear epitope binding antibody mapping. The accessibility and speed of Nanopore allows for in-laboratory sequencing even if core facilities or next-generation sequencing providers are not readily available. Despite a high error rate, we found the sequence distance generated by offset peptide tiles was sufficient for accurate binning and determination of Nanopore read counts required to calculate enrichment factors. While we have used only a limited set of peptide tiles here (308) relative to other examples in literature (Rajan et al., 2021; Yaffe et al., 2023), the length of 20 amino acids means it is very unlikely sequencing errors could cause binning errors in larger libraries, with the exception of antigens with repeats or regions of high sequence homology. The finding that most antibody binders were directed to the epitope tag was unexpected, especially considering that His tags are typically considered to have low immunogenicity. While there are reports of His tags influencing antigen immunogenicity (Lin et al., 2022; Singh et al., 2020), raising antibodies against His tags frequently requires conjugation to carrier proteins to elicit sufficient antibody titers (Przedpelski et al., 2020). It should be noted the strongest antibody binding signal we observed did require residues from the TEV cleavage site in addition to those from the 6-His tag. An obvious mitigation strategy is to remove the His tag via TEV cleavage, but this step is inconsistently performed in preclinical vaccine studies. The generation of non-productive antibodies against epitope tags could impact observed vaccine efficacy in early phases. It is also worth mentioning that TEV cleavage leaves a residual serine residue (ENLYFQ\S) that is not likely to be immunogenic by itself but could contribute to the development of a neoepitope.

The background binding of polyclonal antibodies to the library phage presents one potential source of error in the method presented here. While we sought to minimize this influence through pre-enrichment of the antibody pool by immunoprecipitation against TeeVax3 antigen,

we recommend the inclusion of an experimental arm panned against unimmunized sera or an antibody isotype control to enable identification of these tile sequences. A further challenge with using multiple rounds of phage panning is the resulting enrichment of high expressing rather than high-affinity peptide clones. Sequences with a growth advantage can bias tile frequency and lead to the false positive identification of linear epitopes. Here, we took steps to mitigate this issue by monitoring tile abundance before and after phage expression as well as using synthetic peptides to validate our top N-terminal hits. We suggest a maximum of two rounds of panning are required to detect tile enrichment when using next-generation sequencing approaches, as strong binders rapidly dominate the library. Further, employing two rounds of panning could provide additional information on the relative affinity of interactions assuming a limited influence from expression bias. When we removed the high abundance epitope tag peptides from the analysis and mapped the remaining tiles above the 95th percentile, we found a small number of linear epitopes unevenly distributed across the TeeVax3 antigen. Previous work has also found biases in antibody binding to T-antigens upon immunization. Raynes *et al.* (2023) report conformational binders with a bias towards the N-terminal ends of T18.1 N-domains exposed in a temperature-dependent manner. While these data are from a limited subset of T-antigens, it is clear further exploration of immunodominant motifs would be valuable for developing these proteins as vaccines.

Epitope mapping studies provide insights into immune response bias that inform antigen design in vaccines. Here, we identify the epitope tag on TeeVax3 as having unanticipated immunogenicity. Beyond the scope of the findings presented here, our method could be used to further profile larger peptide libraries or be generalized for the analysis of other bacterial or viral pathogens. For example, to establish TeeVax3 correlates of protection across all 21 known T antigen groups encoded by the tee gene from Group A Strep pathogens (Falugi et al., 2008; Steemson et al., 2014). This data would best be paired with experimentally generated information on conformational epitope binding. Nonetheless, the method presented here provides an accessible and low-cost approach to linear epitope panning using tools available to most molecular biology laboratories.

References

Antanasijevic, A., Bowman, C. A., Kirchdoerfer, R. N., Cottrell, C. A., Ozorowski, G., Upadhyay, A. A., Cirelli, K. M., Carnathan, D. G., Enemuo, C. A., Sewall, L. M., Nogal, B., Zhao, F., Groschel, B., Schief, W. R., Sok, D., Silvestri, G., Crotty, S., Bosinger, S. E., & Ward, A. B. (2022). From structure to sequence: Antibody discovery using cryoEM. *Science Advances*, *8*(3), eabk2039. https://doi.org/10.1126/sciadv.abk2039

Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., Du, S., Wang, J., Li, Q., Chen, X., Yu, Y., Wang, P., Zhang, Z., Liu, P., An, R., Hao, X., Wang, Y., Wang, J., Feng, R., … Xie, X. S. (2022). BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature*, *608*(7923), 593–602. https://doi.org/10.1038/s41586-022-04980-y

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, *2*(2), e107. https://doi.org/10.1002/imt2.107

Cia, G., Pucci, F., & Rooman, M. (2023). Critical review of conformational B-cell epitope prediction methods. *Briefings in Bioinformatics*, *24*(1), bbac567. https://doi.org/10.1093/bib/bbac567

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

De Leon, A. J., Tjiam, M. C., & Yu, Y. (2024). B cell epitope mapping: The journey to better vaccines and therapeutic antibodies. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1868*(10), 130674. https://doi.org/10.1016/j.bbagen.2024.130674

Falugi, F., Zingaretti, C., Pinto, V., Mariani, M., Amodeo, L., Manetti, A. G. O., Capo, S., Musser, J. M., Orefici, G., Margarit, I., Telford, J. L., Grandi, G., & Mora, M. (2008). Sequence Variation in Group A Streptococcus Pili and Association of Pilus Backbone Types with Lancefield T Serotypes. *The Journal of Infectious Diseases*, *198*(12), 1834–1841. https://doi.org/10.1086/593176

Fan, J., Toth, I., & Stephenson, R. J. (2024). Recent Scientific Advancements towards a Vaccine against Group A Streptococcus. *Vaccines*, *12*(3), Article 3. https://doi.org/10.3390/vaccines12030272

Garrett, M. E., Itell, H. L., Crawford, K. H. D., Basom, R., Bloom, J. D., & Overbaugh, J. (2020). Phage-DMS: A Comprehensive Method for Fine Mapping of Antibody Epitopes. *iScience*, *23*(10), 101622. https://doi.org/10.1016/j.isci.2020.101622

Gulati, S., McQuillen, D. P., Mandrell, R. E., Jani, D. B., & Rice, P. A. (1996). Immunogenicity of Neisseria gonorrhoeae Lipooligosaccharide Epitope 2C7, Widely Expressed In Vivo with No Immunochemical Similarity to Human Glycosphingolipids. *The Journal of Infectious Diseases*, *174*(6), 1223–1237. https://doi.org/10.1093/infdis/174.6.1223

Hanning, K. R., Minot, M., Warrender, A. K., Kelton, W., & Reddy, S. T. (2022). Deep mutational scanning for therapeutic antibody engineering. *Trends in Pharmacological Sciences*, *43*(2), 123–135. https://doi.org/10.1016/j.tips.2021.11.010

Hu, D., & Irving, A. T. (2023). Massively-multiplexed epitope mapping techniques for viral antigen discovery. *Frontiers in Immunology*, *14*, 1192385. https://doi.org/10.3389/fimmu.2023.1192385

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., & Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R., & Plückthun, A. (1997). Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *Journal of Immunological Methods*, *201*(1), 35–55. https://doi.org/10.1016/S0022-1759(96)00208-6

Larocca, R. A., Abbink, P., Peron, J. P. S., De A. Zanotto, P. M., Iampietro, M. J., Badamchi-Zadeh, A., Boyd, M., Ng'ang'a, D., Kirilova, M., Nityanandam, R., Mercado, N. B., Li, Z., Moseley, E. T., Bricault, C. A., Borducchi, E. N., Giglio, P. B., Jetton, D., Neubauer, G., Nkolola, J. P., … Barouch, D. H. (2016). Vaccine protection against Zika virus from Brazil. *Nature*, *536*(7617), 474–478. https://doi.org/10.1038/nature18952

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, *37*(23), 4572–4574. https://doi.org/10.1093/bioinformatics/btab705

Liao, J. J. Z., & Liu, R. (2009). Re-parameterization of five-parameter logistic function. *Journal of Chemometrics*, *23*(5), 248–253. https://doi.org/10.1002/cem.1218

Lin, T.-W., Huang, P.-H., Liao, B.-H., Chao, T.-L., Tsai, Y.-M., Chang, S.-C., Chang, S.-Y., & Chen, H.-W. (2022). Tag-Free SARS-CoV-2 Receptor Binding Domain (RBD), but Not C-Terminal Tagged SARS-CoV-2 RBD, Induces a Rapid and Potent Neutralizing Antibody Response. *Vaccines*, *10*(11), Article 11. https://doi.org/10.3390/vaccines10111839

Liu Yihao, Zeng, Q., Deng, C., Li, M., Li, L., Liu, D., Liu, M., Ruan, X., Mei, J., Mo, R., Zhou, Q., Liu, M., Peng, S., Wang, J., Zhang, H., & Xiao, H. (2022). Robust induction of B cell and T cell responses by a third dose of inactivated SARS-CoV-2 vaccine. *Cell Discovery*, *8*(1), 10. https://doi.org/10.1038/s41421-022-00373-7

Loh, J. M., Rivera-Hernandez, T., McGregor, R., Khemlani, A. H. J., Tay, M. L., Cork, A. J., M. Raynes, J., Moreland, N. J., Walker, M. J., & Proft, T. (2021). A multivalent T-antigen-based vaccine for Group A Streptococcus. *Scientific Reports*, *11*(1), 4353.

López-Laguna, H., Voltà-Durán, E., Parladé, E., Villaverde, A., Vázquez, E., & Unzueta, U. (2022). Insights on the emerging biotechnology of histidine-rich peptides. *Biotechnology Advances*, *54*, 107817. https://doi.org/10.1016/j.biotechadv.2021.107817

McNamara, H. A., Idris, A. H., Sutton, H. J., Vistein, R., Flynn, B. J., Cai, Y., Wiehe, K., Lyke, K. E., Chatterjee, D., Kc, N., Chakravarty, S., Lee Sim, B. K., Hoffman, S. L., Bonsignori, M., Seder, R. A., & Cockburn, I. A. (2020). Antibody Feedback Limits the Expansion of B Cell Responses to Malaria Vaccination but Drives Diversification of the Humoral Response. *Cell Host & Microbe*, *28*(4), 572-585.e7. https://doi.org/10.1016/j.chom.2020.07.001

Mohan, D., Wansley, D. L., Sie, B. M., Noon, M. S., Baer, A. N., Laserson, U., & Larman, H. B. (2018). PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nature Protocols*, *13*(9), 1958–1978. https://doi.org/10.1038/s41596-018-0025-6

Ngampasutadol, J., Rice, P. A., Walsh, M. T., & Gulati, S. (2006). Characterization of a peptide vaccine candidate mimicking an oligosaccharide epitope of *Neisseria gonorrhoeae* and resultant immune responses and function. *Vaccine*, *24*(2), 157–170. https://doi.org/10.1016/j.vaccine.2005.07.065

Otsyula, N., Angov, E., Bergmann-Leitner, E., Koech, M., Khan, F., Bennett, J., Otieno, L., Cummings, J., Andagalu, B., Tosh, D., Waitumbi, J., Richie, N., Shi, M., Miller, L., Otieno, W., Otieno, G. A., Ware, L., House, B., Godeaux, O., … Spring, M. D. (2013). Results from tandem Phase 1 studies evaluating the safety, reactogenicity and immunogenicity of the vaccine candidate antigen Plasmodium falciparum FVO merozoite surface protein-1 (MSP142) administered intramuscularly with adjuvant system AS01. *Malaria Journal*, *12*(1), 29. https://doi.org/10.1186/1475-2875-12-29

Paull, M. L., Bozekowski, J. D., & Daugherty, P. S. (2021). Mapping antibody binding using multiplexed epitope substitution analysis. *Journal of Immunological Methods*, *499*, 113178. https://doi.org/10.1016/j.jim.2021.113178

Przedpelski, A., Tepp, W. H., Pellett, S., Johnson, E. A., & Barbieri, J. T. (2020). A Novel High-Potency Tetanus Vaccine. *mBio*, *11*(4), e01668-20. https://doi.org/10.1128/mBio.01668-20

Qi, H., Ma, M., Hu, C., Xu, Z., Wu, F., Wang, N., Lai, D., Li, Y., Zhang, H., Jiang, H., Meng, Q., Guo, S., Kang, Y., Zhao, X., Li, H., & Tao, S. (2021). Antibody Binding Epitope Mapping (AbMap) of Hundred Antibodies in a Single Run. *Molecular & Cellular Proteomics*, *20*, 100059. https://doi.org/10.1074/mcp.RA120.002314

Rajan, J. V., McCracken, M., Mandel-Brehm, C., Gromowski, G., Pollett, S., Jarman, R., & DeRisi, J. L. (2021). Phage display demonstrates durable differences in serological profile by route of inoculation in primary infections of non-human primates with Dengue Virus 1. *Scientific Reports*, *11*(1), 10823. https://doi.org/10.1038/s41598-021-90318-z

Raynes, J. M., Young, P. G., Lorenz, N., Loh, J. M. S., McGregor, R., Baker, E. N., Proft, T., & Moreland, N. J. (2023). Identification of an immunodominant region on a group A Streptococcus T-antigen reveals temperature-dependent motion in pili. *Virulence*, *14*(1), 2180228. https://doi.org/10.1080/21505594.2023.2180228

Rodrigues, M. X., Yang, Y., de Souza Meira, E. B., do Carmo Silva, J., & Bicalho, R. C. (2020). Development and evaluation of a new recombinant protein vaccine (YidR) against *Klebsiella pneumoniae* infection. *Vaccine*, *38*(29), 4640–4648. https://doi.org/10.1016/j.vaccine.2020.03.057

Sevvana, M., & Kuhn, R. J. (2020). Mapping the diverse structural landscape of the flavivirus antibody repertoire. *Current Opinion in Virology*, *45*, 51–64. https://doi.org/10.1016/j.coviro.2020.07.006

Shrock, E. L., Shrock, C. L., & Elledge, S. J. (2022). VirScan: High-throughput profiling of antiviral antibody epitopes. *Bio-Protocol*, *12*(13), e4464–e4464.

Singh, M., Sori, H., Ahuja, R., Meena, J., Sehgal, D., & Panda, A. K. (2020). Effect of N-terminal poly histidine-tag on immunogenicity of *Streptococcus pneumoniae* surface protein SP0845. *International Journal of Biological Macromolecules*, *163*, 1240–1248. https://doi.org/10.1016/j.ijbiomac.2020.07.056

Smith, G. P., & Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. In *Methods in Enzymology* (Vol. 217, pp. 228–257). Academic Press. https://doi.org/10.1016/0076-6879(93)17065-D

Steemson, J. D., Moreland, N. J., Williamson, D., Morgan, J., Carter, P. E., & Proft, T. (2014). Survey of the bp/tee genes from clinical group A streptococcus isolates in New Zealand – implications for vaccine

development. *Journal of Medical Microbiology*, *63*(12), 1670–1678. https://doi.org/10.1099/jmm.0.080804-0

Xu, G. J., Kula, T., Xu, Q., Li, M. Z., Vernon, S. D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R. T., O'Connor, K. C., Walker, B., Larman, H. B., & Elledge, S. J. (2015). Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. *Science (New York, N.Y.)*, *348*(6239), aaa0698. https://doi.org/10.1126/science.aaa0698

Yaffe, Z. A., Sung, K., Bosire, R., Farquhar, C., Ngacha, D. M., Lohman-Payne, B., Nduati, R., John-Stewart, G., Matsen, F. A., & Overbaugh, J. (2023). Passively Acquired Constant Region 5–Specific Antibodies Associated With Improved Survival in Infants Who Acquire Human Immunodeficiency Virus. *Open Forum Infectious Diseases*, *10*(7), ofad316. https://doi.org/10.1093/ofid/ofad316

Zang, J., Zhu, Y., Zhou, Y., Gu, C., Yi, Y., Wang, S., Xu, S., Hu, G., Du, S., Yin, Y., Wang, Y., Yang, Y., Zhang, X., Wang, H., Yin, F., Zhang, C., Deng, Q., Xie, Y., & Huang, Z. (2021). Yeast-produced RBD-based recombinant protein vaccines elicit broadly neutralizing antibodies and durable protective immunity against SARS-CoV-2 infection. *Cell Discovery*, *7*(1), 1–16. https://doi.org/10.1038/s41421-021-00315-9

Zhao, X., Li, G., & Liang, S. (2013). Several Affinity Tags Commonly Used in Chromatographic Purification. *Journal of Analytical Methods in Chemistry*, *2013*, 1–8. https://doi.org/10.1155/2013/581093

Chapter Four

Simple high-throughput encoding of deep mutational scanning libraries by oligo-based golden gate assembly

Preface

As alluded to in chapter two, the affordable, efficient assembly of site-specific DMS variant libraries can prove a challenge - especially when multi-mutant libraries are desired. Type IIS restriction cloning approaches, such as golden gate assembly, are beginning to be utilised for the modular assembly of protein CDS. So far, techniques to generate mutagenised assembly parts have relied on a PCR step in some form, or expensive synthesis solutions. Here, we outline an affordable, PCR-free library generation approach that incorporates golden gate assembly with mutant oligo hybridisation. We demonstrate the approach by assembling several multi-mutant nanobody libraries for affinity maturation.

The research is presented in this chapter as a manuscript ready for submission. Supplementary material associated with this manuscript is presented in Appendix B.

Hanning, K.R., Walker, E.J., Beijerling, K., Irvine., E.B., Steel, J.J., and Kelton, W. (2025). Simple high-throughput encoding of deep mutational scanning libraries by oligo-based golden gate assembly. *bioRxiv,* https://doi.org/10.1101/2025.07.16.665225

Author contributions

K.R.H., and W.K. conceived of and designed the study. K.R.H., E.J.W., K.B, J.J.S, and W.K. performed laboratory experiments and created reagents used in the study. K.R.H., W.K. and E.B.I. performed the Illumina sequencing. K.R.H undertook the bioinformatic analysis. K.R.H., and W.K. wrote the manuscript.

Simple high-throughput encoding of deep mutational scanning libraries by oligo-based golden gate assembly

Kyrin R. Hanning[a], Emma J. Walker[a], Kevin Beijerling[a], Edward B, Irvine[c], J. Jordan Steel[d], William Kelton[a,b*]

[a]*School of Pharmacy and Biomedical Science, University of Waikato, Hamilton, New Zealand*
[b]*Te Aka Mātuatua School of Science, University of Waikato, Hamilton, New Zealand*
[c]*Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland*
[d]*Department of Biology, United States Air Force Academy, Colorado Springs, CO, United States*

Abstract

Control over mutational library diversity is essential consideration when engineering proteins, but is often fraught with trade-offs between diversity, specificity and affordability. Contemporary library assembly approaches often incorporate oligonucleotide pool synthesis to achieve affordable, precise mutagenesis, however, these oligos are often reliant on complex design to facilitate downstream PCR and/or restriction digests. Direct hybridisation of oligo pools is an overlooked strategy to simplify mutagenesis, especially when paired with a type IIS restriction cloning approach. We validate this approach through the design, hybridisation and deep sequencing of single and dual substitution CDR region parts derived from a nanobody gA10. Assembly of these parts into full-length nanobody CDS facilitated the phage-display of variant libraries for affinity maturation against its cyclic peptide target. Variants identified through enrichment analysis were expressed in isolation and yielded improved affinities in excess of 100-fold. Recent advances in machine learning have successfully inferred improved variants outside of screened library space, but require controlled, multi-mutant libraries. The library assembly approach outlined in this research is well-suited for such approaches.

Introduction

Engineering proteins via directed evolution has been instrumental for developing enzymes and antibody molecules as drugs and diagnostics. In these processes, point mutations are created within protein-coding sequences and linked via a protein display system (Starr et al., 2020) to protein function, or partitioned (Ma et al., 2018), for functional screening of a desired phenotype. Several methods have become popular for diversity generation within protein sequences, such as the use of error prone PCR to create distributed mutations throughout a sequence (Cadwell & Joyce, 1992). Challenges with mutational bias and the difficulty of guaranteeing full sequence coverage with error-prone approaches have led to newer methods with synthetic oligo assembly increasing in popularity. Typically, oligonucleotide pools are created, and the second strand is synthesised by PCR prior to cloning into a display vector (Kosuri & Church, 2014). Commercial providers will also create targeted libraries from oligonucleotide pools, but the costs increase drastically with the sequence diversity required. In all cases, a major challenge for directed evolution experiments is covering the enormous potential sequence space of mutations available for screening. For example, the potential combinatorial sequence diversity within a 10 amino-acid antibody CDR region alone exceeds 1×10^{13} variants (20^{10} variants), far surpassing the largest protein engineering libraries created (Bradbury et al., 2011; Plückthun, 2012). Given the size of most proteins being engineered vastly exceed 100 amino acids in size, new approaches are needed to strategically access protein diversity within the limits of experimental library construction.

One method to systematically tackle this diversity issue is to employ deep mutational scanning (DMS), a successor to alanine scanning that involves the systematic creation of diversity at defined positions within a protein framework (Fowler et al., 2014). In this approach, degenerate codons, such as NNK, are used to saturate mutations at desired locations, or even an entire protein sequence, for screening. Screening of DMS libraries has been widely used as an initial screen to eliminate mutations that are deleterious to protein function and constrain the diversity of subsequent protein engineering libraries (Hanning et al., 2022). However, creating libraries beyond first order DMS, where single sites are target, to include two or even three fully saturated positions (second or third order DMS) becomes technically challenging and expensive. Higher order libraries are highly desirable to facilitate the capture of epistatic interactions between mutation sites distal in the primary sequence but linked in a three-dimensional folded protein. The information contained within these interactions becomes more

important for the training of machine learning models that can assist with protein engineering schemes. Even more desirable are methods that can create a variety of high and low edit distances for robust model training (Taft et al., 2022).

Here, we develop a rapid and cost-effective method for building targeted higher order DMS protein engineering libraries. Degenerate oligonucleotides are designed and assembled by a coiled annealing approach to create overhangs compatible with golden gate assembly processes. A directional, multi-part assembly enables the creation of large libraries with targeted diversity for screening in protein display systems. As proof of concept, we apply our method to the affinity maturation of a phage displayed nanobody against a peptide mimetic from the obligate pathogen *N. gonorrhoea* and demonstrate several orders of magnitude improvement in affinity. We expect our approach could be used to generate libraries with customised diversity for optimal training of machine learning models and will have utility for other protein engineering applications beyond nanobodies.

Methods

Oligos and nanobody sequences

The GA10 nanobody sequence was obtained in-house from a prior screen of a synthetic library (Zimmermann et al., 2020) against PEP-1, a peptide mimetic of the 2C7 lipooligosaccharide epitope from the pathogen *Neisseria gonorrhoea* (Gulati et al., 1996). The crystal structure of the nanobody used in this work has been submitted to the PDB database (9B0A). All oligonucleotide pools were ordered from Integrated DNA Technologies at a scale of 50 pM per oligo. Wildtype and framework regions were ordered via gene synthesis (Twist Bioscience) for cloning into pUC19 or pAK200-derived vectors using restriction enzyme cloning. All full length nanobody sequences for expression were also ordered from Twist Biosciences.

Library design

The mutational targeting was designed to encompass both the residues originally defined by the established protocol and additional residues identified by crystallographic analysis as potentially interacting with the antigen, collectively targeting the CDR loops and adjacent regions for diversification.

Oligo pool duplexing and cloning

Lyophilised oligo pools were resuspended in Low TE buffer to 37.5 µM for single NNK pools, or 75 µM for dual NNK pools. Equimolar concentrations of complementary oligo pools were combined to a total of 300 pmol with T4 Polynucleotide Kinase and supplied buffer (New England Biolabs, USA) according to manufacturer instructions. The duplexing mixture was then incubated at 37 °C for 40 minutes, and 65 °C for 20 minutes to facilitate 5' phosphorylation of the ssDNA. Reactions were then immediately heated to 95 °C for five minutes followed by a cyclic, gradual cooling step. This comprised of an initial one-minute incubation at 90 °C, cooling at a constant rate of -0.1 °C/second until reaching 65 °C, at which point the reaction would be rapidly heated to the initial temperature at an increment of -0.5 °C/cycle. These cooling steps were repeated a total of 50 times, at which point the initial temperature was reduced to 65 °C. The reaction was then cooled to 4°C at a rate of -0.1 °C/second, before finally being diluted to a final concentration of 30 nM dsDNA, based on the reaction concentration of 3 µM dsDNA.

Duplexed oligo pools were then ligated with pre-prepared plasmids (Appendix 1) generated by BsaI-HFv2 digestion with concurrent quick-CIP (NEB) dephosphorylation followed by agarose gel excision. Plasmid ligations were performed using T4 DNA Ligase (New England Biolabs, USA) at a ratio of 3:1, followed by a standard incubation period of 20 °C for 20 minutes, and 65 °C for 20 minutes. Ligation products were then purified using the QIAquick Nucleotide Removal Kit protocol (Qiagen, Germany) with Monarch® DNA Cleanup Columns (5 µg; NEB, USA) for a final elution volume of 7 µL in ultra-pure water. 2 µL of purified ligation product was transformed into fresh, electrocompetent *E. coli* MC1061 cells via MicroPulser Electroporator (Bio-Rad, USA) according to the manufacturer's manual. Transformed cells were plated across several 140 mm LB agar plates (with ampicillin), with a 1:100 and 1:1000 dilution plated on 90 mm LB agar plates (with ampicillin) for quantification. Following overnight incubation at 37 °C, colonies were quantified and resuspended in 1 mL of LB media with 25% glycerol. Purified plasmid was extracted from approximately 2-15 OD_{600}·mL of this culture via QIAprep Spin Miniprep Kit (Qiagen, Germany).

Nanobody CDS assembly

Integration of parts into the full nanobody CDS utilised protocols outlined by New England Biolabs for a golden gate assembly strategy utilising BsaI-HFv2 and T4 DNA Ligase (NEB,

USA). Briefly, 30 fmol of destination plasmid was combined with 90 fmol of each other plasmid part, BsaI-HFv2, T4 DNA Ligase and T4 DNA Ligase buffer in a total reaction volume of 50 µL. The reaction was then incubated at 37 °C for five minutes, 16 °C for 5 minutes for 60 cycles total, terminating with 60 °C for five minutes. Ligation products were then purified using the QIAquick Nucleotide Removal Kit protocol (Qiagen, Germany) with Monarch® DNA Cleanup Columns (5 µg; New England Biolabs, USA) and the entire purification was transformed into fresh, electrocompetent NEB 5αF'I^q *E. coli* cells. After a brief recovery period (≤one hour; 37°C; 220 rpm), the entire transformation was plated both undiluted over large (140 mm or 245 mm^2 plates) LB plates supplemented with Chloramphenicol and glucose (25 μgmL^{-1}, 2% (w/v) respectively), and as a dilution (1:100, 1:1000) on standard 90 mm plates for quantification. Following overnight incubation at 37°C, colony numbers were quantified from the diluted plates, and scrape harvested with one mL chilled 2xYT containing glycerol (25% v/v). The harvested cultures OD_{600} determined prior to being snap frozen for storage at -80°C.

Variant library pooling

A total of seven assemblies were performed, each substituting out wild-type CDR part(s) for their equivalent duplexed-oligo-derived NNK part(s). The dual deep mutational scanning library was generated by proportionally pooling six assemblies; three assemblies simply substituted a CDR for the dual-NNK part to generate the intra-CDR NNK-containing variant sub-libraries; the remaining three assemblies combined the single NNK-containing CDR parts to generate all combinations of the dual inter-CDR variant sub-libraries. In contrast, the high-diversity combinatorial library containing up to six mutations (CMB) was created from only a single assembly of dual-NNK parts for each CDR.

Phage production and recovery

Variant libraries, thawed from frozen stocks, were inoculated into 50 mL of 2xYT medium containing chloramphenicol at an initial OD of 0.1. The cells were grown at 37°C to OD 0.4 - 0.6 before the additional of M13K07 helper phage at an MOI of 20. After resting the cultures for one hour, Kanamycin and IPTG were supplemented (35 μgmL^{-1}, 1 mM respectively) and growth continued at 30°C for 8 hours. Phage were harvested by collecting cultures supernatants after centrifugation for 20 mins at 11,000 xg. Each supernatant was filtered through a 0.22 µm syringe filter and $1/5^{th}$ the total volume of 20% PEG/2.5 M NaCl was added. The mixture was left on ice for at least 30 mins and centrifuged at 11,000 xg to pellet the phage before a second

round of PEG/NaCl precipitation. All phage was resuspended in 10 mM Tris-EDTA buffer and phage concentration was determined by plating a serial dilution series of phage incubated with mid-log E. coli (OD ~0.5) on chloramphenicol containing agar plates.

Phage was recovered via infection of 20 mL early-log phase (OD_{600} ~0.1) NEB 5αF'I^q *E. coli* in 2xYT medium supplemented with glucose (2% w/v) incubating at 37°C. After a one-hour rest period, Chloramphenicol was supplemented (25 μgmL^{-1}) and growth continued at 37°C until OD_{600} ~0.3-0.5. Infected cultures were centrifuged at 4,000 xg (4°C; five minutes) to pellet cells before resuspending them in ~1 mL of 2xYT containing glycerol (25% v/v) for storage at -80°C. Aliquots were taken for plasmid extraction and to quantify OD_{600} to facilitate production of the next generation of phage.

Phage panning

Phage libraries were subjected to a single, non-selective, reinfection round prior to surface panning using a protocol modified from the NEB Ph.D. Phage Display Libraries manual (NEB cat #E8100S) using 96-well ELISA plates. Briefly, coated antigen wells consisted of 50 µL of 10 μgmL^{-1} neutravidin with 50 nM 2C7 mimitope (PEP1) in 50 mM $NaHCO_3$, pH 9.5, and were incubated at 4°C overnight. Subsequent blocking was performed with PBS pH 7.4 w/ 0.05% (v/v) Tween-20 (T), and 2% (w/v) skim milk powder (M) for one hour, shaking at room temperature, followed by six washes with PBST. Phage libraries were diluted to 1 x 10^{12} CFU/mL in PBST, with 1 x 10^{11} per well. Phage were first depleted by incubating for 30 minutes in blocked wells (no coated antigen), before being transferred to coated wells and incubated for one hour (shaking, room temperature). Liquid was then discarded and the plate washed three times with PBST. Bound phage was eluted via incubation with 0.2 M Glycine-HCl (pH 2.2; 1 $mgmL^{-1}$ BSA) for five minutes before neutralisation with 1M Tris-HCL (pH 9.1) and transfer into a pre-blocked microfuge tube. The entirety of the neutralised phage eluate was subsequently recovered and propagated. Four rounds of selective screening were performed in total. For each round, per replicate, 8 x 10^{11} phage of the dual DMS library, and 1.2 x 10^{12} phage of the combinatorial library was screened.

Deep sequencing

Purified plasmids of CDR variant parts (cloned, duplexed-oligos), the fully assembled pre-selection libraries (after a non-selective reinfection step), and the post-selection libraries (after four rounds of screening) were used as templates to prepare amplicons for deep sequencing. Briefly, the variant regions were PCR amplified using custom-designed primers. In a

subsequent amplification step, sample-specific barcodes (Illumina Nextera) were introduced to allow pooling of the CDR variant parts and pre-/post-selection libraries into two sequencing runs using (Illumina MiSeq 3; 2x300 bp PE). All sequencing was performed by the Genomics Facility Basel, Switzerland. Quality control and preprocessing of sequencing reads was performed with Fastp (Chen, 2023) and aligned to reference via BWA (Li & Durbin, 2009). Aligned reads underwent a final trimming and polishing step for uniformity with custom Julia scripts (https://github.com/hanningk/MMAtools.jl). Several of these core custom Julia utilities were bundled within a module for downstream analysis.

Analysis

Aligned, polished reads were classified according to the following criteria: quantity of substitutions within and between the targeted CDR positions, and substitution content. Per position statistics were determined for substitution frequency and sum unique variants. Enrichment was determined using a simplified Bayesian framework adapted from Bloom (2015). For each variant, the minimum pre- or post-screen count (x) determines the sigmoidal pseudo-count prior (Eq. 1) where $a = 50$ is the maximum pseudo-count (applied when $x \approx 0$), $c = 1$ the inflection point, and $d = 1$ controls the slope. Given an observed count n for the variant in a library size of N, its posterior frequency θ is drawn (Eq. 2). Fold-change enrichment, R, is calculated normalised to the wild-type fold-change (Eq. 3). The reported enrichment score, $\langle \log_2 R \rangle$, is the arithmetic mean of the $M = 1000$ individual $\log_2$ ratios (Eq. 4).

$$p(x) = \frac{a}{1 + \exp\left(\frac{x - c}{d}\right)} \qquad (Equation\ 1)$$

$$\theta^{(i)} \sim \mathrm{Beta}\left(\underbrace{\mathrm{n} + \mathrm{p}(x)}_{\alpha}, \underbrace{(\mathrm{N} - \mathrm{n}) + \mathrm{p}(x)}_{\beta}\right) \qquad (Equation\ 2)$$

$$R^{(i)} = \frac{\theta^{(i)}_{post,\mathrm{var}} / \theta^{(i)}_{post,\mathrm{WT}}}{\theta^{(i)}_{pre,\mathrm{var}} / \theta^{(i)}_{pre,\mathrm{WT}}} \qquad (Equation\ 3)$$

$$\langle \log_2 R \rangle = \frac{1}{M} \sum_i \log_2 R^{(i)} \qquad (Equation\ 4)$$

Nanobody expression and ELISA analysis

Nanobody variants for experimental testing were chosen based on the degree of enrichment and final sequence abundance. Each are shown as red points data points on enrichment plots. Nanobodies were ordered in pET28a vectors for periplasmic expression in BL21(DE3) *E. coli* via the inclusion of a PelB leader sequence. Overnight cultures were used to inoculate 50 mL Terrific Broth media and then grown at 37°C until an O.D. of 0.6 was reached. Cultures were

transferred to 22°C, induced with 1 mM IPTG, and grown overnight. The next day, the cultures were spun at 4000 g and the cell pellet resuspended in 5 mL of 20 mM Tris, 150 mM NaCl Tris Buffered Saline (TBS) supplemented with 1 mM $MgCl_2$ for periplasmic extraction. The mixture was centrifuged at 11,000 g for 20 mins and filtered through 0.22 μm syringe filters. Nanobodies were purified using a 5 mL MabSelect column (Cytiva). Briefly, the column was equilibrated with TBS buffer before loading of the sample. After washing to baseline with TBS, nanobodies were eluted with a gradient from 0-100% over 1 column volume (5 mL) of 0.1 M Glycine. The eluted fraction with immediately neutralised with 10% v/v 1M Tris and buffer exchanged with 10 kDa spin columns back into TBS. Nanobody purity was evaluated by 4-20% Mini Protean TGX Precast gel (Bio-Rad).

To evaluate nanobody binding to the cyclic PEP-1 peptide, ELISA plates were coated with 50 μL volumes of 4 ug/mL neutravidin precomplexed with a 1:4 molar ratio of PEP-1 in 50 mM $NaHCO_3$. The plates were incubated overnight and blocked with 300 μL of TBS containing 2% skim milk powder (TBSM) w/v for 1 hr at room temperature. Nanobodies were prepared in TBSM and serially diluted down the plate, left at room temperature to bind for a further hour, and removed by washing 3X with TBS containing 0.05% Tween20 (TBST). Bound nanobodies were detected by incubation for an hour with 50 μL of anti-VHH in PBSM at a 1:10,000 dilution (Clone: 96A3F5, Genscript). The plate was washed 3X in TBST, developed with 50 uL Ultra TMB substrate (ThermoFisher), and quenched with 50 μL of 1M H_2SO_4. Absorbance was quantified at 450 nm via a plate reader.

Results

Efficient and Controlled Duplex Formation of Mutagenic Oligonucleotides

To create a modular approach for the flexible generation of diversity within targeted regions of protein sequence libraries, we demonstrate a gene assembly method using degenerate oligo hybridisation without PCR amplification. We selected a candidate nanobody with moderate affinity (PDB: 9B0A) raised against a cyclic peptide vaccine candidate against *Neisseria gonorrhoea* for affinity maturation via phage display. The nanobody was originally derived from a synthetic nanobody library (Zimmermann et al., 2020). The residues in that library were targeted and expanded upon based on the crystal structure antigen interface, collectively representing the complementarity-determining regions (CDRs; Fig. 4.1A). The CDRs represent logical parts for golden gate assembly methods, and their lengths ideal for synthesis as

oligonucleotides. The nanobody CDS was split into six contributing parts; three target CDRs and three non-targeted parts, optimising assembly fidelity of the overhangs (Weber et al., 2011). Each CDR part was synthesised as “single” and “dual” substitution oligo pools, within which targeted codons are substituted with single and dual-concurrent NNK degeneracy, respectively. To facilitate PCR-free mutagenesis, the equivalent, reverse complement oligo pool was also synthesised, incorporating NNM degeneracy. Duplexing of the two cognate pools would yield dsDNA with 5’-4nt overhangs – facilitating immediate post-duplexing ligation. To further improve fidelity of this ligation, 5’ phosphorylation was performed on the ssDNA oligos immediately prior to duplex formation (Fig. 4.1B). All parts were co-assembled with wildtype framework parts to generate full nanobody CDS sequences (Fig. 4.1C). To establish the flexibility of diversity profiles achieved with the method, we created libraries with tailored high and low diversity profiles of up to 6 (combinatorial) and up to 2 (dual DMS) mutations per variant respectively. All nanobody variants were expressed on M13 bacteriophage and screened against immobilised cyclic peptide antigen for 4 rounds of selection (Fig. 4.1D).

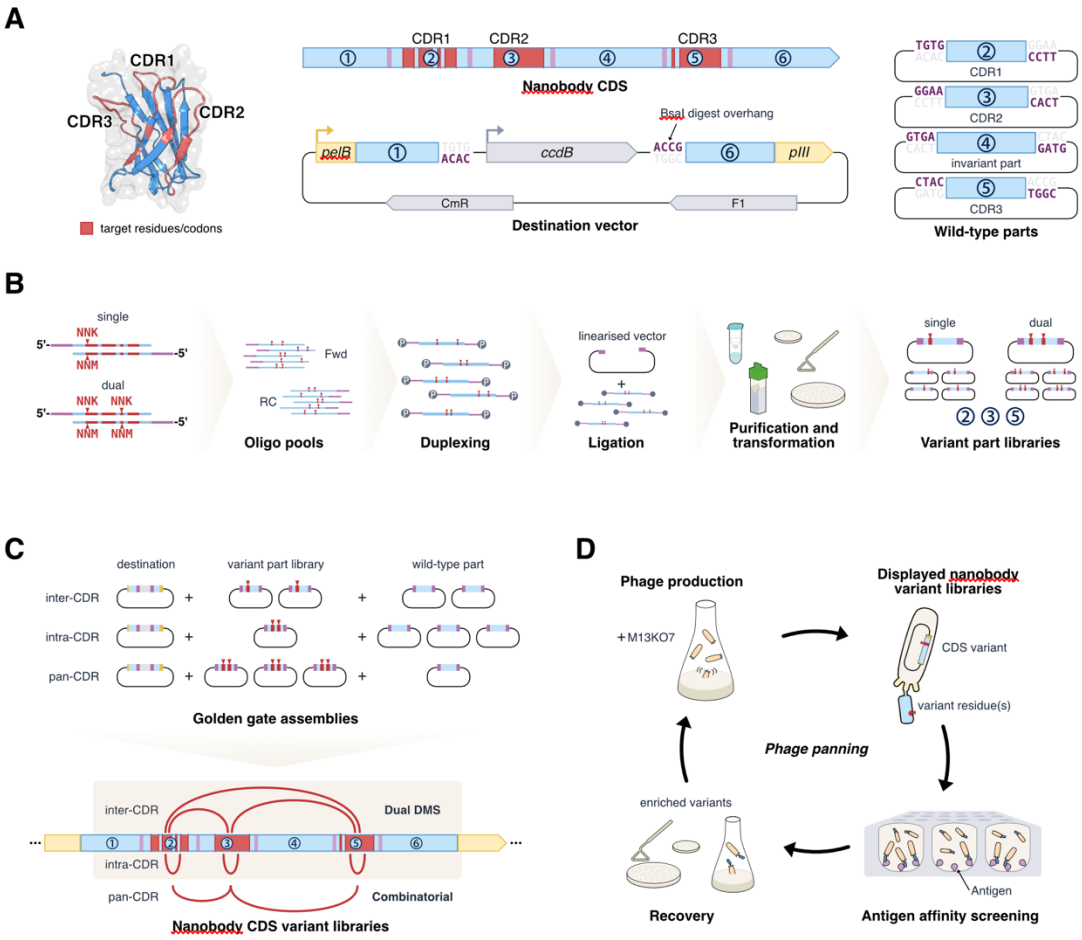


Figure 4.1. Variant nanobody library design and construction approach. Library design (A): Residues encompassing the CDR regions were targeted for mutation (red). The nanobody CDS was split into six parts to facilitate effective synthesis of the CDRs (parts 2, 3, and 5) as separate oligo pools. Part boundaries were determined by suitability of overhangs (purple/pink) for golden gate assembly via BsaI restriction-enzyme digest. The nanobody CDS termini (parts 1 and 6) were synthesised with an internal counter-selection cassette (*ccdB*) flanked by BsaI recognition sites. The fragment was cloned into a phagemid (F1 ori) destination vector, with 5' *pelB* and 3' M13 phage *pIII* fusions (yellow) to facilitate phage expression upon assembly with the remaining CDS parts (2-5). These parts were synthesised with flanking BsaI sites and cloned into pUC19-derived vectors prior to golden gate assembly. **Variant part generation (B):** Oligos of the three parts representing the CDRs (2, 3, 5) were designed as single or dual NNK-codon containing libraries and synthesised as seperate forward (FWD) and reverse complement (RC) oligo pools (IDT oPools). Each oligo contained a unique 5' golden gate overhang, which after 5' phosphorylation and duplexing via heat-cool cycles, enabled their immediate ligation into overhang-matched pre-digested vectors. The ligation products were immediately purified and electro-transformed to yield separate single and dual variant libraries for each part. **Golden gate assemblies (C):** By substituting wild type parts with variant library parts during golden gate assembly, seven nanobody CDS variant libraries were generated. Six of these formed the dual DMS library; ≤2 substitutions total, either inter- or intra-CDR. The seventh

library was assembled from the dual intra-CDR parts to yield a high diversity combinatorial library; ≤ 6 substitutions total, ≤2 intra-CDR. The assembled libraries were electro-transformed into NEB 5-alpha F'Iq *E. coli*. **Variant screening (D):** All bacterial libraries were cultured and infected with M13KO7 helper phage to produce nanobody-pIII fused phage. This recombinant phage was then screened against an immobilised antigen (PEP1) before recovery via infection and plating of fresh *E. coli*. This panning process was repeated for multiple rounds.

We first sought to validate the depth of coverage and fidelity of the oligo hybridisation method. Each of the six hybridisations for the single NKK and dual NNK variant pools were independently repeated before cloning into storage plasmids. Colony counts revealed a coverage depth of the theoretical diversity of between 294-, and 1300-fold for the single NNK parts and between 5.8-, and 50-fold for the higher diversity dual NNK parts (Table 1) and we observed a reasonable degree of consistency across both replicates. Minimising the time between oligo hybridisation and subsequent ligation was critical to ensuring a high coverage of variant diversity.

Table 1. Transformation yields for replicate (#1, #2) oligo pool duplexing. Yield represents the total approximate colonies harvested derived from dilution series, and assumed depth of total, targeted, non-redundant NNK variants.

	Cloned duplex	Theoretical Diversity	Yield (depth) #1	Yield (depth) #2
Single NNK	CDR1	346	3.0×10^5 (867×)	1.1×10^5 (318×)
	CDR2	408	1.2×10^5 (294×)	2.7×10^5 (662×)
	CDR3	377	4.9×10^5 (1300×)	1.1×10^5 (292×)
Dual NNK	CDR1	54,447	2.7×10^6 (50×)	9.0×10^5 (17×)
	CDR2	76,860	4.5×10^5 (5.8×)	9.5×10^5 (12×)
	CDR3	65,173	8.3×10^5 (13×)	1.8×10^6 (28×)

Characterisation of Initial Variant Space by Deep Sequencing

To probe coverage of the theoretical diversity at single nucleotide resolution and evaluate positional bias due to hybridisation, we next deep sequenced the individual part libraries. As the R2 reads were of insufficient quality, the R1 reads were instead processed, aligned and polished as single end reads (Fig. S1, S2). For the single site libraries, we observed 100% coverage of the theoretical diversity across all parts sequenced in both replicates. In the dual NNK libraries, we observed >97% coverage of the expected variant codons (>99% when

translated). We next interrogated the mutational frequencies at each position within the CDRs (Fig. 4.2A). As expected, the mean mutational frequency within the dual site assemblies is much lower than in the single site assemblies. Likewise, single site assemblies have fewer unique reads than their dual site counterparts. One notable exception was the CDR2 assembly which showed a reduced mutational frequency towards the 3' end of the assembly indicative of less diversity in this region. The use of deep sequencing also enabled evaluation of background or non-expected sequences that fell outside the design constrains of NKK codons. While most background sequences in non-mutated positions were consistently low, higher frequencies were observed in positions where NNK codons were used. Additionally, no obvious increase in mutational frequency is observed in codons that overlapped the golden gate overhang.

A global evaluation of mutations within each hybridised part also confirms a strong bias towards positions where diversity is desired (Fig. 4.2B). Variants falling outside the expected numbers of mutations with each part typically contained more mutations than expected up to a maximum of six. When the dual site libraries were investigated in greater detail by mapping linkage between mutational positions (Fig. 4.2C), the CDR2 assembly again presented with a strong bias of low mutational frequencies towards the 3' end. The CDR1 assembly also displayed a stronger tendency for linked mutations than the other two CDR assemblies. We lastly looked at individual codon frequencies, which were remarkably consistent for single and dual site assemblies suggesting good representation of diversity is maintained at a codon level noting that translation of the sequences to protein will eliminate some redundancy (Fig. 4.2D).

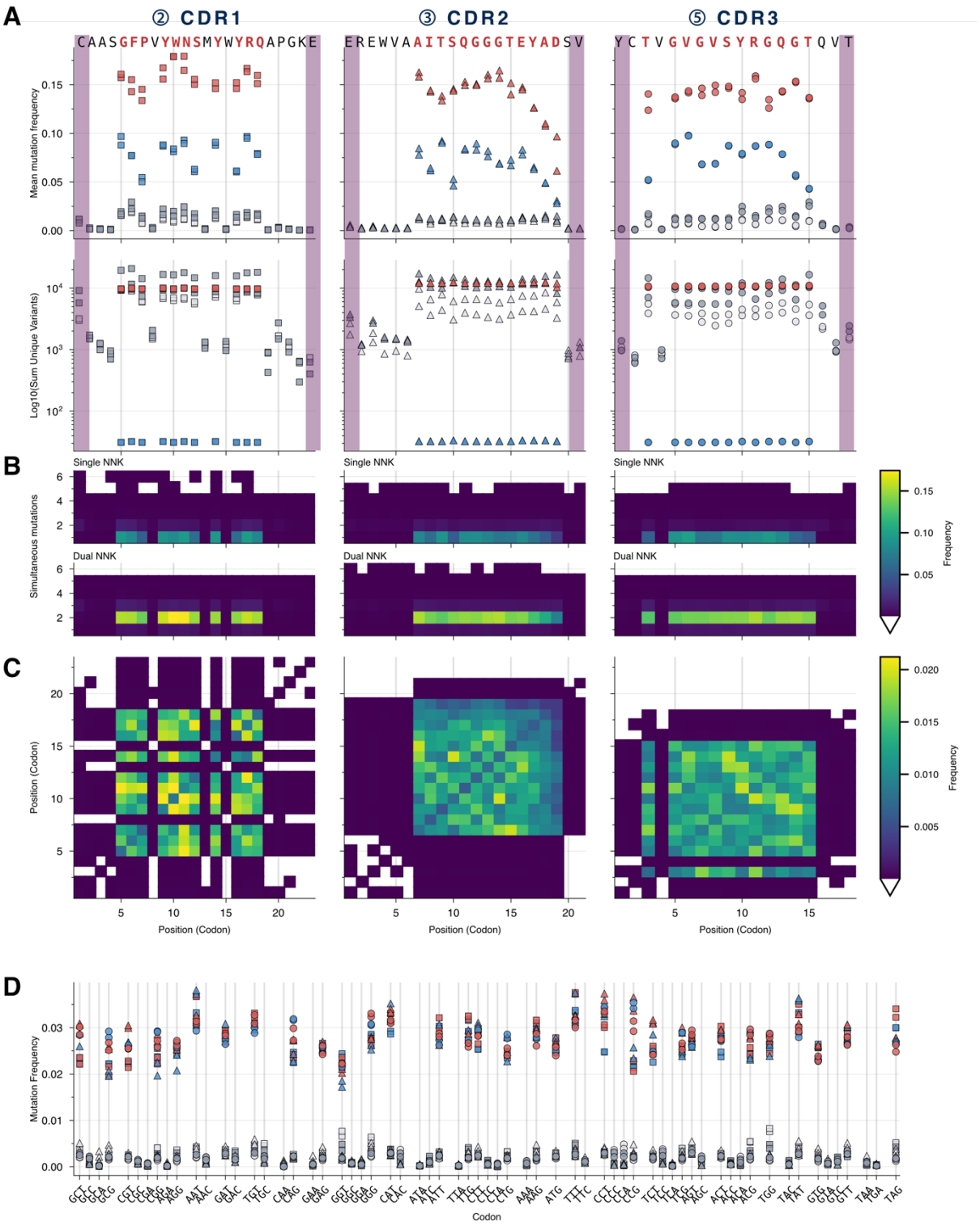


Figure 4.2.

Codon sequence variance on a per position basis within duplexed NNK oligo pools. Parts 2,3 and 5, corresponding to CDRs 1, 2 and 3, respectively were cloned into storage plasmids and deep sequenced. (**A**) Amongst variant sequences, the per position mutational frequency and contributing unique variants were calculated for the single (blue) and Dual (red) NNK libraries, and their replicates. Plotted are variants that met the intended design constraints (including permitted NNK codons, number of mutant codons) (light grey), and

variants that did not (dark grey). The wild-type amino acid sequence is shown, with targeted residues highlighted in red, and the type IIs restriction site overhang positions in pink. (**B**) Heatmaps represent mean frequency of mutant codon counts for each position within each single and dual NNK library cloned. (**C**) Within the dual NNK library (≤2 codon mutants), the mean frequency of variants affecting each position within each part/CDR were calculated and displayed by heatmap. (**D**) Codon preference of variants within the single (blue, light grey) and dual (red, dark grey) design constraints were calculated across the each of the three CDR parts (1: square; 2: triangle; 3: circle).

Golden gate assembly was used to assemble full-length constructs. Post-assembly transformation yielded a range of library coverage efficiencies (Table 2), while still enabling construction of a large-scale dual DMS library with minimal bottlenecking. Based on colony count, the diversity of the assembled, dual DMS sub-libraries exceeds the theoretical diversity (≥4.7×), whereas the combinatorial library reached ≤1.8^{-8} % of theoretical maximum.

Table 2. Golden gate assembly yields for replicate (#1, #2) assemblies contributing to the dual DMS libraries and combinatorial (CMB) libraries. Yield represents the total approximate colonies harvested derived from dilution series, and assumed depth of total, targeted, non-redundant NNK variants.

Assembly			Theoretical Diversity	Yield (depth) #1	Yield (depth) #2
Dual DMS	Single NNK	CDRs 1, 2	141,168	6.6×10^5 (4.7×)	4.4×10^6 (31×)
		CDRs 1, 3	130,442	7.2×10^5 (5.5×)	6.4×10^6 (49×)
		CDRs 2, 3	153,816	1.12×10^6 (7.3×)	7.6×10^6 (49×)
	Dual NNK	CDR1	54,447	1.0×10^6 (18×)	≥6.7×10^6 (123×)
		CDR2	76,860	1.7×10^6 (22×)	≥5.6×10^6 (73×)
		CDR3	65,173	≥5.6×10^6 (86×)	3.7×10^6 (57×)
CMB		CDRs 1, 2, 3	~3.04×10^{14}	8.3×10^5 (2.7×10^{-9}×)	≥5.6×10^6 (1.8×10^{-8}×)

Deep sequencing of the pre- and post-selection libraries facilitated a discrete analysis of the assembly products, in addition to the intended enrichment analysis. These libraries underwent identical processing, alignment and polishing as overlapping paired end reads (Fig. S1, S2). Dual DMS and combinatorial assemblies largely reflect the targeted per position bias (Fig. 4.3A, 4.3B) seen with the CDR parts (Fig. 4.2A). For both dual DMS and combinatorial libraries, most reads represented variants that conformed to the intended design criteria of the respective libraries (68% of dual DMS and 93% of combinatorial library reads). The larger

population of non-conformant variants in the dual DMS libraries was due to a number of variants containing > 2 substitutions. Regarding coverage of the theoretical library diversity, the dual DMS libraries achieved 36% (225,170 variants) and 62.5% (387,429 variants) for the two replicates. In contrast, the combinatorial library replicates showed much lower coverage, with 178,909 variants (5.9×10^{-8}%) and 952,365 variants (3.13×10^{-7}%) respectively.

Due to comparatively low off-target substitution, and to aid with enrichment analysis, reads were condensed into spans of the 36 target residues, comprising CDR1 (1-11), CDR2 (12-24) and CDR3 (25-36). Amongst the single and dual-substitution variants within the dual DMS library, there is a bias towards dual-substitutions with CDR3 residues, and internally within CDR2 (Fig. 4.3C). Multi-substitution trends within the condensed variants remained comparable to the full-length assembled sequence (Fig. 4.3D).

The enrichment of each variant was determined as the fold change from wild-type abundance. Typically, DMS analysis is restricted to those variants observed in the pre-screening library, this was enforced for analysis of the dual DMS and combinatorial libraries (Fig. 4.4A). Here, due to potential under sampling in the combinatorial library, we also considered the situation where variants were observed in the post-screened library but not in the pre-screened library. This approach captures a wider sequence space at the risk of reducing accuracy of enrichment prediction. For the dual DMS library replicates, both approaches showed a high degree of enrichment convergence, with Pearson correlation values of 0.84 and 0.83 for the restrictive and relaxed approaches. Deeper analysis of replicate sequence similarity in pre-screened libraries revealed a more modest Jaccard index (fraction of shared sequences) of 0.18 and 0.24 for the restrictive and relaxed approaches across the two replicates. For the combinatorial library replicates, undersampling of the potential library diversity meant a narrow pool of variants were observed in each of the replicate. The restrictive analysis approach only analysed a pool of 999

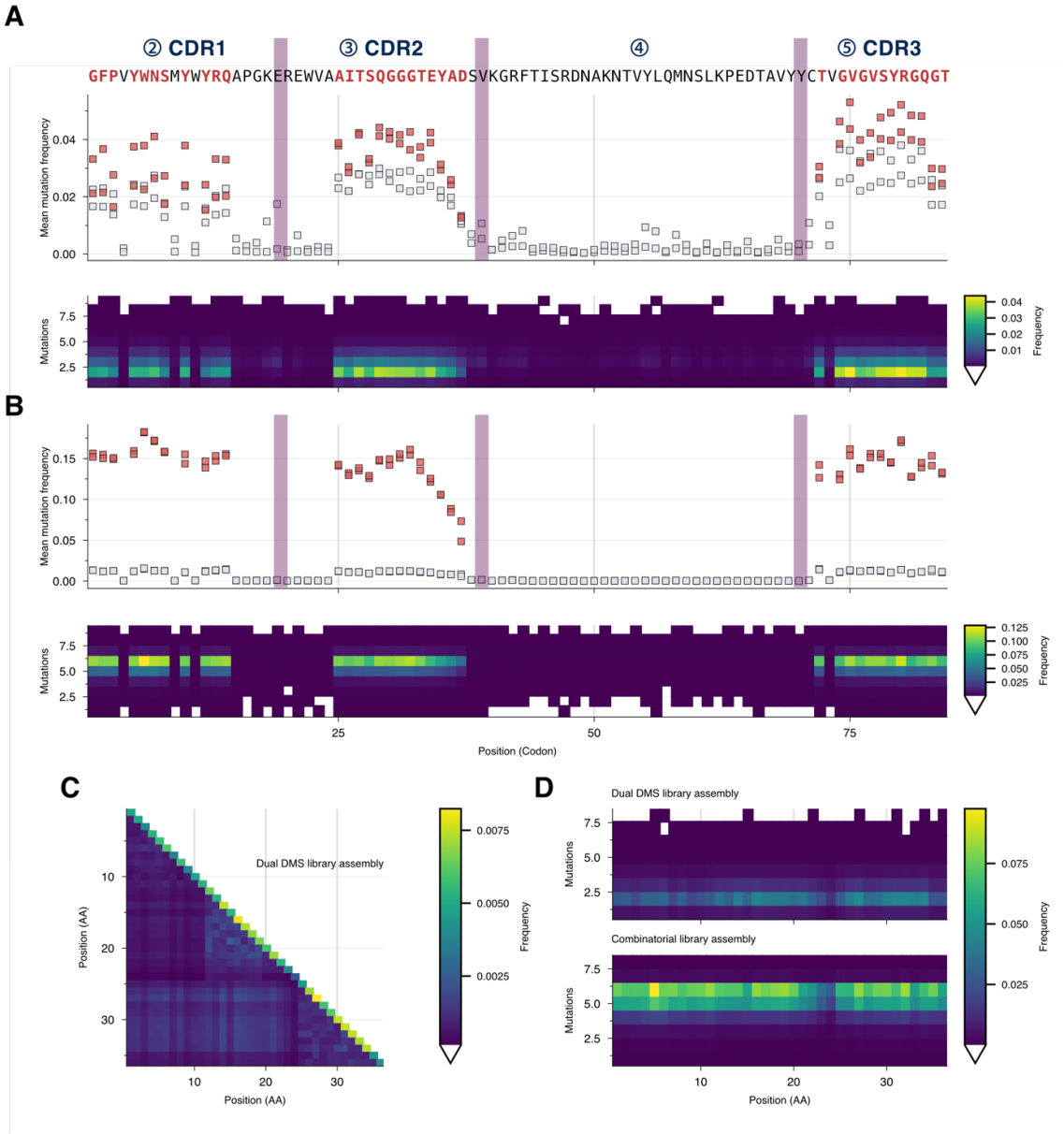
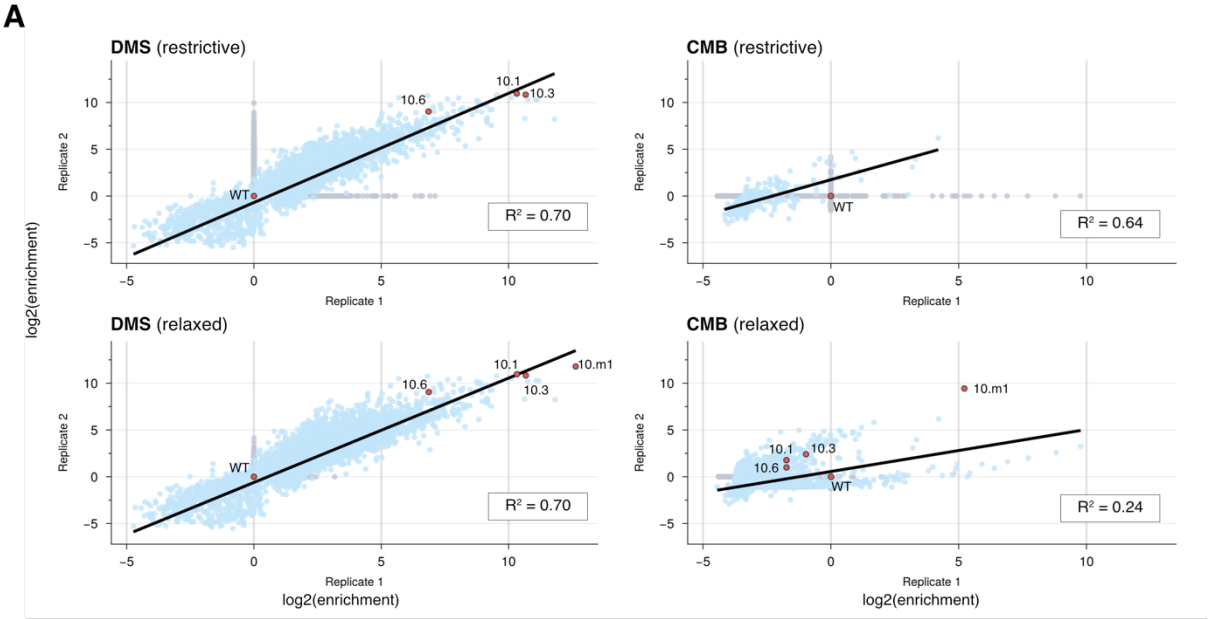


Figure 4.3. Per position variance derived from sequencing of fully assembled variants post one round of non-selective phage reinfection, representing the pre-selection variant library. (A) Amongst variants within the dual DMS library, the per position frequency of mutations at codons that fall within the intended design constraints (Red; ≤2 intra-CDR variants; ≤2 inter-CDR variants; no TGA or TAA stop codons) and variants that did not (Grey). The heatmap represents the mean frequency of simultaneous mutant codon counts at each position within the dual DMS library. (B) Amongst variants within the combinatorial library, the per position frequency of mutations at codons that fall within the intended design constraints (Red; ≤2 intra-CDR variants; ≤6 inter-CDR variants; no TGA or TAA stop codons) and variants that did not (Grey). The heatmap represents the mean frequency of simultaneous mutant codon counts at each position within the combinatorial library. (C, D) Variants were translated, and target residues concatenated. (C) Per position mean mutational frequency of dual DMS library variants with ≤2 substitutions. (D) Mean frequency of simultaneous mutant codon counts at each position within the dual DMS (top) combinatorial (bottom) libraries.

unique variants, which was expanded to 44,173 when using the relaxed approach. Overall, the Jaccard indices were < 0.01 irrespective of the analysis method between these replicates.



B

		log2(enrichment)	
		DMS	CMB
WT	GFPVYWNSMYWYRQAPGKEREWVAAITSQGGGTEYADSVK...YYCTVGVGVSYRGQGT		
10.1	.....R.....K.....	10.63	0.93
10.3	.....R.....E.....	10.75	1.74
10.6	..L.....K.....	7.97	0.54
10.m1	..L.....I.....SA.....VR.....	12.21	8.45

Figure 4.4. Summary of dual DMS and combinatorial enrichment. (A) Distribution of replicate enrichment ratios for the DMS (left) and combinatorial (right) libraries. Enrichment ratios were calculated by taking the mean posterior $\log_2$ enrichment relative to wild type (WT). Unique sequences observed in both replicates are coloured blue, whereas variants only observed in a single replicate are in grey. A linear regression was also fitted to shared replicate data using ordinary least squares (solid line) and the associated R^2 values reported. (B) Nanobody variants selected for expression. Each variant is represented as a red point on the enrichment plots.

The enrichment of each variant was determined as the fold change from wild-type abundance. Typically, DMS analysis is restricted to those variants observed in the pre-screening library, this was enforced for analysis of the dual DMS and combinatorial libraries (Fig. 4.4A). Here, due to potential under sampling in the combinatorial library, we also considered the situation where variants were observed in the post-screened library but not in the pre-screened library. This approach captures a wider sequence space at the risk of reducing accuracy of enrichment prediction. For the dual DMS library replicates, both approaches showed a high degree of enrichment convergence, with Pearson correlation values of 0.84 and 0.83 for the restrictive and relaxed approaches. Deeper analysis of replicate sequence similarity in pre-screened libraries revealed a more modest Jaccard index (fraction of shared sequences) of 0.18 and 0.24 for the restrictive and relaxed approaches across the two replicates. For the combinatorial

library replicates, undersampling of the potential library diversity meant a narrow pool of variants were observed in each of the replicate. The restrictive analysis approach only analysed a pool of 999 unique variants, which was expanded to 44,173 when using the relaxed approach. Overall, the Jaccard indices were < 0.01 irrespective of the analysis method between these replicates.

To determine whether nanobody variants represented with high enrichment scores possess higher affinity, we selected a small panel of variants for experimental affinity evaluation (Fig. 4.4B). These variants were chosen based on high abundance in the raw read counts and high ranking in enrichment scores across the two dual DMS replicates. Despite undersampling, we also selected one variant with six unique mutations (10.m1) that appeared in all libraries that had significantly higher raw read counts and enrichment than any other variant.

Nanobody variants were expressed in the periplasm of *E. coli* and purified using a conformationally selective protein A-based resin. SDS-PAGE analysis showed nanobodies at the expected sizes noting that the WT variant retains additional C-terminal sequences as result of cloning from the original synthetic library (Fig. 4.5A). Variants were evaluated for binding to the cyclic PEP-1 peptide relative to the wildtype nanobody. From the dual DMS library, the 10.1 and 10.3 nanobody clones displayed similar binding and showed improvements over the wildtype variant of more than 100-fold (Fig. 4.5B). The third ranked clone 10.6 also displayed improved affinity with a more than 50-fold increase relative to wildtype. A single variant with six mutations was chosen from the combinatorial library for analysis. Surprisingly, and despite a high enrichment score, the 10.m1 variant had a more modest but still significant affinity increase over WT of more than 20-fold. Nonetheless, all selected variants displayed large improvements over the base nanobody sequence highlighting the utility of this approach for rapid nanobody affinity maturation.

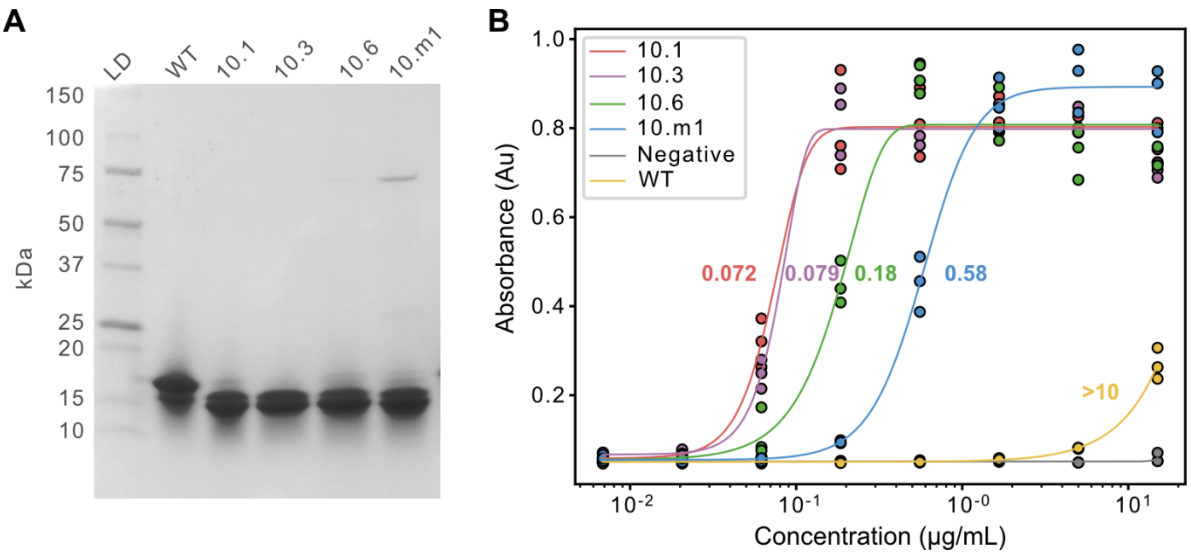


FIGURE 5.

ELISA analysis of engineered nanobody variant affinity. (**A**) Nanobody variants were selected with the highest-ranking enrichment scores from both the dual site (3 variants) and combinatorial libraries (1 variant). Nanobodies were expressed in E. coli, purified with MabSelect™ resin, and run by SDS-PAGE using a 4-20% gradient gel. The wild type variant was expressed in parallel. (**B**) Nanobody affinity was measured by ELISA against 2C7 peptide coated plates. Nanobodies were diluted from a starting concentration of 15 µg/mL with the anti-EGFR nanobody 7D12 used as a negative control. EC50 values were obtained by fitting of 5-parameter logistic (10.1, 10.6, 10.m1) or 4-parameter logistic models (WT).

Discussion

Synthetic biology approaches such as golden gate cloning are providing increased flexibility for protein engineers when designing libraries. In this study, we developed a method to create protein libraries with targeted and precisely controllable diversity profiles within larger protein sequences. We applied this method to create high and low edit distance libraries for the affinity maturation of a candidate nanobody with moderate affinity against a peptide vaccine candidate. Deep sequencing was used to characterise each step of the library construction process, revealing a minimal degree of codon bias and high coverage of the theoretical diversity at the individual part level. The final assembled nanobody libraries were panned against the cognate peptide resulting in binders with greater than 100-fold improvements in affinity.

Unlike traditional library construction approaches in which error prone methods are employed (Lee et al., 2018), targeted diversity allows a more constrained library focused on regions most likely to deliver a desired phenotype. This focusing becomes increasingly important as the size

of the protein being engineered increases. More recent approaches using gene synthesis in larger oligonucleotide pools have greatly improved the ease of targeted library construction containing saturation mutagenesis but lack flexibility in the control of edit distances (Fowler et al., 2014; Melnikov et al., 2014). However, obtaining large library sizes remains challenging and many of these methods use PCR to amplify regions of diversity which increases the risk of variant bias/drop-out before screening. A key difference in our approach has been to use complementary oligonucleotide pools for direct assembly with overhangs ready for Golden Gate Cloning. While we first cloned these pools into intermediate vectors to evaluate hybridisation fidelity by deep sequencing, direct assembly without prior cloning may further improve the speed of library generation. Access to alternative deep sequencing techniques, such as Oxford Nanopore, may reduce screening timelines while enabling an entirely PCR-free solution to data collection.

During oligonucleotide assembly, deep sequencing led us to notice a drop in mutation frequency towards the 3' end of the CDR2 parts in both the single and dual site assemblies. The CDR2 part possesses Golden Gate overhang sequences closest to the positions of introduced NNK diversity (3 bases). Depending on the GC content of the sequence, these results suggest increasing the distance between the proximal NNK codon and the Golden Gate overhang to at least five bases to improve oligo pool annealing. The use of deep sequencing further allowed us to evaluate the appearance of codons falling outside NNK degeneracy design profiles. It was surprising to note that background mutations were low in positions not targeted for variation but more concentrated where NNK codons were present. We speculate these background rates could be derived from mismatch repair processes occurring when annealed oligos do not find perfectly complementary sequences (Khilko et al., 2018; Kosuri & Church, 2014), rather than stemming from sequencing or oligo synthesis errors which would distribute errors more evenly across the sequence. However, given codon degeneracy during translation, and the ability to bioinformatically filter out these variants, they pose little issue when selecting for enriched variants.

In this work, the highest affinity nanobody variants were retrieved from the dual DMS libraries with up to two mutations. Under-sampling in the combinatorial library replicates resulted in comparison of enrichment scores across a very limited pool of shared variants. For high diversity libraries, this means replicate analysis is unlikely to add much additional accuracy over the screening of a single library. Further, our finding that the 10.m1 variant with six

mutations from wildtype did not have improved affinity over the variants selected from the dual DMS libraries also highlights another potential challenge with phage display methods whereby affinity and expression improvements are challenging to decouple.

While single-site DMS is a relatively accessible method, there are few examples of higher order DMS owing to the challenge of library construction. Accessing this diversity will enable easier systematic profiling of epistatic interactions within proteins. Alongside epistasis, the ability to control the depth of diversity within select regions of proteins has particular promise in machine learning guided protein engineering. These methods allow extrapolation of the binding landscape well beyond the experimental sequence space screened. Evidence is emerging that the edit distance distribution of libraries is a significant contributor to the accuracy of deep learning models predicting protein-protein binding interactions (Erlach et al., 2025; Frei et al., 2025; Taft et al., 2022). The modular system developed here allows diversity to be controlled by the mix of individual parts combined to form final library assemblies. Further strategies such as the incorporation of triple site NNK parts and the intentional doping of wildtype parts could provide further control when creating libraries. Further, noise in the form of false positive and false negative binders is a particular challenge during model training. Although approaches to mitigate noise with small subsets of high-quality data have emerged (Minot & Reddy, 2024), the ability to eliminate sources of noise using deep sequenced replicates may assist with model accuracy.

Control over protein library diversity has historically been challenging. Here, we present a PCR-free library building approach using short oligonucleotides that is both affordable and accessible to laboratories with molecular biology skill. Furthermore, we expect the modularity and scalability of the method to be broadly applicable beyond protein display and engineering applications.

References

Bloom, J. D. (2015). Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics*, *16*(1), 168. https://doi.org/10.1186/s12859-015-0590-4

Bradbury, A. R. M., Sidhu, S., Dübel, S., & McCafferty, J. (2011). Beyond natural antibodies: The power of in vitro display technologies. *Nature Biotechnology*, *29*(3), 245–254. https://doi.org/10.1038/nbt.1791

Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *Genome Research*, *2*(1), 28–33. https://doi.org/10.1101/gr.2.1.28

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, *2*(2), e107. https://doi.org/10.1002/imt2.107

Erlach, L., Friedensohn, S., Neumeier, D., Mason, D. M., & Reddy, S. T. (2025). *Antibody affinity engineering using antibody repertoire data and machine learning* (p. 2025.01.10.632313). bioRxiv. https://doi.org/10.1101/2025.01.10.632313

Fowler, D. M., Stephany, J. J., & Fields, S. (2014). Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nature Protocols*, *9*(9), 2267–2284. https://doi.org/10.1038/nprot.2014.153

Frei, L., Gao, B., Han, J., Taft, J. M., Irvine, E. B., Weber, C. R., Kumar, R. K., Eisinger, B. N., Ignatov, A., Yang, Z., & Reddy, S. T. (2025). Deep mutational learning for the selection of therapeutic antibodies resistant to the evolution of Omicron variants of SARS-CoV-2. *Nature Biomedical Engineering*, *9*(4), 552–565. https://doi.org/10.1038/s41551-025-01353-4

Gulati, S., McQuillen, D. P., Mandrell, R. E., Jani, D. B., & Rice, P. A. (1996). Immunogenicity of Neisseria gonorrhoeae Lipooligosaccharide Epitope 2C7, Widely Expressed In Vivo with No Immunochemical Similarity to Human Glycosphingolipids. *The Journal of Infectious Diseases*, *174*(6), 1223–1237. https://doi.org/10.1093/infdis/174.6.1223

Hanning, K. R., Minot, M., Warrender, A. K., Kelton, W., & Reddy, S. T. (2022). Deep mutational scanning for therapeutic antibody engineering. *Trends in Pharmacological Sciences*, *43*(2), 123–135. https://doi.org/10.1016/j.tips.2021.11.010

Khilko, Y., Weyman, P. D., Glass, J. I., Adams, M. D., McNeil, M. A., & Griffin, P. B. (2018). DNA assembly with error correction on a droplet digital microfluidics platform. *BMC Biotechnology*, *18*(1), 37. https://doi.org/10.1186/s12896-018-0439-9

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: Technologies and applications. *Nature Methods*, *11*(5), 499–507. https://doi.org/10.1038/nmeth.2918

Lee, J. M., Huddleston, J., Doud, M. B., Hooper, K. A., Wu, N. C., Bedford, T., & Bloom, J. D. (2018). Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proceedings of the National Academy of Sciences*, *115*(35), E8276–E8285. https://doi.org/10.1073/pnas.1806133115

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Ma, F., Chung, M. T., Yao, Y., Nidetz, R., Lee, L. M., Liu, A. P., Feng, Y., Kurabayashi, K., & Yang, G.-Y. (2018). Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nature Communications*, *9*(1), 1030. https://doi.org/10.1038/s41467-018-03492-6

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., & Mikkelsen, T. S. (2014). Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Research*, *42*(14), e112. https://doi.org/10.1093/nar/gku511

Minot, M., & Reddy, S. T. (2024). Meta learning addresses noisy and under-labeled data in machine learning-guided antibody engineering. *Cell Systems*, *15*(1), 4-18.e4. https://doi.org/10.1016/j.cels.2023.12.003

Plückthun, A. (2012). Ribosome Display: A Perspective. In J. A. Douthwaite & R. H. Jackson (Eds.), *Ribosome Display and Related Technologies: Methods and Protocols* (pp. 3–28). Springer. https://doi.org/10.1007/978-1-61779-379-0_1

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J. D. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, *182*(5), 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012

Taft, J. M., Weber, C. R., Gao, B., Ehling, R. A., Han, J., Frei, L., Metcalfe, S. W., Overath, M. D., Yermanos, A., Kelton, W., & Reddy, S. T. (2022). Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain. *Cell*, *185*(21), 4008-4022.e14. https://doi.org/10.1016/j.cell.2022.08.024

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, *6*(2), e16765. https://doi.org/10.1371/journal.pone.0016765

Zimmermann, I., Egloff, P., Hutter, C. A. J., Kuhn, B. T., Bräuer, P., Newstead, S., Dawson, R. J. P., Geertsma, E. R., & Seeger, M. A. (2020). Generation of synthetic nanobodies against delicate proteins. *Nature Protocols*, *15*(5), 1707–1741. https://doi.org/10.1038/s41596-020-0304-x

Chapter Five
Fc engineering and Future directions

Preface

As discussed in chapter two, functional screening of mutagenised complementarity-determining regions is an integral step of contemporary antibody development. This artificial affinity maturation approach aims to improve the nascent affinity of a candidate antibody to therapeutically relevant levels, and substantial gains can rapidly be obtained as observed with the nanobody example in chapter four. Here, we showcase preliminary data for a more complex library design of the IgA Fc designed for parallel functional screening against more than one Fc receptor.

The research in this chapter was deemed too preliminary for a complete manuscript and continues to be under development.

Author contributions

All presented data was collected by the first author of this thesis. K.R.H., and W.K. conceived of and designed the study.

Abstract

Despite the success of the therapeutic antibodies, there are still many unmet medical needs requiring the development of next-generation antibodies with expanded functionality. Immunoglobulin A (IgA) is an antibody subclass that has not yet been widely explored as a therapeutic scaffold. Recent studies suggest IgA has promise as a next-generation anti-cancer drug, with function being mediated primarily by its interaction with the immune receptor FcαRI. The therapeutic viability of IgA is dependent on resolving several pharmacokinetic problems. Augmenting IgA potency by optimising its interaction with FcαRI is one approach to improve its therapeutic potential. This can be achieved by mapping the tolerated mutational landscape of IgA for FcαRI, achievable through deep mutational scanning approaches. By targeting the interaction-interface of IgA and FcαRI, the variant library may also be amendable to mapping against other Fc receptors, such as TRIM21, providing a pathway for simultaneous, parallel screening.

Design of higher complexity combinatorial libraries

A key avenue of future research will be the construction of libraries with higher complexity (more assembled parts). Here, we focused on the Fc domain from IgA. As all approved therapeutic antibodies to date are based on the IgG Cγ scaffold there is a limited number of effective immune responses available (Mullard, 2021). The development of next-generation therapeutic antibodies should optimise immune responses, and IgA possesses several characteristics that make it attractive to develop as a therapeutic scaffold (Breedveld & van Egmond, 2019; Sterlin & Gorochov, 2021; van Tetering et al., 2020). However, several obstacles need to be overcome in order to realise IgA's potential.

FcαRI

The most potent immune response available to IgA is activated by FcαRI (Bakema & van Egmond, 2011; Mkaddem et al., 2014). Expression of FcαRI is constitutive, and limited to immune cells of the myeloid lineage, including neutrophils, eosinophils, most macrophages, and some dendritic cells. FcαRI binds within the C_{H2}-C_{H3} interface of each Fc monomer. Upon sustained interaction with IgA-antigen complexes, multiple FcαRI monomers cross-link,

resulting in full phosphorylation of the immunoreceptor tyrosine-based activation motif (ITAM). Phosphorylated ITAM triggers a signalling cascade that ultimately leads to immune cell activation and responses. The immune complex is critical for cross-linking and activation and provides IgA with high avidity for FcαRI. Interactions with singular IgA only partially phosphorylates ITAM, resulting in a long-lasting signalling cascade that inhibit local immune cell responses (Pasquier et al., 2005). The responses FcαRI activates and inhibits vary by immune cell, but include phagocytosis, cytokine synthesis, proliferation, and degranulation.

Unlike many antibody receptors, such as the well-studied IgG FcRs, FcαRI will bind IgA regardless of glycosylation (Gomes et al., 2008). Interestingly, the IgA2 subclass has greater affinity for FcαRI, but deglycosylation of the IgA1 subclass can improve its affinity for FcαRI to similar levels (Steffen et al., 2020). The affinity of FcαRI for monomeric IgA in general is considered low-medium among receptors (*K*d = ~176 nM) (Herr et al., 2003). The dual activatory-inhibitory nature of FcαRI suggests it plays an important role in maintaining immune homeostasis, and indeed many autoimmune diseases are associated with aberrant IgA and FcαRI interactions (Breedveld & van Egmond, 2019).

Neutrophil response

Neutrophils are the most abundant myeloid cell and constitute up to 70% of circulating white blood cells (Burn et al., 2021; Rosales, 2018). With a half-life of only six to twelve hours, large quantities of neutrophils are constantly produced (~10^{11} cells per day) and several mechanisms exist to increase production rates further. Circulating neutrophils migrate to tissues where they perform their immune functions before expiring.

Both IgA and IgG can mediate neutrophil responses such as antibody-independent cellular cytotoxicity (ADCC), phagocytosis, superoxide generation, and cytokine release, by binding their cognate receptors FcαRI and FcγRIIA (Huls et al., 1999). Effective destruction of cancer cells by therapeutic IgG is thought to include neutrophil ADCC, during which neutrophils engulf a stretch of the target cell surface to mechanically destroy the cell membrane causing cell death (Matlung et al., 2018). IgA and FcαRI are also thought to trigger cancer cell destruction via neutrophil ADCC, however overall neutrophil immune response is significantly more potent (Wehrli et al., 2020). This is best demonstrated in a direct comparison *in vitro* between IgA2 and IgG1 antibodies that possess the same cancer cell antigen target. The immune response initiated by IgA2 was similar or better to IgG1, yet five-fold more IgG1 was detected in serum (Boross et al., 2013).

This is likely due, at least in part, to ability of FcαRI to induce the release of chemokines such as leukotriene B4 (LTB4), a powerful neutrophil attractant. FcαRI-mediated release of LTB4 simulates neutrophil migration and is capable of attracting thousands of neutrophil cells towards an IgA-bound target within 20 minutes *in vitro* (van der Steen et al., 2009). Increasing the neutrophil titre at the target site magnifies the quantity of neutrophil-mediated immune responses. This is further exacerbated by the abundance of circulating neutrophils, resulting in rapid neutrophil migration to the target. Additionally, increased neutrophil production rates can be artificially induced, enabling a larger neutrophil pool to be recruited for target destruction (Roberts, 2005). IgA, FcαRI and neutrophils make a uniquely potent trio, and are the principal reason for IgA's potential as a therapeutic antibody.

Obstacles

Several obstacles have waylaid the development and successful deployment of therapeutic IgA, including: a lack of suitable animal model, aggregate and glycosylation issues when producing recombinant IgA, and IgA's short serum half-life. Mice lack a homolog of FcαRI, and efficient *in vivo* testing of IgA monoclonal antibodies (mAb) has had to wait for the development of transgenic FcαRI-expressing mice (Boross et al., 2013; van Egmond et al., 1999). Preventing aggregation and obtaining desired glycosylation patterns for recombinant IgA has proven difficult, and some promising approaches to production result in detrimental side-effects to IgA half-life and stability (Plomp et al., 2018; Rifai et al., 2000; Rouwendal et al., 2016). The already lacklustre half-life of IgA is perhaps the most significant obstacle to overcome. Continual fine-tuning of glycan profiles, or their removal entirely, and attempts to impart aspects of the IgG antibody recycling mechanism have proven to be rather successful in increasing the IgA half-life (Meyer et al., 2016). Alternatively, the development of IgG-IgA hybrids and other FcαRI-engaging molecules can circumvent some obstacles inherent with IgA (van Tetering et al., 2020).

Another route to enhance the therapeutic viability of IgA is by augmenting its ability to interact with FcαRI. Improved IgA affinity for FcαRI could achieve desired immune responses with lower does (Chen et al., 1999). The impact of glycosylation is suggested to be largely restrictive or irrelevant of IgA to bind FcαRI, depending on the subclass (Gomes et al., 2008; Steffen et al., 2020). Several IgA amino acid residues common to IgA1 and IgA2 have been demonstrated as important for binding FcαRI (Herr et al., 2003). Further work is required to establish which IgA residues are responsible for the FcαRI interaction can be developed for optimised affinity. Too high affinity for FcαRI could result in rapid clearance of IgA as

neutrophils expire, hence a range of affinity improvements would need to be assessed for their therapeutic viability.

Methods

Library design and construction

Crystallographic analysis suggests the interaction interface of IgA1 Fc bound to FcαRI (PDB ID: 1OW0) contains 60 proximate residues (≤2.5 Å), including several residues known to influence affinity. This also overlaps with the predicted interaction interface with TRIM21 (PDB ID: 21WG). These residues are collectively represented by five discontinuous regions spanning a total length of 196 residues, neatly forming part bounds amendable for oligo-based golden gate assembly (Hanning et al., 2025). All 60 residues were targeted for mutagenesis, with single and dual NNK oligo pools synthesised (IDT, USA) and duplexed for each of the five regions as previously described. Briefly, variant IgA1-Fc CDS were assembled through 15 parallel golden gate assemblies to generate ≤2 substitution libraries (dual DMS) and a single assembly to generate a ≤10 substitution (≤2 within each region) library (combinatorial; CMB). These assemblies were transformed into *Escherichia coli* XL1-Blue, quantified, scrape-harvested, and subsequently snap-frozen.

Phage production

To circumvent the potential length limitations of 2x300 paired end deep sequencing, four discrete dual DMS libraries were proportionally pooled from bacterial stocks wherein mutations were restricted to specific regions: regions 1,2,3,4 (DMS-A), regions 2,3,4,5 (DMS-B), regions 1,2,4,5 (DMS-C), and all regions (DMS-D). These libraries were each mixed directly in 50 mL of 2xYT medium containing chloramphenicol to an initial OD of 0.1. These cultures were then grown and phage produced, purified and quantified as previously described (Hanning et al., 2025).

Receptor expression and biotinylation

hFcαRI-GST were produced by transient transfection of Expi293 cells (Invitrogen, USA) as previously described (Kelton et al., 2014). This was purified via Glutathione Sepharose (Cytiva, USA) affinity chromatography according to manufacturer's instructions. HLTV-hTRIM21 was produced in *E. coli* BL21 as previously described (Clift et al., 2017) and purified via Nickel Sepharose (Cytiva) affinity chromatography followed by Superdex^{TM} 200 10/300

GL (Cytiva) size exclusion chromatography according to manufacturer's instructions. Biotinylation of purified receptor proteins was achieved via two hour incubation on ice with a 20-fold molar excess of Sulfo-NHS-LC-Biotin (Abcam, UK) in Phosphate-buffered solution (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) pH 7.4 followed by three 10 minute 4,000 g centrifugation steps at 4°C in pre-rinsed Amicon® Ultra 10 kDa MWCO 2 mL concentrators (Merk, Germany) to remove residual free biotin, topping up with fresh PBS throughout. Confirmation of biotinylation was achieved using a colorimetric assay with 4'-hydroxyazobenzene-2-carboxylic acid (ThermoFisher, USA) according to manufacturer's guidelines.

ELISA

To evaluate phage-displayed Fc binding to immune receptors, ELISA plates were coated with 50 µL of receptor in 50 mM $NaHCO_3$. Receptor concentrations varied according to the assay, with 0.5-15.0 μgmL^{-1} used for FcαRI and 7.5-15.0 μgmL^{-1} FcRn assays. Depending on the assay, PBS at either pH 7.4 or pH 5.8 (150 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH with HCl) were utilised. Plates were incubated overnight at 4°C, blocked with 300 µL PBS containing 2% skim milk powder (w/v; PBSM) for one hour at room temperature. Plates were washed via three sequential one minute incubations with PBS containing 0.05% Tween-20. Phage were prepared in PBSM and serially diluted across the plate, incubated for one hour at room temperature. Plates were washed once again, before detection via one hour incubation at room temperature with anti-M13 phage coat protein in PBSM at a 1:20,000 dilution (Clone: MM05, Sino-Biological). The plate was finally washed five times, developed in 50 µL Ultra TMB substrate (ThermoFisher), and quenched with 50 µL 1 M H_2SO_4. Absorbance was quantified at 450 nm via a plate reader.

Phage panning

Phage libraries were subjected to a single, non-selective reinfection round prior to solution-phase panning using a protocol modified from the NEB Ph.D Phage Display Libraries manual (NEB cat #E8100S). All incubation steps were performed on a rotational mixer at room temperature unless otherwise stated, PBS at either pH 7.4 or pH 5.8 was utilised depending on the desired screen. Briefly, all microfuge tubes used with and without StreptAvidin C1 or T1 (ThermoFisher) magnetic beads were first blocked with 1 mL 50 mM $NaHCO_3$ containing 5 $mgmL^{-1}$ bovine serum albumin and 0.1 μgmL^{-1} NeutrAvidin (ThermoFisher) for one hour at 4°C. These were then washed four times with PBS containing 0.5% Tween-20 (PBST). Phage

libraries were diluted to 2 x 10^{11} in PBST to a total volume of 200 µL and incubated for 15 minutes with washed magnetic beads. The beads were then pelleted, the supernatant transferred to a clean, pre-blocked microfuge tube and the desired amount of target Fc receptor was added. This phage-target complex was incubated for 20 minutes before being transferred to new washed magnetic beads. These were pelleted, the supernatant discarded, before being washed 10 times with PBST. Elution of phage was achieved by resuspending pelleted beads with 1 mL of 0.2 M Glycine-HCl (pH 2.2; 1 $mgmL^{-1}$ bovine serum albumin), incubating for 10 minutes, before pelleting beads and transferring the supernatant to a clean pre-blocked microfuge tube followed by immediate neutralisation with 150 µL 1 M Tris-HCl (pH 9.1). The entirety of the phage eluate was subsequently recovered and propagated.

Preliminary data

As a first step towards generating IgA libraries, we first determined whether aglycosylated IgA Fc domains could be displayed on M13 bacteriophage and, most importantly, whether they retain binding for hFcαRI-GST when glycans are absent. ELISA data showed strong binding by wildtype IgA Fc domains to recombinant FcαRI expressed in Expi293 cells (Fig. 5.1). IgG was used as a negative control and did not bind to FcαRI.

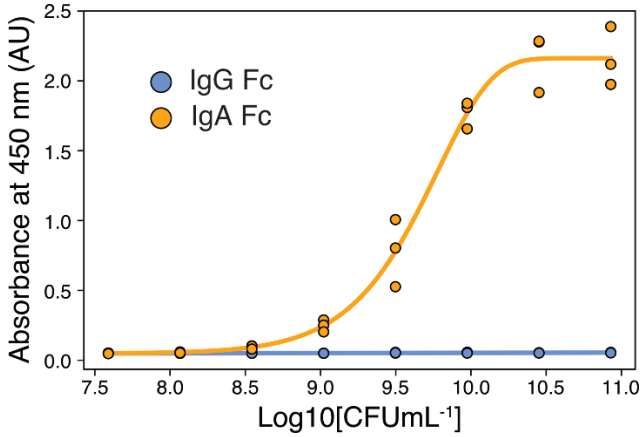


Figure 5.1. ELISA result of displayed IgA Fc and IgG Fc against hFcαRI-GST. Plates were coated with 4 µg/mL-GST in Na_2HCO_3 buffer at pH 9.5. After blocking with PBS containing 2% skim milk, IgA expressing bacteriophage was diluted down the plate for 1 hr. The plates were washed with PBS containing 0.05% Tween20, and anti-M13 HRP (Clone: MM05, Sino Biological) was added at 1:20,000. ELISAs were washed again before development with TMB substrate and neutralisation with 1M H_2SO_4. Absorbance was read at 450 nm. A five-parameter logistic model was fit to the ELISA data of displayed IgA Fc (R^2= 0.98; RSME = 0.11). $CFUmL^{-1}$ is a measurement of phage titre.

We then built a dual DMS library encompassing > 60 positions within the human IgA antibody (Fig. 5.2A) sequence targeting the CH2/CH3 interface responsible for IgA receptor binding (Fig. 5.2B). The library design resulted in five non-continuous regions or ‘parts’ where mutational diversity was targeted (Fig. 5.2C). A dual DMS library was constructed by direct oligonucleotide annealing using the method presented in chapter four (Fig. 4.1). Annealed oligos were combined with wildtype sequences not targeted for mutation and assembled by Golden Gate cloning into phage display vectors. The final library is ready for screening against IgA receptors as described in the methods.

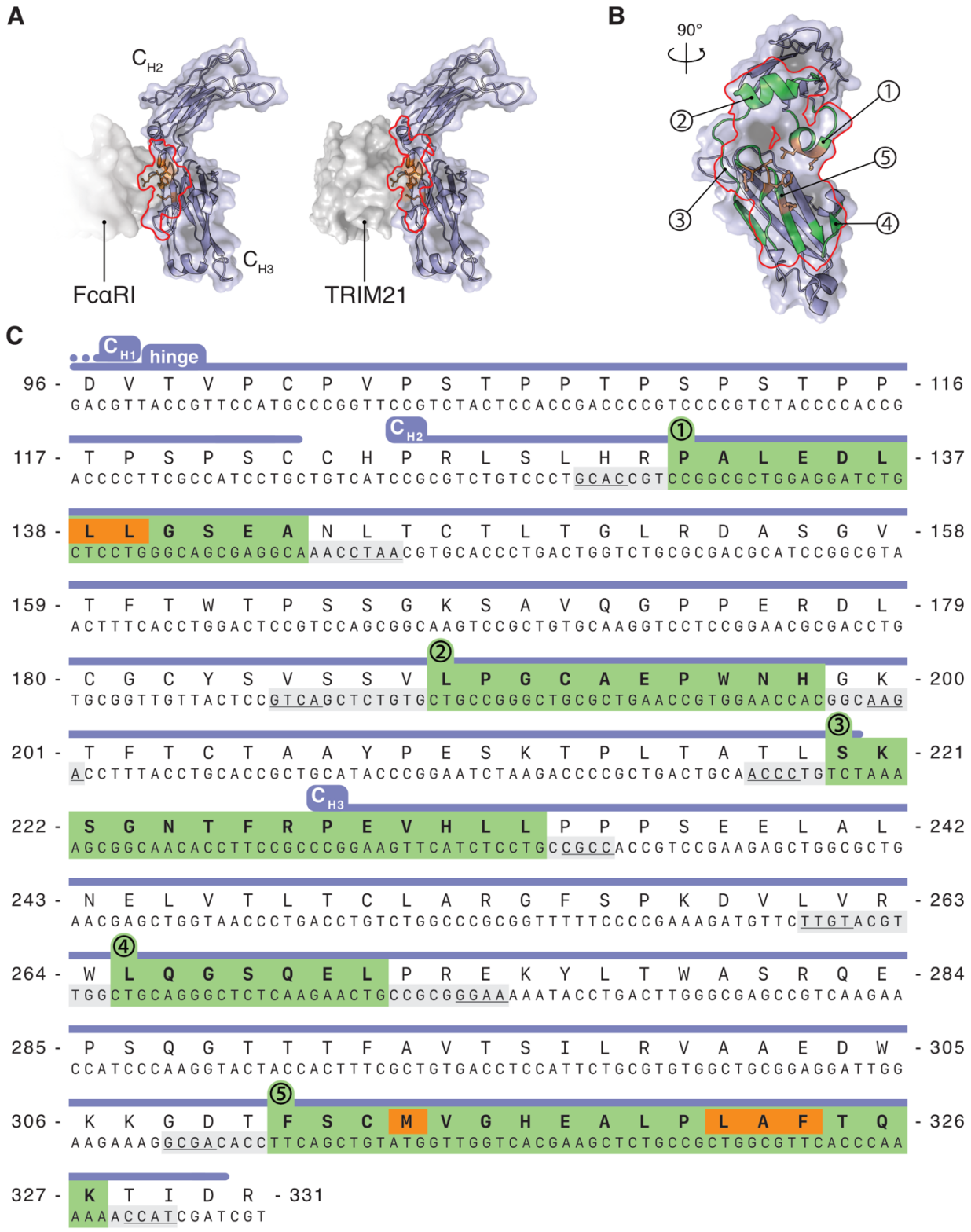


Figure 5.2. Generating an IgA1 Fc deep mutational scanning (DMS) library. **A)** IgA1 Fc interface (red outline) for FcαRI (PDB ID: 1OW0) and TRIM21 (predicted, based on PDB ID: 2IWG). Orange residues are known to be important for affinity for FcαRI. **B)** The sum interface for FcαRI and TRIM21 (red outline) includes five

nonlinear regions (green, 1-5) that include previously identified residues (orange). **C)** IgA1 Fc CDS (UniProt: P01876), beginning with the hinge domain and terminating after C_{H3} domain (purple). Residues targeted for DMS are bold highlighted in green, encompassing some known influential residues (orange). Targeted regions are synthesised as oligonucleotide pools, the bounds of which are grey. Underlined are the overhangs used to achieve scarless assembly of the full CDS.

Table 1. Transformation yields for IgA Fc library construction. Yield represents the total approximate colonies harvested derived from dilution series, and assumed depth of total, targeted, non-redundant NNK variants.

	Cloned duplex	Theoretical Diversity	Yield (depth)
Single NNK	IgA1	67,584	1.6×10^7 (236.8×)
	IgA2	46,080	1.4×10^7 (303.9×)
	IgA3	93,184	2.2×10^6 (24.4×)
	IgA4	21,504	$>1.7 \times 10^7$ (>800×)
	IgA5	139,264	3.9×10^6 (28.0×)
Dual NNK	IgA1-2	122,880	5.1×10^6 (41.6×)
	IgA1-3	172,032	1.6×10^6 (9.3×)
	IgA1-4	86,016	1.5×10^6 (17.7×)
	IgA1-5	208,896	4.5×10^6 (21.5×)
	IgA2-3	143,360	2.7×10^6 (19.0×)
	IgA2-4	71,680	1.8×10^6 (25.7×)
	IgA2-5	174,080	6.3×10^6 (36×)
	IgA3-4	100,352	1.4×10^7 (138.2×)
	IgA3-5	248,192	1.2×10^7 (49.1×)
	IgA4-5	121,856	7.1×10^6 (58.3×)

During the library construction process, we again quantified the depth of assembly of each of the unique parts required for dual DMS (15 in total). We obtained a minimum of 9-fold coverage of diversity, which provides confidence most of the diversity is captured even prior to deep sequencing (Table 1). The final IgA library has been stored in glycerol stocks ready for panning against IgA Fc receptors.

Conclusions and future directions

This thesis has developed new approaches using traditional phage display methods to improve protein engineering workflows. The ability to control diversity with oligonucleotides enables a range of applications in vaccine epitope mapping and antibody affinity maturation. Furthermore, the libraries are well attuned for subsequent computational analyses with machine learning where applicable, compounding their utility. Phage display remains an effective compromise between library sizes, affordability, and protein expression limitations for functional protein screens.

References

Bakema, J. E., & van Egmond, M. (2011). The human immunoglobulin A Fc receptor FcαRI: A multifaceted regulator of mucosal immunity. *Mucosal Immunology*, *4*(6), 612–624. https://doi.org/10.1038/mi.2011.36

Boross, P., Lohse, S., Nederend, M., Jansen, J. H. M., van Tetering, G., Dechant, M., Peipp, M., Royle, L., Liew, L. P., Boon, L., van Rooijen, N., Bleeker, W. K., Parren, P. W. H. I., van de Winkel, J. G. J., Valerius, T., & Leusen, J. H. W. (2013). IgA EGFR antibodies mediate tumour killing in vivo. *EMBO Molecular Medicine*, *5*(8), 1213–1226. https://doi.org/10.1002/emmm.201201929

Breedveld, A., & van Egmond, M. (2019). IgA and FcαRI: Pathological Roles and Therapeutic Opportunities. *Frontiers in Immunology*, *10*, 553. https://doi.org/10.3389/fimmu.2019.00553

Burn, G. L., Foti, A., Marsman, G., Patel, D. F., & Zychlinsky, A. (2021). The Neutrophil. *Immunity*, *54*(7), 1377–1391. https://doi.org/10.1016/j.immuni.2021.06.006

Chen, G., Dubrawsky, I., Mendez, P., Georgiou, G., & Iverson, B. L. (1999). In vitro scanning saturation mutagenesis of all the specificity determining residues in an antibody binding site. *Protein Engineering, Design and Selection*, *12*(4), 349–356. https://doi.org/10.1093/protein/12.4.349

Clift, D., McEwan, W. A., Labzin, L. I., Konieczny, V., Mogessie, B., James, L. C., & Schuh, M. (2017). A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell*, *171*(7), 1692-1706.e18. https://doi.org/10.1016/j.cell.2017.10.033

Gomes, M. M., Wall, S. B., Takahashi, K., Novak, J., Renfrow, M. B., & Herr, A. B. (2008). Analysis of IgA1 N-glycosylation and its contribution to FcαRI binding. *Biochemistry*, *47*(43), 11285–11299. https://doi.org/10.1021/bi801185b

Herr, A. B., Ballister, E. R., & Bjorkman, P. J. (2003). Insights into IgA-mediated immune responses from the crystal structures of human FcαRI and its complex with IgA1-Fc. *Nature*, *423*(6940), 614–620. https://doi.org/10.1038/nature01685

Huls, G., Heijnen, I. A., Cuomo, E., van der Linden, J., Boel, E., van de Winkel, J. G., & Logtenberg, T. (1999). Antitumor immune effector mechanisms recruited by phage display-derived fully human IgG1 and IgA1 monoclonal antibodies. *Cancer Research*, *59*(22), 5778–5784.

Kelton, W., Mehta, N., Charab, W., Lee, J., Lee, C., Kojima, T., Kang, T. H., & Georgiou, G. (2014). IgGA: A "Cross-Isotype" Engineered Human Fc Antibody Domain that Displays Both IgG-like and IgA-like Effector Functions. Chemistry & Biology, 21(12), 1603–1609. https://doi.org/10.1016/j.chembiol.2014.10.017

Matlung, H. L., Babes, L., Zhao, X. W., van Houdt, M., Treffers, L. W., van Rees, D. J., Franke, K., Schornagel, K., Verkuijlen, P., Janssen, H., Halonen, P., Lieftink, C., Beijersbergen, R. L., Leusen, J. H. W., Boelens, J. J., Kuhnle, I., van der Werff Ten Bosch, J., Seeger, K., Rutella, S., … van den Berg, T. K. (2018). Neutrophils Kill Antibody-Opsonized Cancer Cells by Trogoptosis. *Cell Reports*, *23*(13), 3946-3959.e6. https://doi.org/10.1016/j.celrep.2018.05.082

Meyer, S., Nederend, M., Jansen, J. H. M., Reiding, K. R., Jacobino, S. R., Meeldijk, J., Bovenschen, N., Wuhrer, M., Valerius, T., Ubink, R., Boross, P., Rouwendal, G., & Leusen, J. H. W. (2016). Improved in vivo anti-tumor effects of IgA-Her2 antibodies through half-life extension and serum exposure enhancement by FcRn targeting. *mAbs*, *8*(1), 87–98. https://doi.org/10.1080/19420862.2015.1106658

Mkaddem, S. B., Christou, I., Rossato, E., Berthelot, L., Lehuen, A., & Monteiro, R. C. (2014). IgA, IgA receptors, and their anti-inflammatory properties. *Current Topics in Microbiology and Immunology*, *382*, 221–235. https://doi.org/10.1007/978-3-319-07911-0_10

Mullard, A. (2021). FDA approves 100th monoclonal antibody product. *Nature Reviews Drug Discovery*, *20*(7), 491–495. https://doi.org/10.1038/d41573-021-00079-7

Pasquier, B., Launay, P., Kanamaru, Y., Moura, I. C., Pfirsch, S., Ruffié, C., Hénin, D., Benhamou, M., Pretolani, M., Blank, U., & Monteiro, R. C. (2005). Identification of FcαRI as an Inhibitory Receptor that Controls Inflammation: Dual Role of FcRγ ITAM. *Immunity*, *22*(1), 31–42. https://doi.org/10.1016/j.immuni.2004.11.017

Plomp, R., de Haan, N., Bondt, A., Murli, J., Dotz, V., & Wuhrer, M. (2018). Comparative Glycomics of Immunoglobulin A and G From Saliva and Plasma Reveals Biomarker Potential. *Frontiers in Immunology*, *9*, 2436. https://doi.org/10.3389/fimmu.2018.02436

Rifai, A., Fadden, K., Morrison, S. L., & Chintalacharuvu, K. R. (2000). The N-glycans determine the differential blood clearance and hepatic uptake of human immunoglobulin (Ig)A1 and IgA2 isotypes. *The Journal of Experimental Medicine*, *191*(12), 2171–2182. https://doi.org/10.1084/jem.191.12.2171

Roberts, A. W. (2005). G-CSF: A key regulator of neutrophil production, but that's not all! *Growth Factors*, *23*(1), 33–41. https://doi.org/10.1080/08977190500055836

Rosales, C. (2018). Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types? *Frontiers in Physiology*, *9*, 113. https://doi.org/10.3389/fphys.2018.00113

Rouwendal, G. J., van der Lee, M. M., Meyer, S., Reiding, K. R., Schouten, J., de Roo, G., Egging, D. F., Leusen, J. H., Boross, P., Wuhrer, M., Verheijden, G. F., Dokter, W. H., Timmers, M., & Ubink, R. (2016). A comparison of anti-HER2 IgA and IgG1 in vivo efficacy is facilitated by high N-glycan sialylation of the IgA. *mAbs*, *8*(1), 74–86. https://doi.org/10.1080/19420862.2015.1102812

Steffen, U., Koeleman, C. A., Sokolova, M. V., Bang, H., Kleyer, A., Rech, J., Unterweger, H., Schicht, M., Garreis, F., Hahn, J., Andes, F. T., Hartmann, F., Hahn, M., Mahajan, A., Paulsen, F., Hoffmann, M., Lochnit, G., Muñoz, L. E., Wuhrer, M., … Schett, G. (2020). IgA subclasses have different effector functions associated with distinct glycosylation profiles. *Nature Communications*, *11*(1), 120. https://doi.org/10.1038/s41467-019-13992-8

Sterlin, D., & Gorochov, G. (2021). When Therapeutic IgA Antibodies Might Come of Age. *Pharmacology*, *106*(1–2), 9–19. https://doi.org/10.1159/000510251

van der Steen, L., Tuk, C. W., Bakema, J. E., Kooij, G., Reijerkerk, A., Vidarsson, G., Bouma, G., Kraal, G., de Vries, H. E., Beelen, R. H. J., & van Egmond, M. (2009). Immunoglobulin A: FcαRI Interactions Induce Neutrophil Migration Through Release of Leukotriene B4. *Gastroenterology*, *137*(6), 2018-2029.e3. https://doi.org/10.1053/j.gastro.2009.06.047

van Egmond, M., van Vuuren, A. J., Morton, H. C., van Spriel, A. B., Shen, L., Hofhuis, F. M., Saito, T., Mayadas, T. N., Verbeek, J. S., & van de Winkel, J. G. (1999). Human immunoglobulin A receptor (FcalphaRI, CD89) function in transgenic mice requires both FcR gamma chain and CR3 (CD11b/CD18). *Blood*, *93*(12), 4387–4394.

van Tetering, G., Evers, M., Chan, C., Stip, M., & Leusen, J. (2020). Fc Engineering Strategies to Advance IgA Antibodies as Therapeutic Agents. *Antibodies*, *9*(4), Article 4. https://doi.org/10.3390/antib9040070

Wehrli, M., Schneider, C., Cortinas-Elizondo, F., Verschoor, D., Boligan, K. F., Adams, O. J., Hlushchuk, R., Engelmann, C., Daudel, F., Villiger, P. M., Seibold, F., Yawalkar, N., Vonarburg, C., Miescher, S., Lötscher, M., Kaufmann, T., Münz, C., Mueller, C., Djonov, V., … Gunten, S. von. (2020). IgA Triggers Cell Death of Neutrophils When Primed by Inflammatory Mediators. *The Journal of Immunology*, *205*(10), 2640–2648. https://doi.org/10.4049/jimmunol.1900883

Appendix A

Supplementary Tables

Table S1. Amino acid sequence of the TeeVax3 antigen. The underlined sequence is the N-terminal epitope tag

>TeeVax3	MSYYHHHHHHDYDIPTTENLYFQGAMGSPSITKKVTGTIDDVNKKTTSLGSVLSYSLT FELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGI AKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTY DNLDKKPDKGNGITSKEDSKIVYTYRSGSGSGKNTTVNEPKVDKDVTKLGKDDDTYQ IGDKITWFLKSTVPSNIKTLDKFGFTDTLNKGLSFIGDKTQTVTKVQFGTTVLSPDTDY TVEILDSKLTVSLTSAGIEKVSGLVASKQLITEAEKLYKAEDNTDEAAFLSVEVNAKLN ADAVMGSRIENDVELDYGHESDIYKSKVPTNEVPEVHTRSGSGSGPKPGKDVKELGLN HSSYNIGERFSWFLKGTVPKNMLDYEKYSFTDTLDSQLDFISVKSVKYGSQILEKNND YTLSEPTAQNRTLKVELTEAGIKKVAGLYPDRQEVLDTEIEAIKENTDQKPFLEVEFETN INSTVILGKPVTNEVKIEFDNKPDKIAKPVTTPPSDNPEVHTRSGSGSGPTISKSITKSTK DGDKDTASVGEKVDYKLTVQLPSYSKDAINKTVFITDKLSQGLTFLPKSLKIIWNGQTL TKVNEEFKAGDKVIAQLKVENNGFNLNFNYDNLDNHAPEVNYSALLNENAVVGKGG NDNNVDYYYSNNPNKGETHKTTEKPKEGEGTGITKKTDKKTVYTYRSGSGSGESSHK TDVVIHKIKMTSLKGWPKEKNPDGTYTGLGDKNYNGEKIDTITSYFGEGAEELDGVSF TYWSVDKEKYKKLTKNPQNYDTVPKMKAFLQGTEKNKALENSSETIDGKTTGHTAD KGGVKVKDLADGYYWFVENSGSNIANGETLSSSAAVPFGLELPVYKADGSTITELHVY PKNTTTKRS*

Appendix A
Supplementary Figures

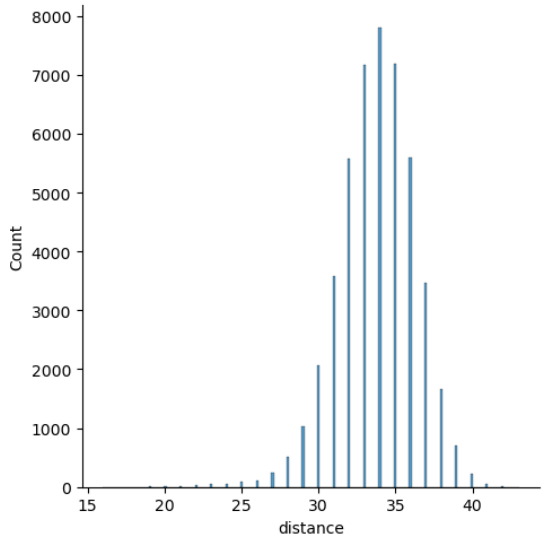


Figure S1. Levenstein distances for tiles in library pool. Each tile was compared with every other tile in the library to compute the number of base changes required to convert one tile into another. The minimum Levenstein distance was 16 with an average of 33.8.

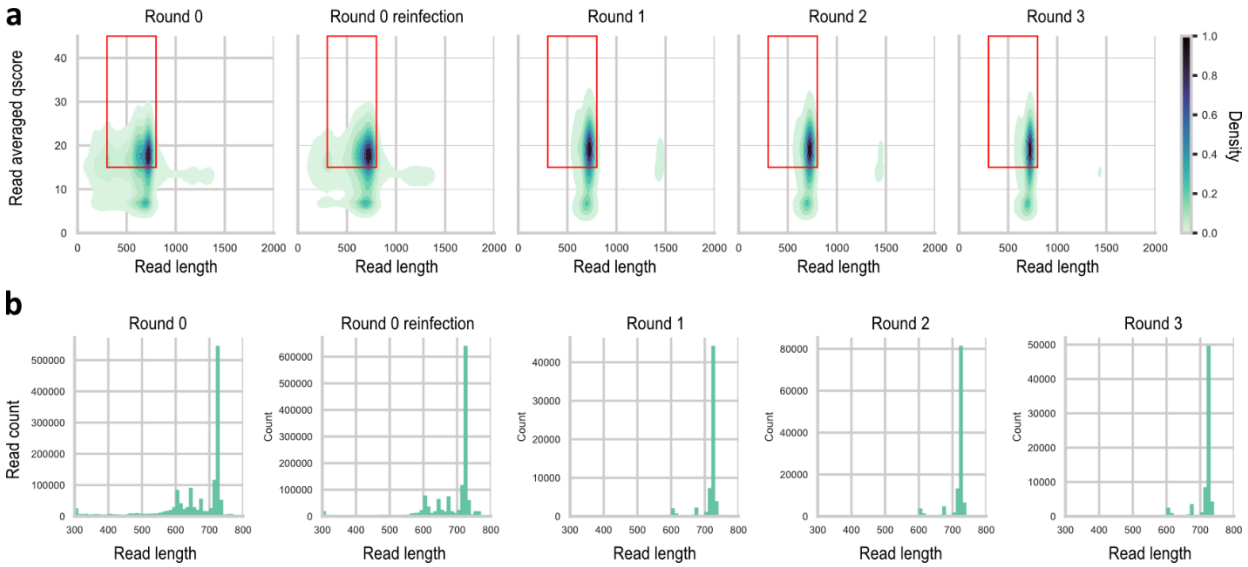


Figure S2. Nanopore sequencing QC data. (a) Density plots of qscores averaged across each read are plotted against read lengths for pre-panning libraries and for each round of panning. The red box indicates the data selected for analysis. (b) Histograms of read counts at each read length for pre-panning libraries and for each round of panning.

Appendix A

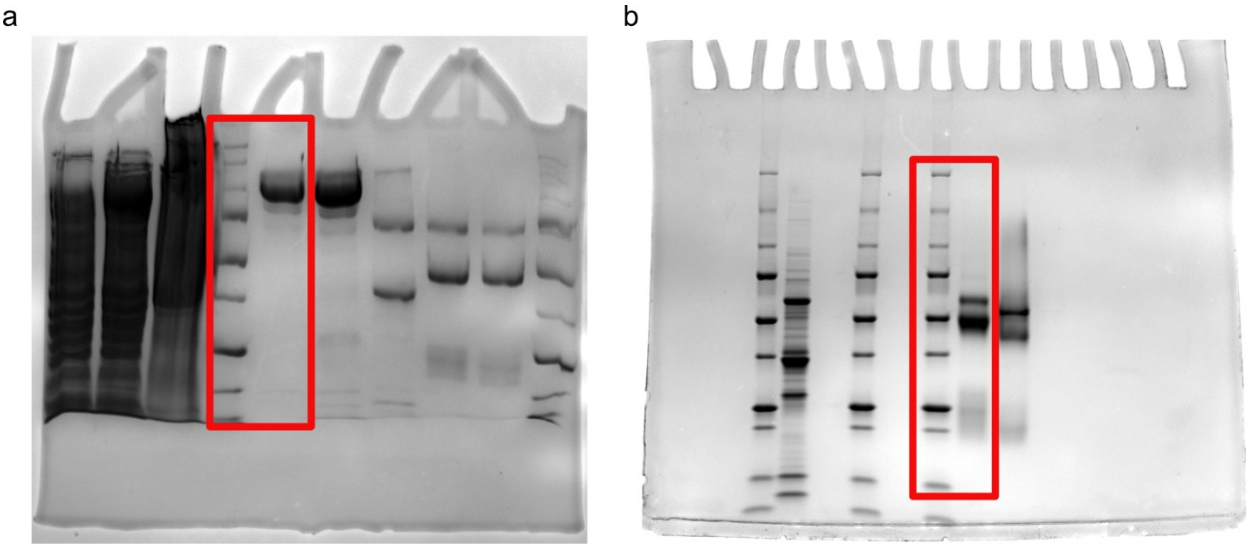


Figure S3. Original SDS-PAGE gel images. (a) Teevax3 protein expression gel. **(b)** Serum purified rabbit anti-TeeVax3 antibodies. Red boxes indicate regions show in main text Figure 3

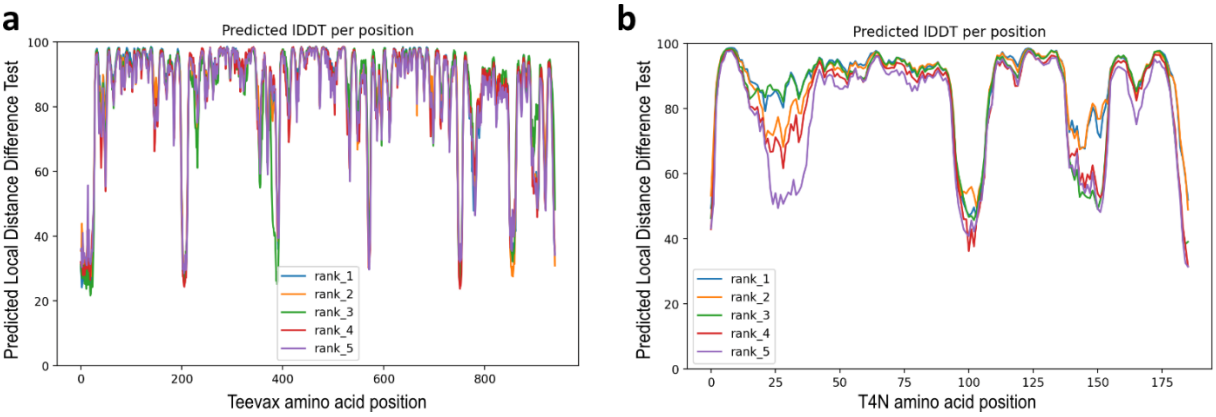


Figure S4. Structural models of TeeVax3 produced using AlphaFold2 software. (a) AlphaFold prediction confidence (Local Distance Difference Test scores) across the full length TeeVax3 amino acid sequence. (b) AlphaFold prediction confidence (Local Distance Difference Test scores) across the T4N domain alone. To improve the accuracy of the C-terminal T4N domain, the PDB structure 3RPK[1] from the Full-Length Major Pilin RrgB from *Streptococcus pneumoniae* was provided as a template and modeled independently. After alignment in Pymol, the rank_1 prediction was used to create a composite model with the other TeeVax3 domains.

References

1. Paterson, N. G. & Baker, E. N. Structure of the Full-Length Major Pilin from Streptococcus pneumoniae: Implications for Isopeptide Bond Formation in Gram-Positive Bacterial Pili. *PLOS ONE* **6**, e22095 (2011).

Appendix B

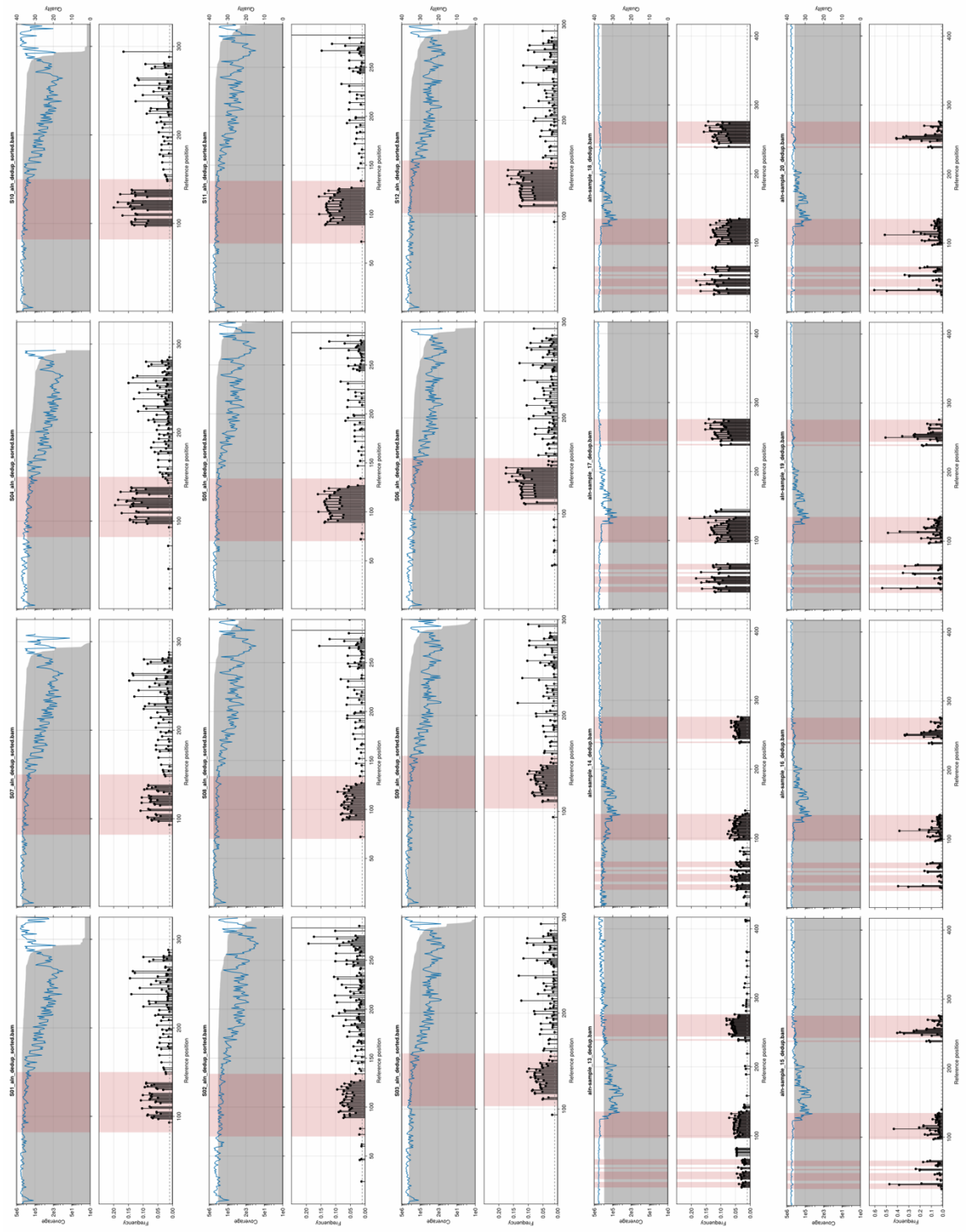


Figure S1. Alignment of deduplicated (post-QC; pre-correction) reads. Coverage (grey fill), quality (blue line) across the reference positions, with mismatches (black stems) displayed per position where frequency > threshold (0.1). No insertions or deletions observed > threshold. Highlighted in red are the duplexed oligo part (S01-12), or targeted residues (S13-20). S01, 07 CDR1 single; S02, 08 CDR2 single; S03, 09 CDR3 single; S04, 10 CDR1 dual; S05, 11 CDR2 dual, S06, 12 CDR3 dual; S13, 14 Dual DMS Pre; S15, 16 Dual DMS post; S17, 18 CMB pre; S19, 20 CMB post.

Appendix B

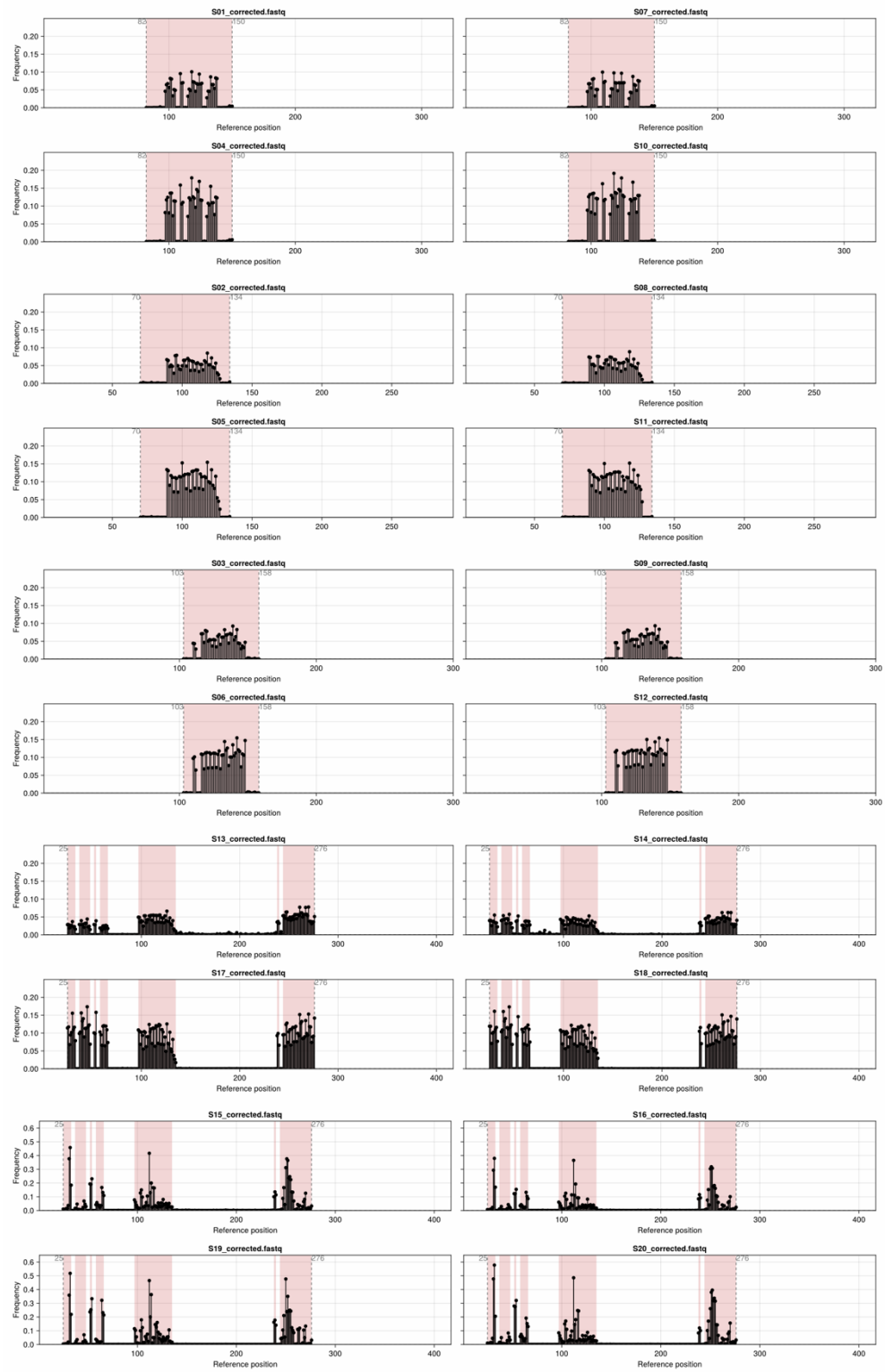


Figure S2. Alignment of corrected reads (Q<20 reverted to WT nt). Mismatches (black stems) displayed per position. Highlighted in red are the duplexed oligo part (S01-12), or targeted residues (S13-20). Read trimming boundaries are also shown (grey dashed line, text). S01, 07 CDR1 single; S02, 08 CDR2 single; S03, 09 CDR3 single; S04, 10 CDR1 dual; S05, 11 CDR2 dual, S06, 12 CDR3 dual; S13, 14 Dual DMS Pre; S15, 16 Dual DMS post; S17, 18 CMB pre; S19, 20 CMB post.

Appendix C

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.	Chapter 1 | Deep mutational scanning for therapeutic antibody engineering
Nature of contribution by PhD candidate	Manuscript writing and editing, Figure creation
Extent of contribution by PhD candidate (%)	60

CO-AUTHORS

Name	Nature of Contribution
Mason Minot	Manuscript writing and editing
Annmaree Warrender	Manuscript writing and editing
William Kelton	Manuscript writing and editing
Sai Reddy	Manuscript writing and editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Mason Minot		
Annmaree Warrender	AWarrender	13/05/2025
William Kelton	wjkelton	13/05/2025
Sai Reddy		

July 2015

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work. Chapter 2 | A rapid approach for linear epitope vaccine profiling reveals unexpected epitope tag immunogenicity

Nature of contribution by PhD candidate	Bioinformatics analysis, Collection of experimental data, Manuscript writing and editing
Extent of contribution by PhD candidate (%)	40

CO-AUTHORS

Name	Nature of Contribution
Kirsten Browne-Cole	Collection of experimental data, Bioinformatic analysis, Manuscript writing and editing
Kevin Beijerling	Collection of experimental data, manuscript editing
Meghan Rousseau	Preparation of key reagents, manuscript editing
Jacelyn Loh	Supply of reagents, conceptialisation, manuscript editing
William Kelton	Conceptialisation, manuscript writing and editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Kirsten Browne-Cole	Kirsten Browne-Cole	14/05/2025
Kevin Beijerling	Kevin B	14/05/2025
Meghan Rousseau		13/05/2025
Jacelyn Loh	Jacelyn Loh	14/05/25
William Kelton	wjkelton	13/05/2025

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3 | Simple high-throughput encoding of deep mutational scanning libraries by oligo-based golde

Nature of contribution by PhD candidate	Conceptialisation, Collection of experimental data, Bioinformatic analysis, Manuscript writing and editing
Extent of contribution by PhD candidate (%)	80

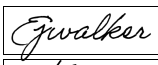
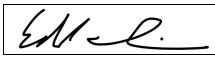
CO-AUTHORS

Name	Nature of Contribution
Emma Walker	Collection of experimental data
Kevin Beijerling	Collection of experimental data
Edward Irvine	Collection of experimental data - sequencing
Jordan Steel	Collection of experimental data
William Kelton	Conceptialisation, Collection of experimental data, Manuscript writing and editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Emma Walker		13/05/2025
Kevin Beijerling		14/05/2025
Edward Irvine		13/05/2025
Jordan Steel		14/05/2025
William Kelton		13/05/2025

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

School of Graduate Research
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 5096
Email: SGR@waikato.ac.nz
Website: http://www.waikato.ac.nz/students/research-degree

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work**. Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4 | Creation of highly diverse IgA libraries by Golden Gate assembly

Nature of contribution by PhD candidate	Collection of experimental data, Manuscript writing and editing
Extent of contribution by PhD candidate (%)	90

CO-AUTHORS

Name	Nature of Contribution
William Kelton	Conceptialisation, Collection of experimental data, Manuscript writing and editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
William Kelton	wjkelton	13/05/2025