ENGINEERING DEEP SEQUENCING-GUIDED PLATFORMS TO EVALUATE SEQUENCE-FUNCTION RELATIONSHIPS BETWEEN PROTEINS FOR THE DEVELOPMENT OF THERAPEUTIC ANTIBODIES

By

Angélica V. Medina-Cucurella

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemical Engineering – Doctor of Philosophy

2019

ABSTRACT

ENGINEERING DEEP SEQUENCING-GUIDED PLATFORMS TO EVALUATE SEQUENCE-FUNCTION RELATIONSHIPS BETWEEN PROTEINS FOR THE DEVELOPMENT OF THERAPEUTIC ANTIBODIES

By

Angélica V. Medina-Cucurella

Over the past two decades, monoclonal antibodies (mAbs) have been used as a major class of therapeutic treatments for cancer and autoimmune diseases given their high specificity against a given target antigen. mAbs can work as antagonists by blocking the downstream signaling pathway through receptors or as agonists by boosting the immune system response to direct tumor cell apoptosis. The understanding of the antibody-mediated recognition of pathogens reveals valuable information related to the immune-protective responses within the host organism. Such information has led scientists to develop new effective vaccines and therapeutics. Nevertheless, understanding the physical basis of affinity and specificity in these interactions is a theoretical and experimental challenge. Subsequently, researchers have developed multiple high-throughput approaches, like deep mutational scanning, to identify the relative binding contribution of individual amino acid residues towards the overall antibody:antigen complex. In this dissertation, I present the successful application of deep sequencing-guided engineering platforms to address numerous aims relevant to the protein engineering and antibody discovery field including the understanding of sequence-function relationships between proteins, antibody conformational epitope mapping, and the development of antibody therapeutics.

First, we use our pipeline utilizing comprehensive mutagenesis, yeast surface display, and deep sequencing to gain insights on the interactions between interleukin-31, a cytokine involved in chronic skin inflammations, and its receptors. Identification of the binding sites on interleukin-

31 by its receptors allows the development of antagonist mAbs to inhibit the downstream signaling pathway. In fact, the mapped conformational epitope of a candidate mAb shows that it inhibits the signaling pathway by binding an overlapping site shared between receptors. A significant limitation of sequence-function mapping by the above method is the requirement that the yeast surface displayed target protein be in a conformation recognizable by the antibody. For example, some proteins such as the neurotrophin family display on the yeast surface in a mostly misfolded or inactive conformation. Consequently, we developed a deep sequencing-guided protein engineering workflow to increase the production of folded canine nerve growth factor, a neurotrophin involved in multiple chronic pain conditions. We identified beneficial mutations within the pro-region of the protein that improved the display of mature, conformationally sensitive protein that enabled the determination of conformational epitopes for multiple antagonist mAbs.

Two fundamental limitations in the creation of large mutagenesis libraries using current template-based mutagenesis is the overrepresentation of specific nucleobases and the difficulty of constructing user-defined libraries beyond single site comprehensive codon scanning. We improve on current methods by using unpurified oligo pools to prepare user-define single and double mutagenesis libraries from plasmid DNA. Results demonstrated a near-complete coverage of desired mutations with even representation of nucleobases and few off-target mutations.

Lastly, we present a new method guided by next-generation sequencing for the selection in cell lysate of agonists mAb for OX40, a costimulatory immune receptor. This project was performed as an industrial internship during Summer 2018. Synthesized OX40 antibodies after deep sequencing selection with cell lysate showed higher therapeutic potentials compared to antibodies enriched by the traditional soluble selection method.

"Success means doing the best we can with what we have. Success is the doing, not the getting; in the trying, not the triumph. Success is a personal standard, reaching for the highest that is in us, becoming all that we can be." – Zig Ziglar

I dedicate this thesis to my family – every little bit of me is a piece of you. Thanks for always supported and encouraged me to keep nourishing my passion for science.

ACKNOWLEDGMENTS

I would like to acknowledge the extraordinary mentorship of my advisor, Dr. Tim Whitehead. Without his valuable guidance this dissertation would not have been possible. Thanks for all the experiences I was given through all these years to become a great researcher and quantitative thinker.

To the past and present lab mates – Dr. Caitlin Kowalsky Stein, Dr. Justin Klesmith, Dr. Emily Wrenbeck, and Matthew Faber – thanks for willingness to help and guide me to success in all my projects. Thanks for conveyed me a day-to-day spirit of adventure regarding research. I would also like to express the deepest appreciation to the Plant Biotechnology for Health and Sustainability graduate training program and the Kinesis Foundation for the financial support granted through my graduate studies. Undoubtedly, these supports gave me valuable skills and experiences that have influenced in my future endeavors and career goals.

I want to thank my graduate school friends, for listening all my frustrations and sharing laughs. Lastly, and most importantly, huge thanks to my eternal cheerleaders - my parents, my siblings, and my fiancé, William - for their never-ending long-distance patience, love, and support through my graduate studies and for always believe in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

FACS	fluorescence activated cell sorter
IL-31	interleukin-31
IL-31RA	gp-130 like receptor
K_D	dissociation constant
mAbs	monoclonal antibodies
NGF	nerve growth factor
OSMR	oncostatin-M receptor
PBS-BSA	phosphate buffered saline with bovine serum albumin
PPIs	protein-protein interactions
scFv	single-chain fragment variable
SSM	single-site saturation mutagenesis
YSD	yeast surface display

CHAPTER 1

1. Introduction to antibodies, epitope mapping, and deep sequencing-guided platforms

Portions of this chapter were adapted from "Characterizing protein-protein interactions using deep sequencing coupled to yeast surface display" in *Methods in Molecular Biology* (**2018**) 1764:101-121

1.1. Introduction

Antibody-antigen interactions are central to understanding humoral immune responses. These non-covalent interactions involve hydrogen bonds, hydrophobic interactions, electrostatic interactions, and van der Waals forces. The surface conformational shape, the residues at the interface, and the sensitivity to the environmental conditions are some examples of the physical characteristics that distinguish individual antibody paratopes[1]. The understanding of the antibody-mediated recognition of pathogens reveals valuable information related to protective immune responses within the host organism. Such information may lead to the development of new effective vaccines and therapeutics against multiple human diseases.

1.2. Background

1.2.1. Monoclonal Antibodies

During the past decades, monoclonal antibodies (mAbs) have been used as a major class of therapeutic treatments against cancer and autoimmune diseases in addition to diagnostic and research applications[2]. mAbs maintain efficacy by binding at unique locations within the target antigen. Since 1986, a total of forty-seven therapeutic mAbs have been approved for commercial sale in the United States and Europe with a worldwide sale of approximately 100 billion by the end of 2017[3]. Given the steadily growing demand, it is expected that the market size for mAbs will be worth $130 - 200 billion by 2022[3]. The low risk of unexpected safety concerns and high-specificity towards targets make them an enhanced therapeutic over other alternative options[2]. These molecules can work as antagonists by blocking the signaling pathway through receptors or as agonists by boosting the immune system response to direct tumor cell apoptosis[4]. Nevertheless, adequate experimental data is required to consider that a therapeutic mAb is capable of mollifying the disease being treated.

1.2.2. Epitope Mapping

An important step for candidate mAb evaluation is to determine the epitope that they target. An epitope is the site of the antigen that is recognized by the antibody. The epitopes can be presented as a linear or as a conformational epitope. A linear epitope is composed of residues that are adjacent in the polypeptide chain while a conformational epitope is comprised of residues located in different fragments of the protein. Multiple epitope mapping studies suggest that natural antibodies recognize a conformational epitope as linear epitopes are not always exposed in the quaternary structure of the antigen[5,6]. The epitope characterization can be used to provide a basis for structural vaccine design, to engineer and improve antibody affinity maturation, to develop agonist molecules for cancer therapy and to predict antigen structure, among others[7–9]. Despite multiple studies over the years, the understanding of the physical basis of affinity and specificity in protein-protein interactions (PPIs) is a theoretical and experimental challenge. Not all residues contribute to the same magnitude in the binding interaction. The residues that show a strong binding interaction give guidance to understand how antibodies can potentially neutralize an infection – a key component in the development of therapeutic mAbs. Consequently, researchers have developed multiple innovative techniques to identify neutralizing epitopes in faster, inexpensive, and high throughput approaches[10–12]. Indeed, most of these methods rely on a powerful high-throughput technology called deep mutational scanning[13] to examine the sequence-function relations between PPIs.

1.2.3. Deep Mutational Scanning

Delineating the sequence determinants of stability, affinity, and specificity of PPIs has been a major research goal for decades. The classical approach to study PPI sequence-function relationships has been “alanine-scanning” mutagenesis in which residues are individually mutated

to alanine and assayed by assorted biophysical techniques[14,15]. The change of binding affinity upon mutation gives a reasonable measure of the importance of the perturbed residue. However, such classical mutagenesis and screening studies are extremely labor-intensive. For example, these methods are limited to the characterization of each mutant separately[14].

Recently, transformative methods utilizing large-scale mutagenesis, surface display, and next generation sequencing, NGS, have been developed to obtain relative binding contributions of individual residues by testing thousands of PPI mutants in a single experiment [8,10–13,16]. The change of binding affinity upon a mutation gives a reasonable measure of the importance of the perturbed residues. All methods share a similar framework: a population containing mutants of one PPI partner is prepared and cloned into a surface-display vector. The population is selected and/or screened for positive or negative binding to the other partner, and then the selected and unselected populations are deep sequenced and analyzed. Finally, the change in frequency for each library member is calculated and converted to a relative binding score[16]. In the method developed by our lab[12] we utilize yeast surface display (YSD)[17,18] as it affords quantitative screening via fluorescence-activated cell sorting (FACS)[5,10,11]. Yeast display has become one of the powerful platforms used for the isolation, the engineering, and the epitope mapping of antibodies and other proteins[17]. In the most common YSD technique, the target protein is displayed as an Aga2 fusion protein, which is attached to yeast cell wall to an Aga1 protein through disulfide bonds (**Figure 1a**). Then, the protein expression is typically detected using a fluorescent conjugated antibody against either hemagglutinin or a c-cmyc tag using flow cytometry. Then, the interaction with a biotinylated secondary protein is detected with a secondary fluorescent conjugated antibody against biotin (**Figure 1b**). This scheme of two-color labeling reaction identifies the stability and affinity of a protein-protein interaction simultaneously, one of the many advantages that yeast

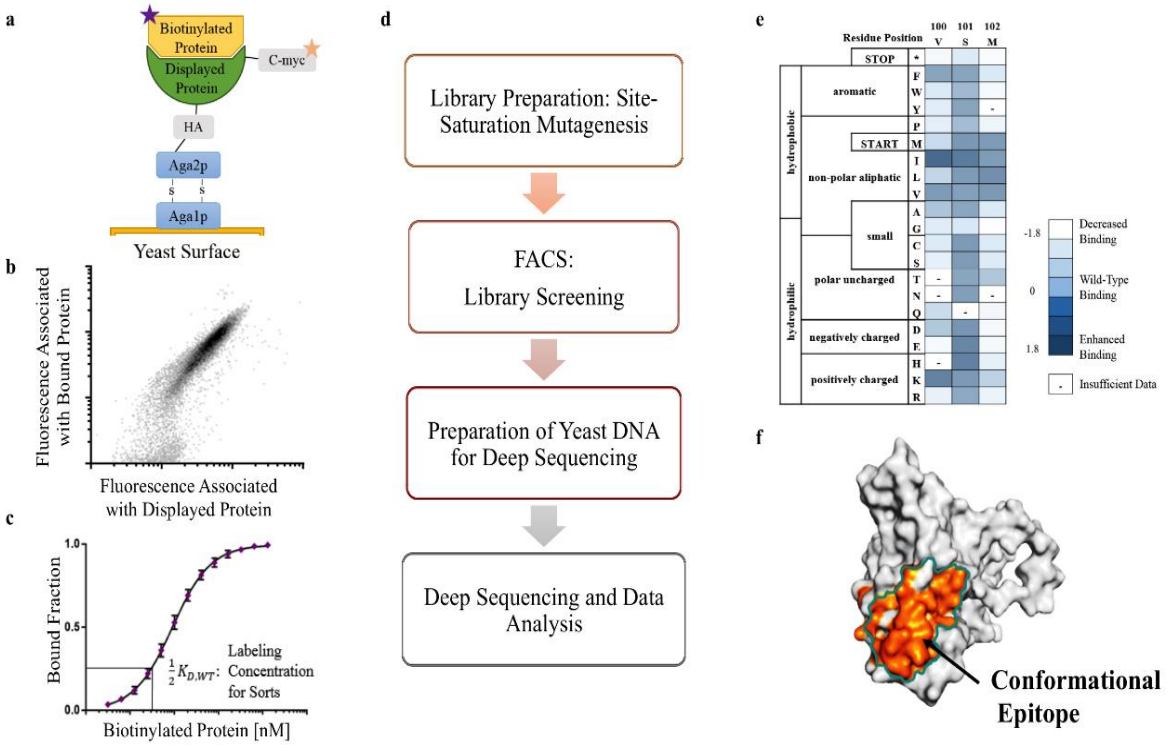


Figure 1. A streamlined process required for PPI characterization using deep sequencing and mutagenesis analysis. (**a.-c.**) Requirements for the pipeline: the binding activity of two proteins is measured using yeast surface display coupled to flow cytometry, and the relative dissociation constant is determined using yeast clonal titrations. The top panel is adapted from Chao et al.[17] (**d.**) The workflow covered in this chapter to characterize protein-protein interactions. (**e.-f.**) Deep sequencing results can be visualized as a heatmap and used to determine the conformational epitope of one member of the interaction.

surface display offers. Furthermore, the proteins can bind at different concentrations allowing the researchers to determine the dissociation constant, K_D, following Hill kinetics (**Figure 1c**). Compared with competing methods using YSD[10,11,19], our approach is faster and less expensive albeit with a limited dynamic range of approximately 10-fold change in binding affinity centered about the wild-type sequence. Accordingly, our method is suitable for fine maturation of PPI affinity and specificity, or to determine fine conformational epitopes. As an example, and as mentioned before, this method can be used to determine if a mAb can function as an antagonist to inhibit the downstream signaling pathway of an antigen. This goal can be achieved by understanding the interactions between a target antigen and its receptors. **Chapter 2** presents a

study that aims to measure the intermolecular interactions between a cytokine involves in chronic skin inflammation, interleukin-31 (IL-31), and its receptors. To evaluate these interactions, we determined the binding sites of each receptor using a predicted feline IL-31 structure model combined with our workflow. Interesting, these sites largely overlapped with the previous sites determined by computational analysis and alanine scanning mutagenesis[20]. However, our binding sites revealed a new overlapping site between both receptors not described before. We also mapped the conformational epitope for an anti-interleukin-31 mAb. Our constructed epitope revealed that this mAb could potentially antagonize feline IL-31 signaling pathway by (i.) binding to a shared site between both receptors, and (ii.) decreasing the binding signals of both receptors. This work is a further evidence of the power of our method for studying the sequence-function relationship between PPIs and for mapping conformational epitopes of potential neutralizing antibody candidates.

Although yeast surface display is an extraordinary platform to evaluate multiples PPIs, a major limitation of this method is the requirement that the displayed protein must be in a conformation recognizable by the antibody. Some complicated proteins display on the surface of the yeast in a mostly inactive conformation, making conformational epitope mapping measurements impossible. As an example, a previous study showed that a member of the neurotrophin family, brain-derived neurotrophic factor, displayed on the yeast surface in a mostly inactive conformation[21]. **Chapter 3** presents a deep sequencing-guided protein engineering workflow developed to improve the production of folded, displayed nerve growth factor (NGF), a related neurotrophin involved in multiple chronic pain conditions[22]. This protein displayed on the yeast surface but was barely recognized by conformationally sensitive mAbs. Thus, we created mutational libraries in the neurotrophin pro region using an engineering pipeline to enhance the

production of folded, displayed canine NGF. Such libraries revealed new insights into the sequence-function relationships of the neurotrophin pro region and allowed us to generate conformational epitopes maps of multiple anti-NGF mAbs using an engineered NGF construct. This study demonstrated the potential of deep sequencing-guided engineering pipelines to assign a functional effect for every possible nonsynonymous point mutant and to map the conformational epitopes for potential anti-NGF mAbs. Our pipeline could be a promising platform to increase the production of other members of neurotrophin family and other complicated proteins.

Another limitation of these methods is the design of oligonucleotides (oligos) for user-defined mutations used in the synthesis of large-scale mutagenesis libraries. These libraries are often produced by using degenerate oligos which can make up to 63 different mutations per codon substitution. However, there are some experimental biases. For example, solid phase synthesis of degenerate oligonucleotides often introduces an overrepresentation of nucleobases. Multiple vendors have recently developed new microarray-synthesized oligo pools technologies which can be used for synthetic biology designs. Although these pools are synthesized at low attomole amounts, given the very low oligo concentration needed in our mutagenesis protocol, we speculate that oligo pools could be directly used in the reaction. **Chapter 4** presents the integration of a single oligo pool with nicking mutagenesis[23] to construct user-defined single and double mutagenesis libraries for three different targeted proteins with low off-target rates. These oligo pools can be designed to encode every possible single mutation in a more precise and homogeneous manner. Indeed, the results showed that the oligo pool libraries had a much more even representation of all 20 amino acid substitutions compared with the traditional 'NNN' degenerate oligos. The constructed libraries show a near-complete coverage of the programmed mutations with a small percent of non-programmed mutations. This technology can be used for

further studies including conformational epitope mapping, protein engineering directed evolution pipelines, and local fitness landscape evaluation.

Appendix A presents our step-by-step protocol to determine the affinity and specificity for full-length protein binders which can be used to determine fine conformational epitopes[24]. This method combines the creation of single site saturation mutagenesis (SSM) libraries using nicking mutagenesis[23], the transformation into yeast, the screening of SSM libraries using FACS, the DNA preparation for deep sequencing, and the data analysis to determine conformational epitopes as shown in **Figure 1**. After deep sequencing, a relative binding term for each mutant is derived from the change in frequency of the bound population compared with the unselected population[25]. These binding terms are often visualized as heatmaps were a positive value represents a beneficial mutation, a value of zero represent a neutral activity, and a negative value represents a deleterious mutation **(Figure 1e**). Then, the Shannon entropy, a measure of sequence conservation, is calculated for each position. First, we remove the structurally conserved positions. These positions have an entropy less than or equal to the midpoint for the displayed population. Next, we map the conformational epitope by discriminating between conserved and non-conserved positions. A conserved or epitope position will have the Shannon entropy of the bound population less than or equal to the midpoint while non-conserved positions will have an entropy value higher than the midpoint.

1.2.4. Next-generation sequencing in antibody discovery applications

In addition to academic research efforts, biopharmaceutical companies are focused on improving the techniques used to discover human mAbs for therapeutic applications[26]. In the conventional approach of hybridoma screening, a mouse is injected with the target antigen, follow by the collection of produced B cells and fusion with myeloma cells to produce hybridomas.

Although this technology preserves the natural antibody DNA sequence and pairing, the screening process remains inefficient[27]. Alternative approaches are now available through the recent advances offered by next-generation sequencing[26]. One example is the comprehensive antibody discovery pipeline based on a parallel single cell platform created by the biopharmaceutical company GigaGen. Their workflow combines microfluidic methods and multiplex PCR to obtain natively paired heavy and light chain single-chain variable fragment (scFv) libraries from isolated B cells, follow by yeast display of scFv libraries with FACS and deep sequencing to identify clones with the highest affinity[28–31].

Appendix B presents a developed method to screen humanized mouse-derived yeast scFv libraries using recombinant OX40 protein (tumor necrosis factor receptor superfamily member 4, TNFRS4) in cell lysate[31]. Their novel pipeline was used to evaluate two methods for mice immunization and two methods for the selection of OX40 agonists. This project was performed at GigaGen in San Francisco, California as an industrial internship during Summer 2018. While I was conducting my graduate internship at this company, I was responsible for the screening of humanized mouse-derived yeast scFv libraries using recombinant OX40 protein in cell lysate and with soluble OX40 protein. Then, deep sequencing was used to compare the pre- and post-sorted yeast scFv libraries obtained from both selection methods. After that, GigaGen's team was responsible for the expression and characterization of forty-one enriched monoclonal antibodies, and further experiments including cell surface binding, kinetics assays, and *in vitro* activity. This study demonstrated that the cell lysate selection methods yielded OX40 antibodies with higher therapeutic potential compared to the antibodies enriched by the soluble selection method[31].

In the following sections, we will show the successful application of deep sequencing-guided engineering methods to address numerous aims relevant to the protein engineering and antibody discovery field. These methods were applied to improve the understanding of protein sequence-function relationships, antibody-epitope mapping, development of antibody therapeutics, and other end uses. The mutational-point analysis obtained from deep sequencing was suitable to identify nearly all beneficial point mutants in a protein in a simplified workflow and to find the optimal conditions for multiple applications. These contributions prove that effective strategies exist to overcome significant limitations in the field of deep mutational scanning.

CHAPTER 2:

2. Feline Interleukin-31 shares overlapping epitopes with Oncostatin M Receptor and IL-31RA

2.1. Abstract

Interleukin-31 (IL-31) is a major protein involved in severe inflammatory skin disorders. Its signaling pathway is mediated through two type I cytokine receptors, IL-31RA (also known as gp130-like receptor) and Oncostatin M receptor (OSMR). Understanding molecular details in these interactions is crucial for the development of antagonist anti-IL-31 monoclonal antibodies (mAbs) as potential therapies. Previous immunoprecipitation studies suggest that human IL-31 binds to IL-31RA and then recruits OSMR to form a ternary complex. In this model, OSMR cannot interact with IL-31 in the absence of IL-31RA. In this work, we show that feline IL-31 (fIL-31) binds independently with feline OSMR using surface plasmon resonance, ELISA, and yeast surface display. Moreover, competition experiments suggest that OSMR shares a partially overlapping epitope with IL-31RA. To map the binding sites of both receptors on fIL-31, we used deep mutational scanning combining comprehensive mutagenesis of yeast surface displayed fIL-31, fluorescence activated cell sorting, and deep sequencing. The constructed binding site for IL-31RA contains fIL-31 positions E20 and K82, while the binding site for OSMR comprises the "PANDFERK" motif and positions G39 and K100. These sites largely agreed with the previous sites identified for human IL-31 and its receptors determined by computational analysis and alanine scanning mutagenesis. However, our results also revealed a new overlapping site, composed of positions R69, R72, P73, D76, D81, and E97, between both receptors to which we called the "shared site". The conformational epitope of an anti- feline IL-31 mAb that inhibits both OSMR and IL-31RA binding and signaling also mapped to this shared site. Together, our results show conclusively that in felines, IL-31 binds IL-31RA and OSMR independently through a partially shared epitope. These results suggest reexamination of the putative canonical mechanisms for IL-31 signaling in higher animals.

2.2. Introduction

Cytokines comprise a large family of small proteins that play a critical role in the development and control of the immune response. Certain cytokines are associated with the initiation and the persistence of pathological pain behavior including nerve and skin injuries. A more recently discovered cytokine, interleukin-31 (IL-31), has been linked to the induction of chronic skin inflammation[32]. Human and murine data have shown high expression of IL-31 associated with severe inflammatory skin disorders including pruritis, alopecia, skin lesion, and atopic dermatitis (AD) and with other regulated allergic diseases such as asthma[32–38]. An experimental animal model for human AD reported a strong correlation between itch-associated scratching behavior in NC/Nga mice and expression of IL-31 mRNA[36]. Elevated IL-31 serum levels were found in adult patients with AD compare to healthy control subjects[39] and in pediatric patients during AD flare and quiescence[40]. Together, these data suggest that IL-31 represents an important target for the development of treatments against such skin inflammatory diseases. Accordingly, antagonist anti-IL-31 mAbs are currently in development for human health[41,42] and animal health[43]. For example, an anti-hIL-31RA mAb, CIM331, binds to IL-31RA, inhibits IL-31 signaling and reduces severe pruritus[42]. In veterinary medicine a "caninized" anti-IL-31 mAb, Lokivetmab, showed efficacy in clinical trials for canine pruritis and is currently approved as an AD therapy for dogs[43].

IL-31 is a member of the IL-6 cytokine superfamily produced preferentially by T helper type 2 cells[32]. Mature human IL-31 (hIL-31) is composed of 141 amino acids[32] with a predicted topology of four antiparallel helices[20]. The IL-31 signaling pathway is thought to be mediated through a gp130-like type 1 cytokine receptor (IL-31RA, also known as GPL) and oncostatin M receptor (OSMR)[20,32,44,45]. Both receptors belong to the type I cytokine receptor family which share

a common cytokine binding domain (CBD) formed by two fibronectin type III-like domains[46]. Previous studies supplied immunoprecipitation evidence that human IL-31RA (hIL-31RA) binds directly to hIL-31. In these same studies immunoprecipitation results failed to detect direct human OSMR (hOSMR) binding to hIL-31[20,44]. However, an increase in binding was distinguished when hIL-31RA and hOMSR were combined, suggesting that hIL-31 binds first to hIL-31RA, at which time hOMSR is recruited to form the ternary complex[20,44]. In this model, the ternary complex activates numerous downstream signaling pathways[20,32,44,47,48] (**Figure 2a**).

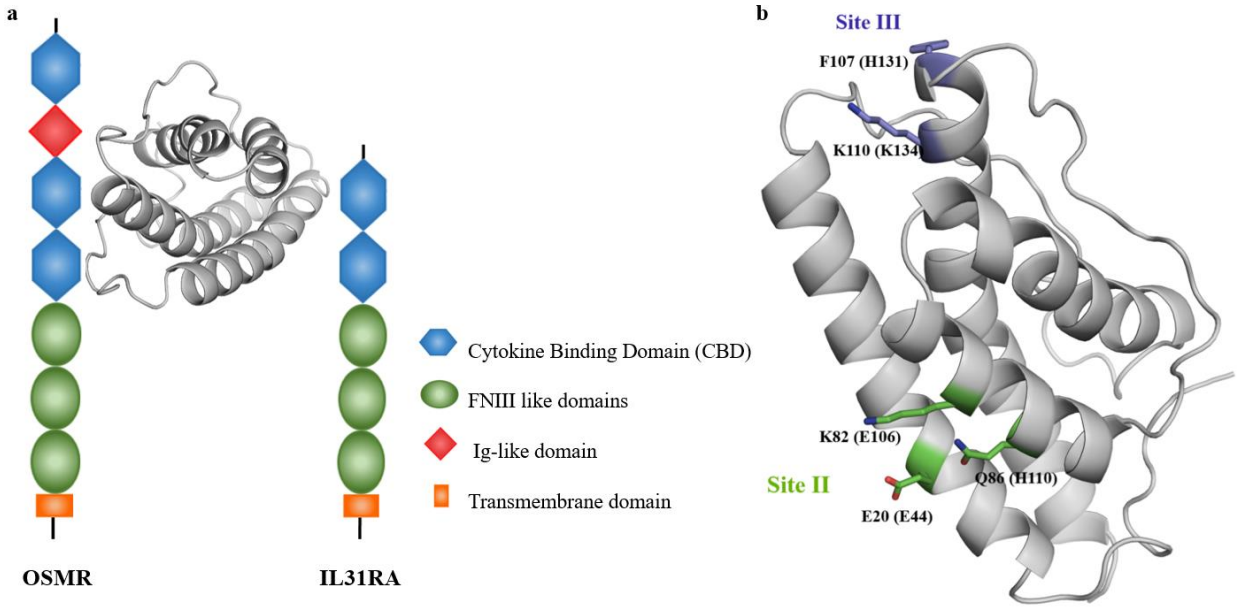


Figure 2. Interactions between IL-31 and its receptors. Adopted from Le Saux et. al[20]. The cartoon model of IL-31 structure was generated using the feline sequence. (**a.**) The ternary complex formed between IL-31, IL-31RA, and OSMR. Different domains of both receptors are indicated in the figure legend. (**b.**) Binding sites for the interaction between human IL-31 (hIL-31) and its receptors as described by Le Saux et at[20]. Residues corresponding to the hIL-31 sequence are indicated in parenthesis.

Specific atomistic knowledge of the binding interactions between IL-31 and its receptors is crucial for the development of antagonist therapeutic mAbs. Based on the structure of IL-6/IL-6 α-Receptor/gp130 complex[49], the IL-6 cytokine superfamily family is thought to interact with their receptors through three different contact binding sites (sites I, II, and III). Le Saux, *et al*. used

computational analysis and sparse alanine scanning to delineate sites II and III only as critical binding sites for the interaction between hIL-31 and its receptors[20]. In particular, Glu44, Glu106, and His110 were identified as critical residues for binding site II while Lys134 was identified within the binding site III (**Figure 2b**).

The aim of this study is to gain insights into the interactions between feline IL-31 (fIL-31) and its feline receptors fOSMR and fIL-31RA and to map the conformational epitope for an anti-fIL-31 mAb (mAb #1). In contrast to previous studies conducted with homologs, we show through multiple biophysical methods that fOSMR directly binds fIL-31 and partially interferes with fIL-31RA binding. We identified the potential binding sites for fOSMR, fIL-31RA, and an anti-fIL-31 mAb (mAb#1) using a predicted fIL-31 structure model combined with fine epitope mapping[24] using yeast surface display[17], nicking mutagenesis[23], and deep sequencing[50]. The constructed binding sites agreed with the sites previously found by Le Saux et *al*[20] and showed an additional overlapping site between both receptors to which we termed the "shared site". Finally, mAb#1 also bound this shared site between both receptors demonstrating its potential to inhibit the signaling pathway. Together, these results suggest that the current model for IL-31 mediated signaling in higher mammals is incomplete and suggest efficient therapeutic strategies to antagonize IL-31-mediated signaling.

2.3. Materials and methods

2.3.1. Strains

The *Escherichia coli* strain used in this study was XL1-Blue (Agilent, Santa Clara, CA*)* endA1 supE44 thi-1 hsdR17 recA1 gyrA96 relA1 lac [F' proAB lacI1^{q}ZLiM15 Tn10 (Tet^{r})]. The *Saccharomyces cerevisiae* strain used in this study was EBY100 (American Type Culture Collection, Manassas, VA) *MATa AGA1::GAL1-AGA1::URA3 ura3-52 trp1 leu2-delta200 his3-*

delta200 pep4::HIS3 prb11.6R can1 GAL.

2.3.2. Plasmid Constructs

pETconNK_fIL-31 was created by inserting a codon-optimized gene encoding the Ser1 – Gln136 from the mature portion of feline interleukin-31 (fIL-31) (GenScript, Piscataway, NJ) into yeast display vector pETconNK[23] (Addgene plasmid #81169) using standard restriction cloning. Sequences were verified by Sanger sequencing (Genewiz, South Plainfield, NJ). The full sequence is listed in **Note C1**.

2.3.3. Preparation of fIL-31, fOSMR, fIL-31RA constructs, and mAb#1

FIL-31 was produced recombinantly in *E. coli* and CHO cells with a C-terminal His-tag and purified by Ni-NTA affinity chromatography. Other proteins were prepared as Fc fusions and produced in mammalian cell culture and purified by protein A affinity chromatography. Each protein was prepared in phosphate-buffered saline (PBS) at a concentration of at least 0.082 mg/mL by Zoetis. Proteins were biotinylated at a molar ratio of 1:20 protein: biotin using the EZ-link NHS-biotin kit following the manufacturer's instruction (Life Technologies, Carlsbad, CA). Biotinylated proteins were desalted using Zeba Spin desalting columns (Thermo Fisher, Waltham, MA) and stored at 4ºC.

2.3.4. Analytical Size Exclusion Chromatography

Size exclusion chromatography was performed using a TSK SuperSW3000 4.6 x 30 mm gel permeation column. 50 μg of fIL-31 was injected at a flow rate of 0.25 mL/min for 25 minutes with a mobile phase of 200 mM sodium phosphate, pH 7.2.

2.3.5. Surface Plasmon Resonance binding assays

Surface Plasmon Resonance was performed on a Biacore T200 (GE Healthcare, Pittsburgh, PA) to measure binding affinities of fOSMR-ECD and mAb#1 to fIL-31. fIL-31 immobilization on CM5 sensor and the binding measurements were conducted as previously described[51]. Data was analyzed with Biacore T200 Evaluation software by using the method of double referencing. The resulting curve was fitted with the 1:1 binding model.

2.3.6. ELISA

The plate was coated with feline IL-31 overnight at 4°C in carbonate-bicarbonate buffer, pH 9.6 (Sigma-Aldrich, St. Louis, MO) follow by a blocking step with 5% skim milk in PBS with 0.05% TWEEN 20 for 1 hour at room temp. Individual proteins were diluted at different concentrations in blocking buffer and were added to the coated plate for 2 hrs at room temp. Next, HRP conjugated secondary antibodies were added to appropriate wells at 1:10000 in blocking buffer for 1 hr at room temp (goat anti-human Fc for receptors, jackson goat anti-cat Fc for mAb#1, KPL anti-mouse). Finally, KPL SureBlue TMO (VWR) was added to develop absorbance at 450 nm.

2.3.7. Yeast Surface Display Expression and Binding Activity

Mean fluorescence intensities (MFI) were measured using a BD Accuri C6 flow cytometer. The expression of fIL-31 on the yeast surface was detected using anti-c-myc-FITC (Miltenyi Biotec, San Diego, CA) and the binding interaction with biotinylated proteins was detected using streptavidin- R- phycoerythrin conjugate (Thermo Fisher). Dissociation constant (K_D) values were determined according to Chao *et al.*[17] Titrations were performed at triplicates on at least two separate days and MFI values were used to calculate the experimental K_D using one site–specific binding equation (Hill coefficient of 1) in Graph Pad Prism software. Labeling concentrations

tested vary from 0.064 nM to 262.1 nM for fOSMR and mAb#1, and from 2.05 nM to 524.3 nM for fIL31RA-1FNIII.

2.3.8. Competition Binding Assays

For YSD $1x10^5$ yeast cells were labeled at twelve times their K_D values with either non-biotinylated fOMSR or non-biotinylated mAb#1 for 30 mins at room temperature in PBS with 1g/liter of bovine serum albumin (PBS-BSA). After spinning down and washing with 200 µl of PBS-BSA, cells were labeled at $1xK_D$ values with either biotinylated mAb#1, fOSMR, or fIL31RA-1FNIII for 30 mins at room temperature in PBS-BSA. Then, after a second step of centrifugation and washing, cells were labeled with 0.6 µl of anti-cymc-FITC and 0.25 µl of streptavidin- R- phycoerythrin conjugate in 49.15 µl of PBS-BSA for 10 mins at 4°C. Cells were washed thoroughly with PBS-BSA and read on a flow cytometer.

2.3.9. Cell-based assays

E. coli-derived or CHO-derived fIL-31 at various concentrations was incubated with the adherent feline macrophage cell line FCWF-4 (ATCC CRL-2787) for 1 hr at 37ºC. Activation of pSTAT signaling was determined using an AlphaLISA *SureFire Ultra* pSTAT3 kit (PerkinElmer). For inhibition experiments, fOSMR-ECD was pre-incubated with 10 µg/mL fIL-31 for 1 hr at 37ºC prior to stimulation before testing for activation of pSTAT signaling.

2.3.10. Preparation of Mutagenesis Libraries

Two comprehensive single-site saturation mutagenesis libraries were constructed using nicking mutagenesis as described[23]. All "NNK" mutagenesis oligos were designed using Quick Change Primer Design Program (www.agilent.com) and were ordered from Integrated DNA Technologies (Coralville, IA). Library 1 covered residues Ser1 – Phe68 and library 2 covered Arg69 – Gln136 of the mature fIL-31. 5 µg of library DNA plasmid was transformed into

chemically competent *Saccharomyces cerevisiae* EBY100 cells, and cells were grown and stored at a concentration of $1x10^7$ cells per ml in yeast storage buffer at -80°C according to published protocols[24].

2.3.11. Determination of receptor binding sites and conformational epitope of mAb#1

The library screening through FACS and deep sequencing preparation was performed exactly as previously described[24]. The sorting was done on an Influx Cell Sorter at the Michigan State University Flow facility. After preparing the plasmid DNA for deep sequencing, libraries were pooled and sequenced on an Illumina MiSeq using 2 x 250bp pair-end reads at the Michigan State University Genomic Sequencing Core facility or the University of Illinois at Chicago DNA Service facility. **Appendix E** contains the sorting conditions (**Table E1**), the primers used for deep sequencing (**Table E2**), and the summary table of statistics (**Table E3**).

2.3.12. Data Analysis

A modified version of Enrich 0.2 software as described in Kowalsky *et al.*[25] was used to compute enrichment ratios from the raw sequencing files. The relative fluorescence (ζ_i) for variant *i* was defined as:

$$\zeta_i = log_2\left(\frac{\bar{F}_i}{\bar{F}_{wt}}\right), \qquad 2.1$$

where $\bar{F}_i$ is the mean fluorescence of variant *i* and $\bar{F}_{wt}$ is the mean fluorescence of wild type fIL-31. This equation can be written in terms of experimental observables according to:

$$\zeta_i = log_2(e)\sqrt{2}\acute{\sigma}\left[erf^{-1}\left(1-\Phi 2^{(\varepsilon_{wt}+1)}\right) - erf^{-1}\left(1-\Phi 2^{(\varepsilon_i+1)}\right)\right] \qquad 2.2$$

where $\acute{\sigma}$ is the log normal fluorescence standard deviation of the clonal population, ε is the enrichment ratio and Φ is the percentage of cells collected by the sorting gate on the flow

cytometer. Custom python scripts available at Github (user: JKlesmith) were used to normalize the relative fluorescence, and to calculate the Shannon Entropy and overall statistics[12].

2.3.13. Data Availability

Raw sequencing reads for this work have been deposited in the SRA (SAMN11289369–72, 79, 81-83, 90-91, SAMN11289422-23).

2.4. Results

2.4.1. fOSMR and fIL-31RA bind independently to fIL-31

We assessed the binding activity of fIL-31 against soluble fOSMR extracellular domain (fOSMR-ECD), four versions of soluble fIL-31RA (fIL31RA-CBD comprising the CBD only, two fIL31RA-1FNIII isoforms comprising CBD and 1 FNIII domain, and fIL31RA-2FNIII comprising CBD and two FNIII domains, **Figure 3a**), and an anti-fIL-31 mAb (mAb#1) (**Figure 3**). All receptors were expressed as fusion proteins with a C-terminal Fc. Recombinant proteins were produced by transient expression in CHO cells and purified by affinity chromatography using Protein A resin. fIL-31 was at the expected molecular weight (**Suppl Figure D1a**) after PNGase treatment. fIL-31 was monomeric with a small dimeric peak as judged by analytical gel permeation chromatography (**Suppl Figure D1b**).

We first assessed the ability of fOSMR-ECD and fIL31-RA to independently recognize soluble fIL-31 using ELISAs. fIL-31 was coated and then receptors or mAb were incubated at indicated concentrations. After washing, receptors were detected with an anti-human IgG. fOSMR-ECD bound fIL-31 in a dose dependent fashion, whereas the fIL-31RA CBD alone did not recognize fIL-31 at any concentration tested (**Figure 3b**). However, both fIL31RA-1FNIII isoforms recognized fIL-31. We also tested mAb#1, which could also bind fIL-31 (**Figure 3b**). Surface plasmon resonance (SPR) measurements using immobilized fIL-31 showed similar

binding results for fOSMR-ECD and mAb#1. (**Figure 3c**). Binding equilibrium measurements were determined using kinetic fitting with 1:1 binding mode, revealing low-nM affinities for both fIL-31-fOSMR-ECD and fIL-31-mAb#1 interactions.

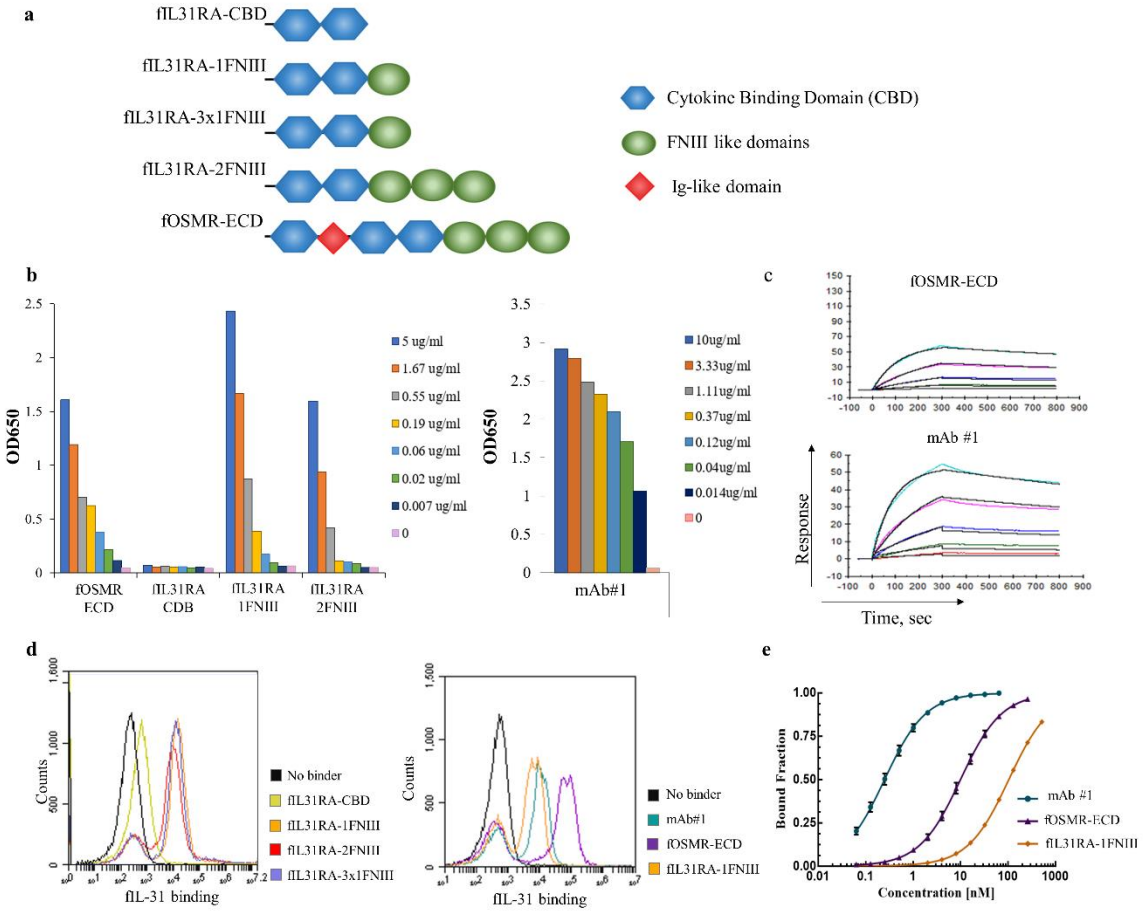


Figure 3. Binding profiles of feline IL-31 (fIL-31) with all different constructs of fIL-31RA, fOMSR-ECD, and mAb#1. (**a.**) fIL-31RA and fOSMR-ECD constructs tested in this work. (**b.**) Surface plasmon resonance sensograms and (**c.**) ELISA plots showing the binding of fIL-31 with all constructs. Done by Zoetis. (**d.**) Flow cytograms showing the increase in fluorescence in the fIL-31 binding channel with all constructs. (**e.**) Binding titrations curves for yeast surface-displayed fIL-31 with mAb#1, fOSMR-ECD, and fIL31RA-1FNIII. Titrations were performed at least in triplicates on different days. (error bars, 1 s.d., n≥3)

We also measured these intermolecular interactions using yeast surface display (YSD), where fIL-31 was displayed on the yeast surface with an N-terminal yeast surface protein Aga2p and a C-terminal *c-myc* epitope tag. Saturating amounts of biotinylated soluble fIL-31RA constructs, fOSMR-ECD, and mAb#1 were incubated with yeast cells displaying fIL-31, followed

by secondary labeling (**Figure 3d**). fOSMR-ECD bound fIL-31 with nanomolar affinity (**Figure 3d-e**). While fIL31RA-CBD did not bind, both fIL31RA-1FNIII isoforms and fIL31RA-2FNIII were able to recognize fIL-31, consistent with ELISA and SPR data. Given that the fIL31RA-1FNIII isoform obtained the highest mean fluorescence intensity (MFI) under saturating amounts of receptor, we used this receptor for the remainder of this work. Dissociation constants ranged from 0.26 ± 0.01 nM for mAb#1 to 106 ± 2.7 nM for IL31RA-1FNIII (**Table 1)**. For all interactions including mAb #1, the binding on the yeast surface could be modeled using 1:1 binding kinetics (best fits of Hill coefficient – 1.0 ± 0.04), consistent with fIL-31 and receptors being mostly monomeric.

Table 1. Summary of experimental determined dissociation constants (KD) using yeast surface display and surface plasmon resonance. Error bars represent standard error of at least 3 independent measurements. ND, not determined.

	Average K_D values [nM]			
Name	**Yeast Surface Display**	**Hill Coefficient**	**Surface Plasmon Resonance**	**Fit**
fOSMR-ECD	10.3 ± 0.5	0.99 ± 0.04	3.8	1:1
fIL31RA-1FNIII	106 ± 2.7	1.00 ± 0.01	ND	1:1
mAb #1	0.26 ± 0.01	0.99 ± 0.04	3.1	1:1

2.4.2. fOSMR partially, but not completely, inhibits binding of fIL31RA-1FNIII to fIL-31

To determine whether the receptors bind to independent sites on fIL-31 or share overlapping sites we performed competition binding assays using SPR and YSD. First, we set-up an SPR-based competitive binding assay where fIL-31 is bound and then either fOSMR-ECD or mAb#1 is captured at saturating amounts. Next, either fOSMR-ECD or mAb#1 is injected. If receptor and mAb bind at different epitopes then there should be no difference in response. When fOSMR-ECD was capture on the surface, mAb#1 binding signal decreased indicating some degree of competition. Similarly, the captured mAb#1 decreased the binding signal of fOSMR-ECD by approximately 50% (**Figure 4a**), indicating at least partially overlapping binding footprints.

Next, using our YSD set-up, fIL-31 yeast cells were labeled with either non-biotinylated fOSMR-ECD or mAb#1 at 12x their experimentally determined K_D values, washed, and subsequently labeled with either biotinylated fOSMR-ECD, fIL31RA-1FNIII or mAb #1 at 1xK_D. At such labeling concentrations and considering 1:1 binding kinetics, it was expected that the non-biotinylated proteins inhibit the binding signal of the biotinylated proteins by approximately 93% if proteins occupy overlapping binding sites and 0% if binding sites are completely non-

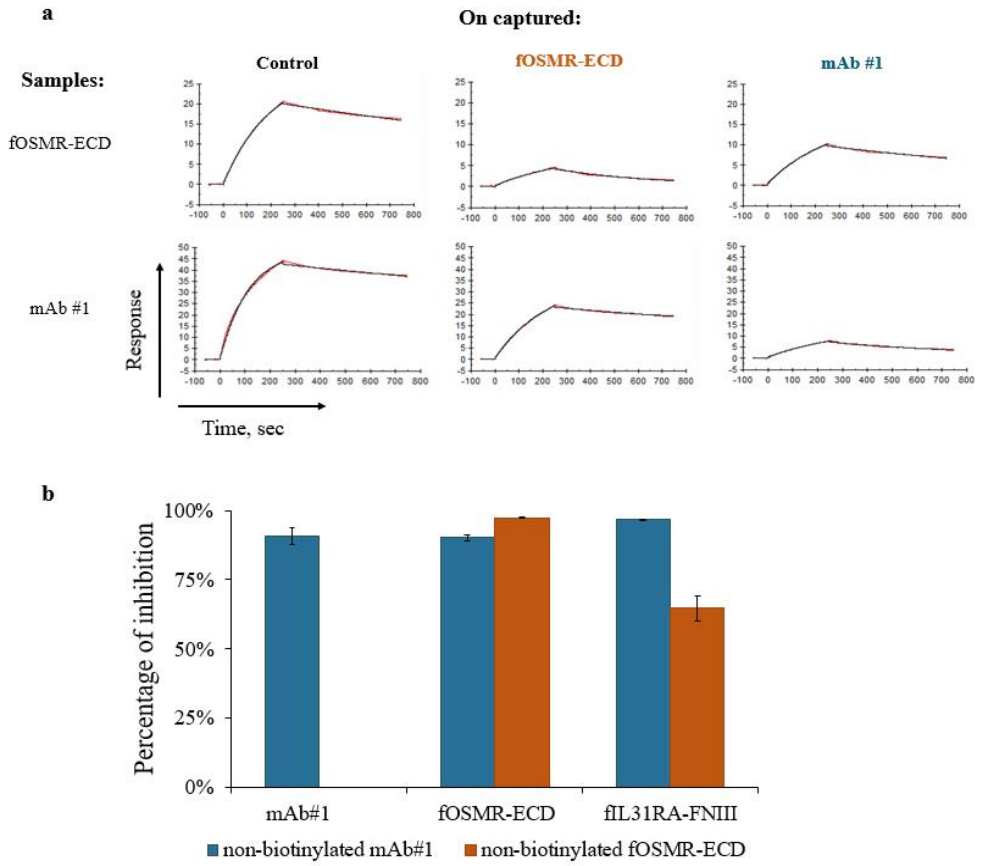


Figure 4. Competition binding assays demonstrated that fOSMR-ECD, fIL-31RA, and mAb#1 share an overlapping epitope using SPR and YSD. (**a.**) SPR sensograms showing a decrease in the binding signal when fOMSR-ECD and mAb#1 were captured on the surface. Done by Zoetis. (**b.**) Percentage of inhibition for fIL-31 yeast cells labeled with either non-biotinylated mAb#1 or non-biotinylated fOSMR-ECD at twelve times their respective Kd values and subsequently labeled with mAb#1, fOSMR-ECD, and fIL31RA-1FNIII at Kd values. (error bars, standard error of the mean, n≥3)

overlapping. The non-biotinylated mAb #1 decreases binding signal of itself, fOSMR-ECD, and fIL31RA-1FNIII by at least 90% (**Figure 4b)**. These results demonstrated that mAb#1 binds an

overlapping epitope or otherwise sterically prevents binding to the two receptors. Non-biotinylated fOSMR-ECD decreases binding to fOSMR-ECD and fIL31RA-1FNIII by 97% and 65%, respectively (**Figure 4b**). The collected data from both measurements suggests that fOSMR-ECD and fIL31RA-1FNIII share partially overlapping binding epitopes on fIL-31 and that mAb#1 blocks both receptors from binding.

2.4.3. fIL-31 can be specifically inhibited by fOSMR-CBD in cell-based assays

To determine whether recombinant fIL-31 could function in cell-based signaling, we incubated varying amounts of either *E. coli*-derived or CHO-derived fIL-31 with the feline macrophage cell line FCWF-4 and determined pSTAT signaling using AlphaLISA pSTAT3 kit. Both recombinant IL-31s activated pSTAT signaling with an EC_{50} < 0.1 μg/mL **(Figure 5a)**, demonstrating fIL-31 functionality. To determine whether fOSMR-CBD can inhibit IL-31, we first pre-incubated receptor with 10 μg/mL fIL-31 for 1 hr at 37ºC prior to stimulation. Increasing amounts of fOSMR-CBD could directly block fIL-31-mediated pSTAT signaling **(Figure 5b)**, indicating that blocking the IL-31 OSMR-binding epitope is sufficient to block signaling.

2.4.4. Structural homology model for fIL-31

Mapping the fIL-31 binding sites and the conformational epitope requires a reasonably accurate homology model. Given that there is no IL-31 crystal structure available, we generated our initial models using I-TASSER[52,53]. While all models contained the predicted four helix up-up-down-down topology consistent with a previous structural model from LeSaux et al.[20], none were likely to be completely accurate. All models contained several buried hydrophilic loop residues and surface exposed hydrophobic residues, and only some models properly paired the disulfide bond between Cys49-Cys132 (data not shown). Based on the initial set of models, we re-ran ITASSER using additional constraints based on putative hydrophobic contacts and requiring a

distance restraint of 2Å for Cys49-Cys123. Then, we used Rosetta to identify the lowest-scoring variant and refine all atoms of the structure. The resulting homology model of the mature protein has a mostly hydrophobic core formed from the four antiparallel helices (**Supp Figure D2)**. Molprobity analysis[54] shows a structure with minimal clashes, over 99% favored rotamers, and only 2 Ramachandran outliers (Glu41, Ser42) on a loop covering the interface between the A and D helix. While there will certainly be deviations between this model and the true structure, we judged this model sufficient to evaluate epitope mapping experiments.

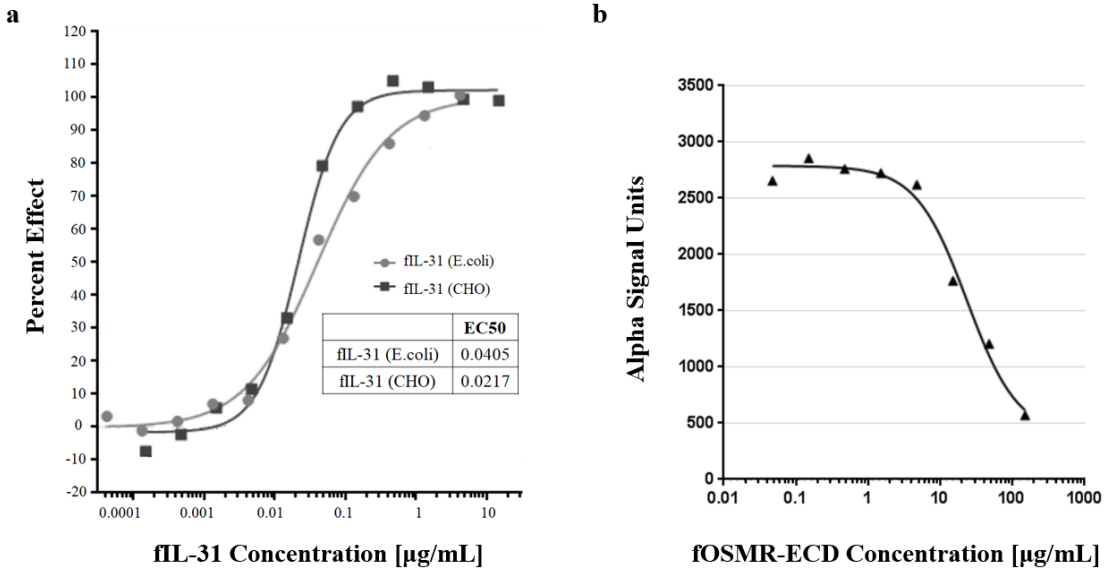


Figure 5. Effect of fIL-31 to induce STAT3 phosphorylation in macrophage cell line FCWF-4 (a.) and fOSMR-ECD inhibition of pSTAT signaling induced by fIL-31 (b.). Done by Zoetis.

2.4.5. YSD and saturation mutagenesis reveals partially overlapping binding sites for fOSMR and fIL-31RA

We determined comformational epitopes on fIL-31 for its binding partners using our previously published method combining yeast surface display, nicking mutagenesis, and deep sequencing[12,24]. The general concept behind our method is that mutations that result in loss of binding will predominantly map to the epitope. To that end, we created two comprehensive single-

site saturation mutagenesis (SSM) libraries for fIL-31 using nicking mutagenesis[23], transformed these libraries into *S. cerevisiae* EBY100, and deep sequenced the population. We observed an average coverage of 79.5% for every possible single missense and nonsense nonsynonymous substitution (2,161 out of 2,720 mutations; **Supplementary Table E3**). The two SSM fIL-31 libraries were expressed on the surface of yeast and each labeled with biotinylated fOSMR-ECD, fIL31RA-1FNIII or mAb #1 at half of the experimentally determined dissociation constant (**Supplementary Table E1**). Libraries were sorted by FACS into two distinct populations: one population corresponding to approximately the top 7% by fluorescence for the channel corresponding to the biotinylated protein (bound population), and a reference population of yeast cells that passed through the cell sorter. Plasmid DNA was extracted and deep sequenced. For each variant we calculated the relative fluorescence values based on the change in frequency between the bound and reference populations[12]. We also determined the per-position Shannon entropy (SE), a measure of sequence conservation, in order to determine the epitope. Summary statistics and complete per-position fIL-31 heatmaps are given in **Table E3** and **Figures D3-D5**, respectively. A subset of the heatmaps for both receptors and mAb#1 are shown in **Figure 6-8**.

Positions with less than 25% accessible surface area (ASA) are removed from analysis, as mutations at these positions often result in misfolded protein. Applied to the structural model of fIL-31, this analysis removed 60/136 positions; we define these positions as structurally conserved. The SE values for these positions are significantly lower than for surface exposed positions (2.17 vs. 1.78 for fOSMR dataset, 2.24 vs, 1.93 for fIL31RA-1FNIII dataset, and 2.30 vs. 2.00 for mAb #1, p-values for all $< 10^{-5}$). Positions with insufficient data for more than ten mutations were also excluded from the analysis, leaving 68 fIL-31 positions for further analysis. Surface exposed residues with lower then midpoint SE values were deemed epitope positions. Of the remaining

non-conserved positions, ones with the highest 10% SE were classified as completely non-conserved.

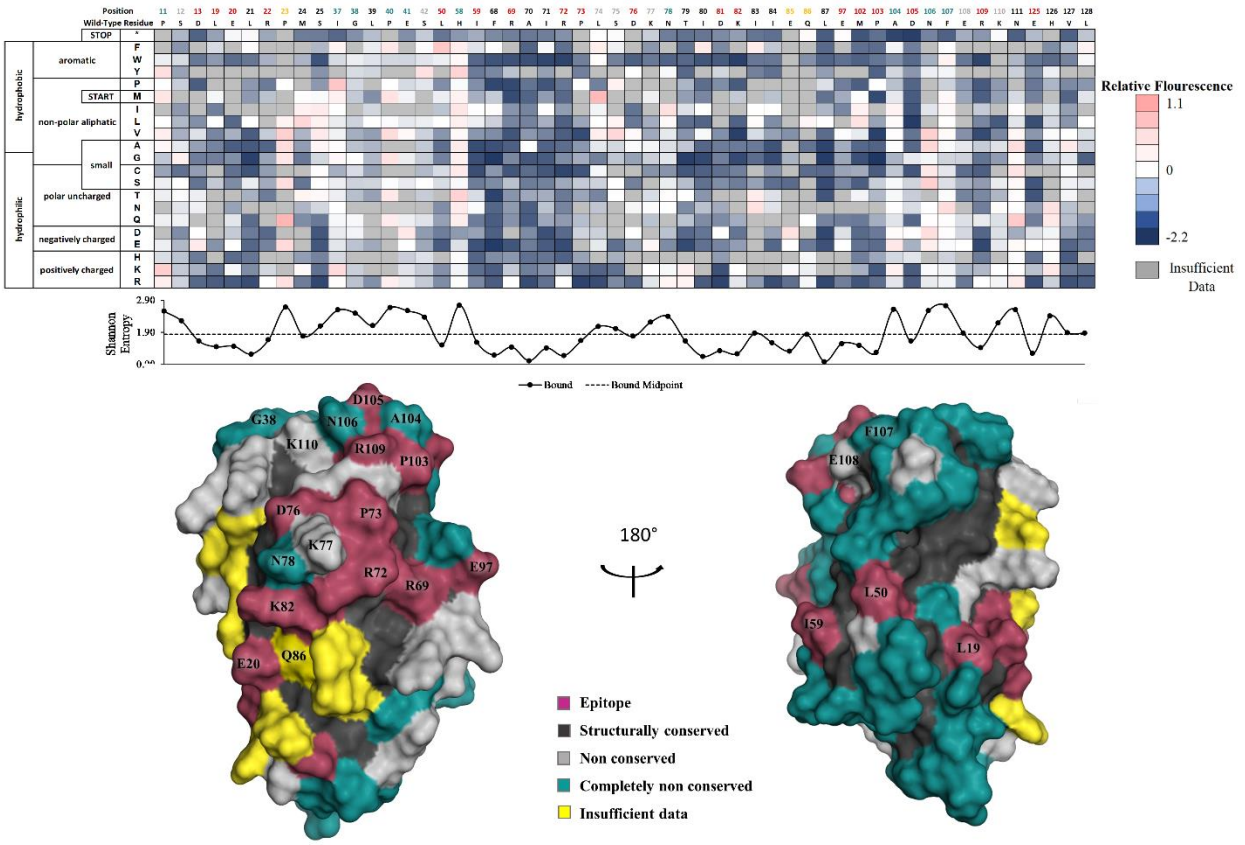


Figure 6. Determination of the fIL31RA-1FNIII binding site using deep sequencing. Shannon entropy with its respective cut-off (dashed lines) is plotted below the heatmap as well as the structural homology model with the determined binding footprint.

Figure 6 shows a restricted per-position heatmap and IL-31 structure for the fIL31RA-1FNIII interaction. 16 IL-31 residues were identified as belonging to the fIL-31RA epitope. These residues form a semi-contiguous patch surrounding the previously described site II from Le Saux et al for hIL31[20]. In particular, E20 on helix A and K82 on helix C were 2/3 previously identified epitope positions on Site II from alanine scanning experiments by Le Saux et al. The third identified residue, Q86, does show reduced binding upon alanine mutation, although there is insufficient data for other mutations at that positions to make a definitive epitope determination.

Unexpectedly, an adjacent contiguous patch on helix B (R69, R72, P73), BC loop (D76), helix C (D81) and CD loop (E97) also appears part of the fIL-31RA epitope, as most mutations at these positions were strongly depleted in the bound library. Other positions identified as belonging to the epitope (L19, L50, and I59) are discontinuous with the rest of the epitope positions and may represent structural conserved positions that reflect inadequacies with the structural model.

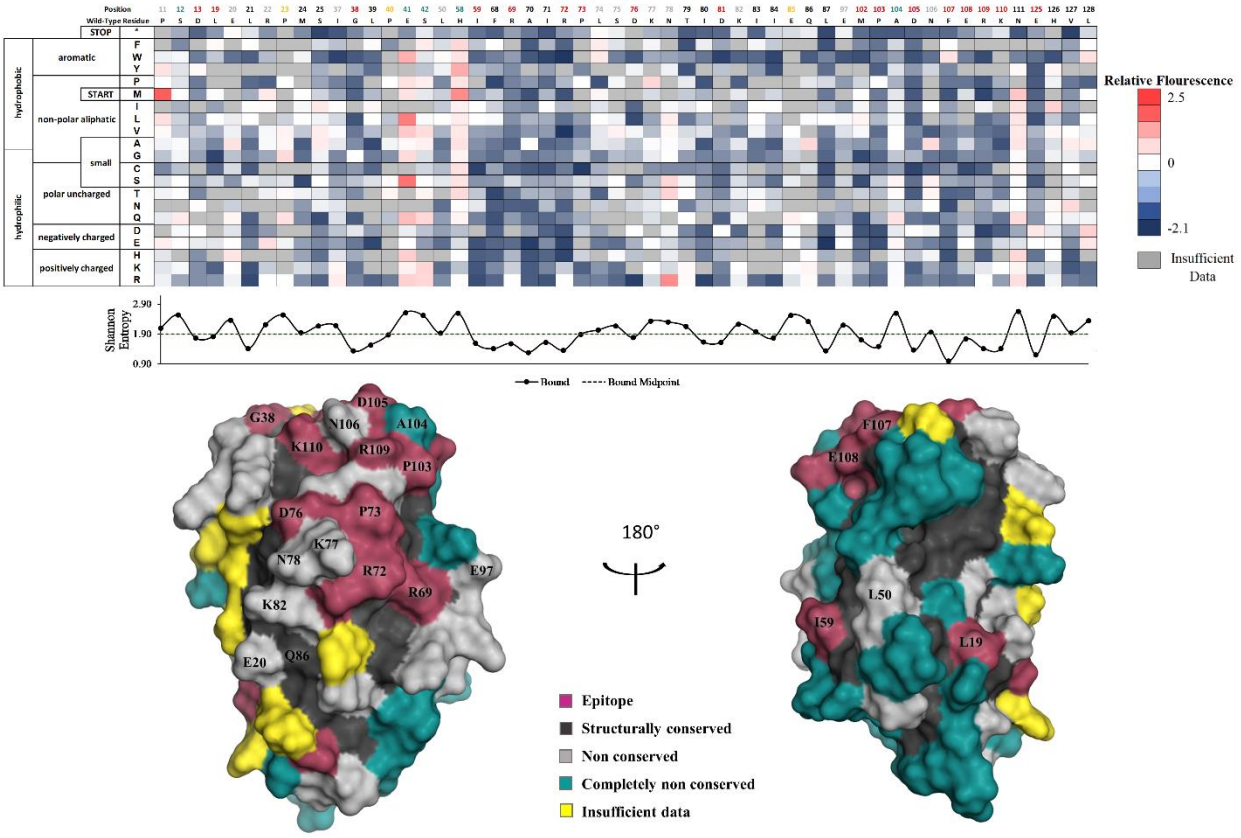


Figure 7. Determination of the fOSMR-ECD binding site using deep sequencing. Shannon entropy with its respective cut-off (dashed lines) is plotted below the heatmap as well as the structural homology model with the determined binding footprint.

Figure 7 shows the fOSMR-fIL31 heatmap and structural model. Overall, 17 positions mapped to the binding site while 28 were completely non-conserved. The perceived epitope covers the expected patch on site III, including G38 and K110 previously identified by Le Saux et al.[20] for the hIL31-hOSMR interaction. This epitope is characterized by strong binding at the

"PADNFERK" motif (P103-K110) at the beginning of the D helix. However, we also observed low SE and thus strong conservation for the same contiguous patch (R69, R72, P73, D76, D81, E97) as for the fIL-31RA interaction. We deemed this region the shared site. Although the sequence entropy of L50 is slightly above the cutoff, the discontinuous L19, L50, and I59 residues observed in the fIL-31RA binding maps are still conserved in this structure and most likely represent structurally conserved residues.

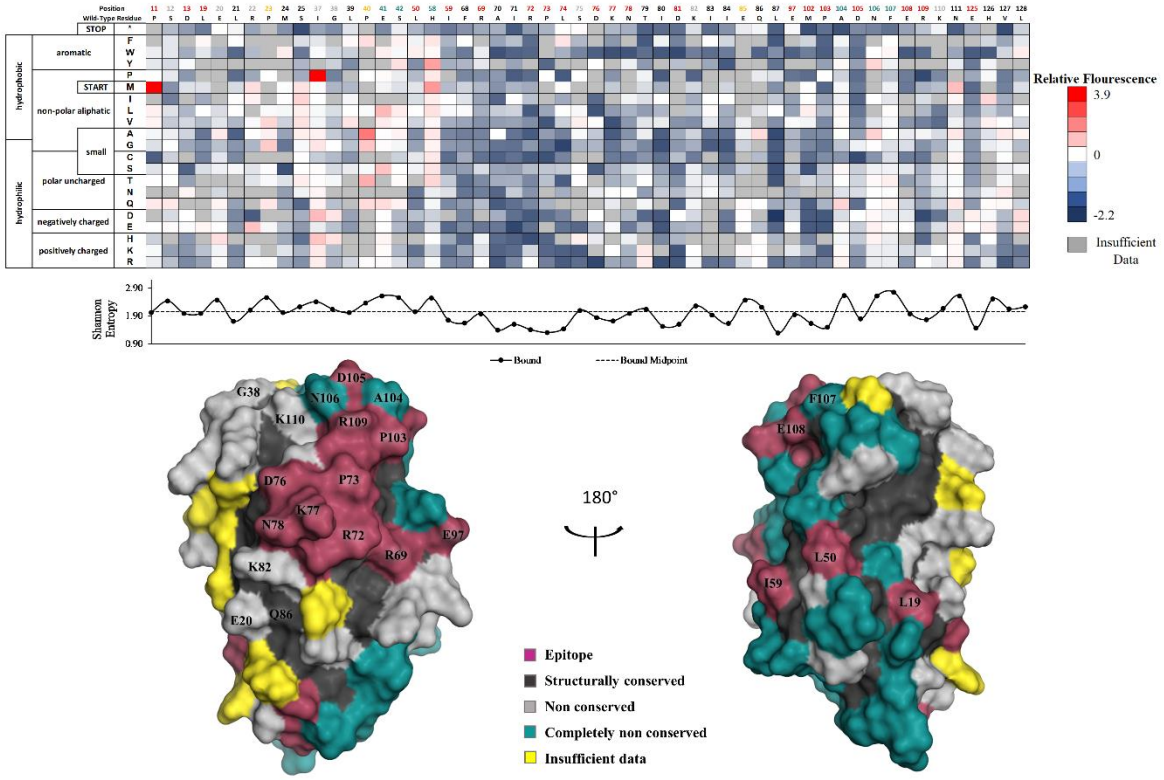


Figure 8. Determination of the mAb#1 conformational epitope using deep sequencing. Shannon entropy with its respective cut-off (dashed lines) is plotted below the heatmap as well as the structural homology model with the determined binding footprint.

2.4.6. mAb#1 conformational epitope reveals nature of inhibition

In our final approach, we determined the conformational epitope for the anti-fIL-31 mAb #1 using the same procedure. 20 positions mapped to the epitope (**Fig 8, Fig D5**). mAb#1 binds to all the positions within the shared site between fOSMR and fIL-31RA as well as adjacent positions K77 and N78. However, high SE was observed at both canonical site II (E20, K82) and site III

(G38, K110) positions, suggesting that mAb#1 does not directly inhibit either canonical IL6-like receptor binding sites. Based on these results, mAb#1 potentially inhibits the signaling of both receptors through the shared site.

2.5. Discussion and conclusion

In this study we have explored the biochemistry for intermolecular interactions between fIL-31, its receptors, and an anti-fIL-31 mAb. Additionally, we used our established deep mutational scanning pipeline with an improved predicted fIL-31 atomic structure to map the binding sites of fIL-31 receptors and the conformational epitope of a potential mAb on fIL-31. We found that, as expected, Site II residues contributed to fIL-31RA recognition while Site III residues were important in fOSMR binding. However, there were two relative surprises resulting from the current work.

First, in contrary to previous results with human orthologs[20,44], our work shows that fOSMR-ECD can directly recognize fIL-31 without the presence of fIL-31RA. Importantly, we used three different independent biophysical methods for assessing binding. We also show that fIL-31 used in binding experiments is functional in cell-based assays and can be specifically inhibited by fOSMR-ECD. Taken together, the biochemical evidence for this interaction is unambiguous. However, we note that our experiments were all performed with feline IL-31 and receptors; thus, it remains to be seen whether there are species-specific differences between IL-31 signaling pathways. In particular, given that the lack if interaction between human OSMR and IL-31 was determined by immunoprecipitation, it would be of interest to perform more stringent biochemical validation on the human orthologs to assess whether the mechanism of interaction is conserved across higher animals. Also importantly, we did not establish in the present work exactly how signaling occurs – are both OSMR and IL-31RA required in felines, or is binding to one

receptor sufficient?

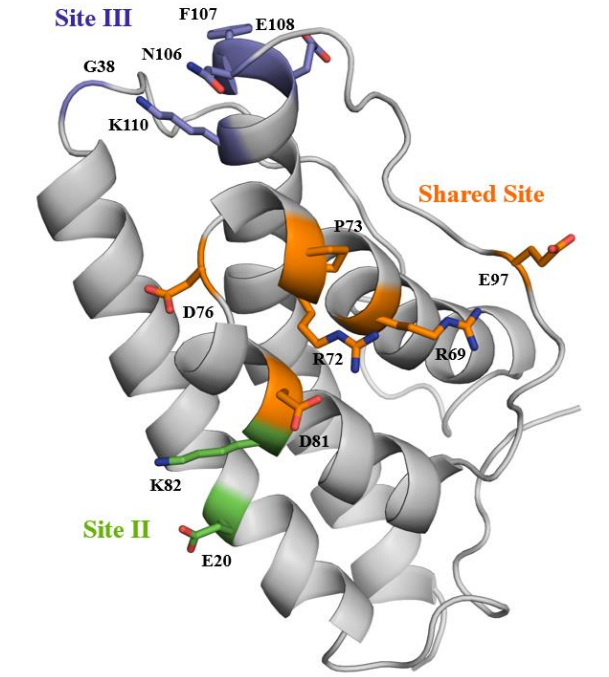


Figure 9. Specific binding sites revealed for fIL31RA-1FNIII and fOSMR-ECD. The shared site (orange) is composed of B-helix, BC loop, C-helix, and CD loop. The binding site for fIL31RA-1FNIII (green), site II, includes A helix and C helix. The binding site for fOSMR-ECD (purple), site III, comprises the N-terminal of AB loop and D-helix.

Second, we found that fOSMR and fIL31-RA could compete for binding on fIL-31, and fine epitope mapping using yeast display pipeline revealed a contiguous patch we deemed the "shared site" between both receptors (**Figure 9**). There are two caveats with an unambiguous determination of this shared site as an epitope for OSMR and IL-31RA recognition. First, we relied on a homology model of IL-31. Thus, it is possible that these shared site positions may not be surface exposed in the monomer. However, we think this unlikely given the contiguous patch revealed by the epitope mapping and the relative hydrophilicity of the involved side-chains. Second, although fIL-31 was mainly produced as a monomeric protein, size exclusion chromatography shows a small percentage of recombinant fIL-31 is dimeric. Such results lead us to ask whether this shared site is exposed on the surface or if it is buried in the interface between

each monomer. While unlikely, this is a real possibility and points to the limitations of fine epitope mapping without a high resolution experimentally determined structure.

For over the past few years, IL-31 and its signaling pathway have been identified as one of the central causes of severe inflammatory skin disorders. Fine epitope information presented here may lead to the development of antagonist mAbs that inhibit the downstream signaling pathway.

CHAPTER 3

3. Pro region engineering of nerve growth factor by deep mutational scanning enables a yeast platform for conformational epitope mapping of anti-NGF monoclonal antibodies

Portions of this chapter were adapted from "Pro region engineering of nerve growth factor by deep mutational scanning enables a yeast platform for conformational epitope mapping of anti-NGF monoclonal antibodies" in *Biotechnology and Bioengineering* (**2018**) 115:1925-1937

3.1. Abstract

Nerve growth factor (NGF) plays a central role in multiple chronic pain conditions. As such, anti-NGF mAbs that function by antagonizing NGF downstream signaling are leading drug candidates for non-opioid pain relief. To evaluate anti-canine NGF (cNGF) mAbs we sought a yeast surface display platform of cNGF. Both mature cNGF and pro-cNGF displayed on the yeast surface but bound conformationally sensitive mAbs at most 2.5-fold in mean fluorescence intensity above background, suggesting that cNGF was mostly misfolded. To improve the amount of folded, displayed cNGF, we used comprehensive mutagenesis, FACS, and deep sequencing to identify point mutants in the pro-region of canine NGF that properly enhance the folded protein displayed on the yeast surface. Out of 1,737 tested single point mutants in the pro region, 49 increased the amount of NGF recognized by conformationally sensitive mAbs. These gain-of-function mutations cluster around residues A-61 – P-26. Gain-of-function mutants were additive, and a construct containing three mutations increased amount of folded cNGF to 23- fold above background. Using this new cNGF construct, fine conformational epitopes for tanezumab and three anti-cNGF mAbs were evaluated. The epitope revealed by the yeast experiments largely overlapped with the tanezumab epitope previously determined by X-ray crystallography. The other mAbs showed site-specific differences with tanezumab. As the number of binding epitopes of functionally neutralizing anti-NGF mAbs on NGF are limited, subtle differences in the individual interacting residues on NGF that bind each mAb contribute to the understanding of each antibody and variations in its neutralizing activity. These results demonstrate the potential of deep sequencing-guided protein engineering to improve the production of folded surface-displayed protein, and the resulting cNGF construct provides a platform to map conformational epitopes for other anti-neurotrophin mAbs.

3.2. Introduction

Nerve growth factor (NGF) was the first discovered member of the neurotrophin family, which also includes brain-derived neurotrophic factor (BDNF), neurotrophin-3, and neurotrophin-4. This family of proteins regulates the development, function, and survival of neurons in the peripheral and central nervous systems[22,55]. Neurotrophins activate downstream signaling pathways by binding the pan-neurotrophin receptor, p75NTR, and to the family of tropomyosin receptor kinases (TrkA, TrkB, and TrkC) with various affinities. NGF is synthesized as a pre-pro protein. The N-terminal pre sequence is released during translocation to the endoplasmic reticulum while the pro-peptide is often but not always cleaved by proprotein convertases prior to secretion. The NGF mature domain is approximately 120 amino acids and arranged as noncovalent homodimers where each monomer conformation possesses a cysteine knot created by three disulfide bonds[56,57].

Multiple studies have demonstrated the high levels of NGF during peripheral nerve injury, inflammation, and chronic pain conditions[58,59]. As a consequence, researchers have developed anti-NGF monoclonal antibodies (mAbs) as potential medicines to modulate chronic pain and many other conditions. These mAbs function by interfering with binding to p75NTR and/or TrkA. Humanized Tanezumab[60] is furthest along in a phase III clinical trial. Comprehensive reviews of clinical studies with tanezumab and others mAbs are found elsewhere[61,62]. As these signaling pathways are highly conserved in higher mammals, here we have investigated a panel of mAbs against canine NGF (cNGF). Canine NGF was used for analysis as this is of interest to Zoetis Animal Health and varies from human NGF by only three amino acids.

An important step in evaluating the neutralizing capacity of antibodies is to determine the epitopes that they target. To that end, yeast surface display[17] is a validated platform to determine

fine conformational epitopes for complicated proteins[11,12,19]. Typically, a set of mutants of the target protein is displayed on the surface of yeast, assessed for antibody binding, with loss of binding mutants mapping to the epitope. A major limitation of the method is the requirement that the displayed target protein be in a conformation recognizable by the antibody. This is an issue with cNGF and consistent with a previous study showing that the related neurotrophin BDNF displayed on the surface of yeast in a mostly inactive conformation[21].

A previous directed evolution study showed that mutations to the pro region could enhance the folding of the related neurotrophin human BDNF in *S. cerevisiae*[63]. Evidence suggests that neurotrophin pro regions act as chaperones to assist folding of a mature neurotrophin as they pass through the secretory pathway[64–67]. The pro peptide is monomeric and highly flexible as shown by the lack of electron density in a solved structure of a proNGF complex[68] and biophysical analysis in vitro [69]. Two domains are sufficient to process and express active mouse NGF[70] (Box 3 and Box 5, shown for canine and human NGF – see **Fig 10a**). Three dibasic sites are proteolytically cleaved during processing of mature NGF through the secretory pathway[65,71].

In the current study we developed a yeast display platform for the production of folded cNGF. We used yeast surface display, saturation mutagenesis, fluorescence activated cell sorting (FACS), and deep sequencing to identify mutations in the pro-region that enhanced display of folded cNGF. Mutational libraries created using this engineering pipeline revealed new insight into the role of the neurotrophin pro region. Combinations of mutations yielded constructs with a 23-fold increase in the signal to noise ratio of display of folded cNGF over background as measured by mean fluorescence intensity (MFI). This engineered pro-cNGF allowed us to generate conformational epitope maps of multiple anti-cNGF antibodies. All anti-cNGF mAbs had an overlapping footprint with tanezumab but each had several site-specific differences. This research

improves our understanding of sequence-function relationships in pro sequences for neurotrophins and highlights the power of deep sequencing to augment classical directed evolution experimental pipelines[72].

3.3. Materials and Methods

3.3.1. Plasmid Constructs

pETconNK_cNGF, pETconNK_Aga2_cNGF, pETconNK_procNGF, and pETconNK Δ1,2-cNGF plasmids were prepared by cloning custom codon-optimized genes (GenScript, Piscataway, NJ) into pETconNK[23] (Addgene plasmid #81169) using standard restriction cloning. Sequences were verified by Sanger sequencing (Genewiz, South Plainfield, NJ), with full sequences listed in **Note C2.**

3.3.2. Preparation of anti- NGF mAbs

Tanezumab, a humanized anti-NGF mAb, was expressed recombinantly in CHO cells based upon the published sequences[60] on hIgG2/kappa constant regions. This antibody was purified using Protein A resin, dialyzed into PBS, and sterile filtered. Three different caninized antibodies (mAb #1, mAb #2, and mAb #3) were supplied by Zoetis Inc. These antibodies were also expressed in CHO cells, affinity purified using Protein A resin, and dialyzed into either PBS or 20 mM Na Acetate, 150 nM NaCl, pH 7.4 buffer. Concentrations were assessed via A280 absorbance using the Edelhoch method and ranged from 1.5 mg/mL to 6.45 mg/mL final concentration. At least 1.5 mg/mL mAbs in PBS were biotinylated at a molar ratio of 1:20 mAb: biotin using the EZ-link NHS-biotin kit following the manufacturer's instruction (Life Technologies). Biotinylated mAbs were then desalted into PBS using Zeba Spin desalting columns (Thermo Fisher) following the manufacturer's instruction and stored at 4^{o}C.

3.3.3. TF-1 Cell Proliferation Assay

Functional potency of antibodies against cNGF was evaluated in a cell proliferation assay utilizing the TF-1 cell line which expresses human TrkA[73]. TF-1 cells (American Type Culture Collection [ATCC], Rockville, MD) were maintained in ATCC modified RPMI 1640 medium (Life Technologies, Carlsbad, CA) supplemented with 10% FBS (Life Technologies, Carlsbad, CA) and 2 ng/ml recombinant human GM-CSF (R&D Systems Inc., Minneapolis, MN) and incubated at 37 °C with 5% CO_2. On the day of experiment, TF-1 cells were washed twice with DPBS (Life Technologies, Carlsbad, CA) before resuspending in proliferation assay medium: ATCC modified RPMI 1640 supplemented with 1% FBS and 10 µg/ml gentamicin. The TF-1 proliferation assay was performed in 96 well microplates (Corning Inc., Corning, NY) by incubating 15,000 cells per well with anti-NGF antibodies at concentrations indicated and 2 ng/ml recombinant cNGF. cNGF was generated at Zoetis in stable CHOK1 cells. After a 72 hours culture period, a CellTiter-GLO luminescent assay kit (Promega, Madison, WI) was employed to evaluate the effects of anti-NGF antibodies on canine β-NGF induced cellular proliferation. After addition of CellTiter-GLO reagent according to manufacturer's instructions, cell lysates were transferred to a white 96 well Optiplate (Perkin Elmer, Waltham, MA) before reading luminescence on a Spectromax M5e microplate reader (Molecular Devices, San Jose, CA). Maximal response in the assay is defined proliferation in the presence of cNGF only (no antibody). Minimal response is defined as measured proliferation without cNGF. Calculated inhibition (NGF neutralization) values for anti-cNGF antibodies are expressed as a percentage of minimal and maximal responses. The resulting percent inhibition/neutralization data was plotted with GraphPad Prism 7 (GraphPad software, San Diego, CA) for IC50 determination using a 4-parameter curve fit.

3.3.4. Surface Plasmon Resonance

Surface Plasmon Resonance was performed on a Biacore T200 (GE Healthcare, Pittsburgh, PA) to measure binding affinities of each antibody to nerve growth factor (NGF). 2.5 µg/ml cNGF in 10 mM Sodium Acetate pH 4 (GE Healthcare, BR-1003-49), 5 µg/ml human NGF (R&D Systems Cat #256-GF/CF) in 10 mM Sodium Acetate pH 4 and 1 µg/ml human proNGF (Alomone labs Cat. # N-280) in 10 mM Sodium Acetate pH 5 (GE Healthcare, BR-1003-51) was immobilized by amine coupling using EDC/NHS for a final density ~250 RU (resonance unit) on CM5 sensor flow cells 2-4, respectively. Flow cell 1 is used as an internal reference to correct running buffer effects. Antibody binding was measured at 15ºC with a contact time of 250 seconds and flow rate of 30 µl/min. The dissociation period was 300 seconds. Regeneration was performed with regeneration buffers (10 mM Glycine pH1.5 and 10 mM NaOH) and flow rate at 20µl/min for 60 seconds each. Running/dilution buffer (1X HBS-EP, GE Healthcare, BR-1006-69, 10X including 100 mM HEPES, 150 mM NaCl, 30 mM EDTA and 0.5% v/v surfactant P20, pH7.4, 1:10 in filtered MQ H2O) was used as negative control at the same assay format. Data was analyzed with Biacore T200 Evaluation software by using the method of double referencing. The resulting curve was fitted with the 1:1 binding model.

3.3.5. Yeast Surface Display Expression and Binding

Cellular fluorescence was measured using a BD Accuri C6 flow cytometer. Yeast cells displaying cNGF variants were detected using anti-cymc-FITC (Miltenyi Biotec, San Diego, CA) and an anti-FLAG tag alexa fluor 647-conjugated antibody (R&D System, Minneapolis, MN). Binding to biotinylated mAbs was detected using streptavidin- R- phycoerythrin conjugate (Thermo Fisher, Waltham, MA). Apparent dissociation constants were determined according to Chao *et al*.[17] by titrating mAb at labeling concentrations from 0.064 nM to 262 nM. Titrations

were performed in triplicate on at least two separate days.

3.3.6. Preparation of Mutagenesis Libraries

Comprehensive single site saturation mutagenesis (SSM) libraries were constructed using nicking mutagenesis exactly as described[23]. All mutagenic oligos were designed using Quik Change primer design Program (Agilent) and were ordered from IDT (Coralville, IA). For pro-cNGF, two separate libraries were prepared: library 1 covered residues Glu-102 – Asn-52 and library 2 covered residues Ile-51 – Arg0. For proΔ1,2-cNGF, the library covered residues Gln-55 – Arg0. For conformational epitope mapping pro.v4-cNGF was split into two libraries, with library 1 covering residues Ser1 – Asp60 and library 2 covering residues Pro61 – Ala120. Library plasmid DNA was transformed into chemically competent *Saccharomyces cerevisiae* EBY100, grown, and stored in yeast storage buffer at -80°C exactly according to published protocols[24] (see **Appendix A**).

3.3.7. Screening of Pro-cNGF and ProΔ1,2-cNGF Libraries

$1x10^7$ cells were grown from freezer stocks in 1 ml of SDCAA for 6 hrs at 30°C and re-inoculated at OD_{600} = 1.0 in 1 ml of SGCAA at 18°C for 16 hrs. $2x10^7$ yeast libraries were labeled with either biotinylated tanezumab or mAb #1 at 5 nM for 30 mins at room temperature in PBS-BSA. After centrifugation and washing, cells were secondarily labeled with 60 µl of anti-cymc-FITC and 50 µl of streptavidin- R- phycoerythrin conjugate in 1.89 ml of PBS-BSA for 10 mins at 4°C. Sorting was done on a BD Influx Cell Sorter at the Michigan State University Flow facility. For each sort 200,000 cells were collected (approx. 100-fold the theoretical diversity at the amino acid level) using a diagonal gate set to collect the top 2-3% of the displaying population (full statistics in **Table E4**). Collected cells from each population were recovered at 30°C for 30 hrs in 10 ml of SDCAA and 100 µl of penicillin-streptomycin, washed, and then stored in 1 ml of yeast

storage buffer at a concentration of $4x10^7$ cells per ml at -80°C.

3.3.8. Determination of Conformational Epitopes

NGF conformational epitopes for all four mAbs were determined using yeast surface display, comprehensive mutagenesis, FACS, and deep sequencing exactly as previously described[24]. **Appendix E** contains the average dissociation constant values (**Table E5**), the percentage collected from library screening (**Table E6**), primers used for deep sequencing (**Table E7**), and statistics results (**Table E8**).

3.3.9. Deep Sequencing Preparation

Libraries were prepared for deep sequencing according to Kowalsky et al. (2015a)[25] using Method B with both PCR reactions set to 14 cycles. The libraries were pooled and sequenced on an Illumina MiSeq using 2 x 250bp pair-end reads at the Michigan State University Genomic Sequencing Core facility or the University of Illinois at Chicago DNA Service facility. Primer sequences used for each library are listed in **Table E7**.

3.3.10. Data Analysis

A modified version of Enrich 0.2 software as described in Kowalsky et al. (2015a)[25] was used to compute enrichment ratios from the raw sequencing files. Custom python scripts available at Github (user: JKlesmith) were used to normalize the enrichment ratios (ER_i) defined as:

$$ER_i = log_2\left(\frac{f_{i,sel}}{f_{i,ref}}\right) \quad 3.1$$

where $f_{i,sel}$ is the frequency of variant *i* in the selected population, and $f_{i,ref}$ is the frequency of variant *i* in the reference population. Libraries statistics results are listed in **Table E4**. For the pro region sorting experiments, we define an enrichment score (ES_i) for each mutant i as the enrichment ratio of the selected mutant minus the wild-type enrichment ratio:

$$ES_i = ER_i - ER_{wt} \quad 3.2$$

For conformational epitope mapping experiments, we define a relative binding term for each mutant as the log transform of the mean fluorescence for variant *i*, $\bar{F}_i$, normalized to the relative mean fluorescence of the wild-type construct, $\bar{F}_{wt}$:

$$\zeta_i = log_2\left(\frac{\bar{F}_i}{\bar{F}_{wt}}\right) \quad 3.3$$

This equation can be written in terms of experimental observables according to:

$$\zeta_i = log_2(e)\sqrt{2}\acute{o}\left[erf^{-1}\left(1 - \Phi 2^{(ER_{wt}+1)}\right) - erf^{-1}\left(1 - \Phi 2^{(ER_i+1)}\right)\right] \quad 3.4$$

where $\acute{o}$ is the log normal fluorescence standard deviation of the clonal population, and Φ is the percentage of cells collected by the sorting gate on the flow cytometer (**Table E6**)[25].

3.3.11. Data Availability

Full datasets including normalized fitness metrics, pre- and post-selection read counts, and raw log base two enrichment scores for each variant can be found in Medina-Cucurella et al.[51]. Raw sequencing reads for this work have been deposited in the SRA (SAMN07693504 – SAMN07693526).

3.4. Results

3.4.1. Initial cNGF constructs are improperly folded on the yeast surface

We sought a yeast surface display platform of cNGF to evaluate binding of candidate anti-cNGF mAbs. Mature cNGF is 97.5% pairwise identical to human NGF (hNGF), with an additional 17 substitutions (out of 103 total residues) on the pro-sequence (**Figure 10a**). We assessed proper folding of cNGF on the yeast surface using two conformationally sensitive mAbs. Soluble

tanezumab – an anti-hNGF mAb – and an anti-cNGF mAb (mAb#1) recognized soluble recombinant cNGF as shown by surface plasmon resonance (**Figure 10b, Table E5**). However, neither tanezumab nor mAb#1 recognized denatured cNGF as demonstrated by lack of signal by Western blotting (data not shown). Recombinant cNGF increased proliferation of TrkA-expressing TF-1 cells[73], and this proliferation could be blocked by the anti-cNGF mAbs (**Figure D6**). Thus, both tanezumab and mAb#1 recognize a conformational epitope on cNGF and all mAbs function as cNGF antagonists. This biochemical data is corroborated with a previously published co-structure of tanezumab with hNGF[60] that reveals binding mainly at the homodimer interface between subunits in a conformation that requires properly folded hNGF. In the remainder of this work we use the term "folded" to describe cNGF that is recognized by these conformationally sensitive mAbs.

We tested folding of four different cNGF yeast display constructs (**Figure 10c**). First, mature cNGF was fused with a N-terminal Aga2p domain, a N-terminal $(G_4S)_3$ linker and a C-terminal c-myc epitope tag (cNGF; pETconNK_cNGF). Second, the mature cNGF was displayed with an N-terminal Aga2p pre sequence, an N-terminal FLAG epitope tag, a C-terminal $(G_4S)_3$ linker, a C-terminal Aga2p, and a terminal c-myc epitope tag (Aga2-cNGF; pETconNK_Aga2_cNGF). Our third construct was identical to Aga2-cNGF but included the full-length pro region between the C-terminus of the Aga2 pre sequence and the N-terminus of the mature cNGF (pro-cNGF; pETconNK_Pro-cNGF). Finally, a classical study defined sections of the pro sequence into five "boxes" on the basis of sequence conservation[70] (**Figure 10a**). Truncation experiments showed that only Box 3 and Box 5 of the pro sequence were necessary and sufficient to produce active mouse NGF. Thus, our fourth construct proΔ1,2-cNGF was identical to pro-cNGF except Box 1 and Box 2 were deleted from the pro sequence. Although all

variants displayed on the yeast surface, yeast cells labeled with saturating amounts of mAbs yielded signals of only 1.1 (cNGF) to 2.2 (pro-cNGF) in the signal:noise ratio (**Fig 10d-e**). This signal:noise ratio was calculated by measuring the ratio of sample MFI over the MFI in the absence of biotinylated antibody ("background") for the subset of cmyc^{+} and fsc/ssc^{+} cells (to ensure measurement of individual yeast cells). Since similar experiments from our research group show 50 to over 100-fold above background for diverse protein-protein interactions[12,74], we conclude that cNGF surface displays in a mostly misfolded form.

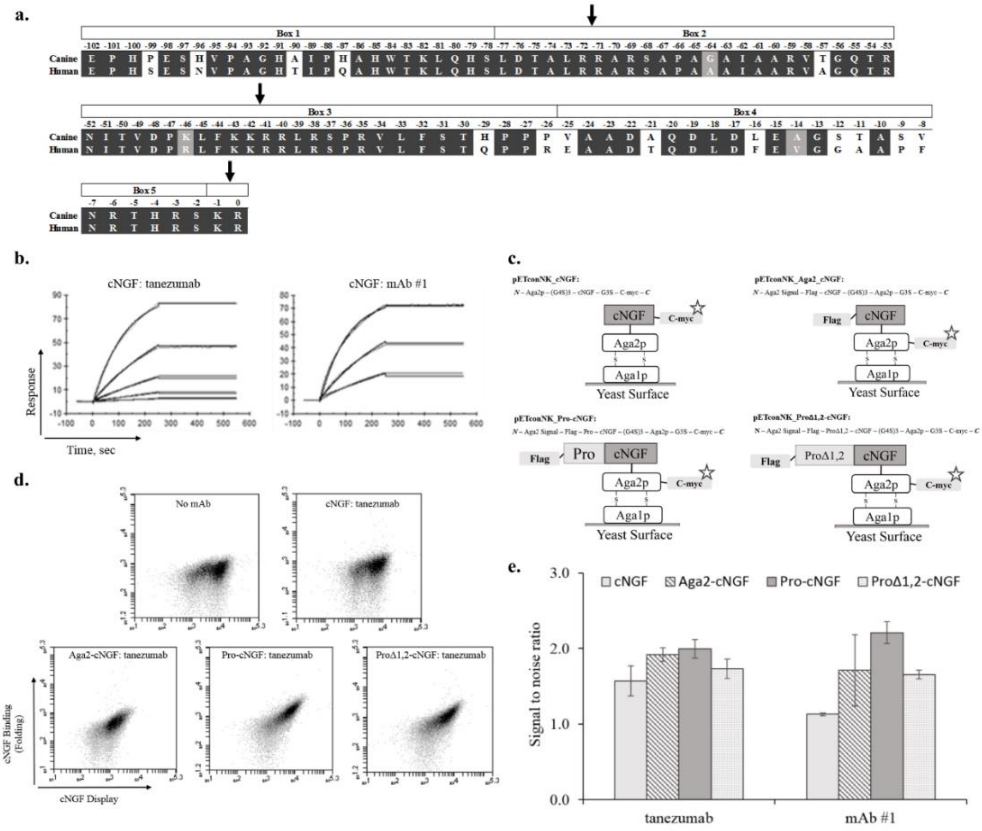


Figure 10. cNGF yeast display constructs are mostly misfolded as probed by conformationally sensitive mAbs. (**a.**) Sequence alignment of the canine and human pro regions of NGF. Domain boundaries and dibasic protease cleavage sites are shown. (**b.**) Surface plasmon resonance sensorgrams of cNGF:conformational mAb binding. cNGF was immobilized on a CM5 surface by amine coupling and either tanezumab or mAb#1 was injected and flowed over surface at various concentrations starting at 100nM and titrating down with 3-fold dilutions, flowed over the chips. Done by Zoetis. (**c.**) Four different cNGF constructs tested in the present work. (**d.-e.**) Flow cytograms (**d.**) and bar charts (**e.**) showing increase in fluorescence in cNGF binding channel probed by tanezumab and mAb #1 (error bars, standard error of the mean, n≥3). The signal:noise ratios were obtained by calculating the ratio of the MFI of the sample to the MFI in the absence of biotinylated mAb.

3.4.2. Comprehensive analysis of pro mutations that improve cNGF folding

A previous directed evolution study showed that mutations within the pro sequence could enhance proper folding of the mature neurotrophin BDNF on the yeast surface[63]. Based on this precedent, we sought mutants that improve the expression of folded cNGF. A flowchart of the experimental pipeline utilizing comprehensive mutagenesis, FACS, and deep sequencing is shown in **Figure 11a**. Comprehensive single-site saturation mutagenesis (SSM) libraries for pro-cNGF and proΔ1,2-cNGF were prepared by nicking mutagenesis[23] and transformed into *S. cerevisiae* EBY100. The SSM libraries covered an average of 84.3% of all possible single missense and nonsense mutations (1737 mutations for pro-cNGF, 943 mutations for proΔ1,2-cNGF; **Table E8**). Libraries were labeled with either biotinylated tanezumab or mAb #1 at 5 nM and sorted by FACS. Tanezumab was chosen as structural information of the mAb-NGF complex is known, while mAb #1 was chosen as a representative anti-cNGF as it had the highest initial signal:noise ratio. We collected the top 3% by cell fluorescence in the mAb channel, along with a reference population of yeast cells that passed through the cell sorter. After each sort, plasmid DNA was isolated, prepared, and deep sequenced. We evaluated each mutant by a relative enrichment score (ES) defined as the enrichment ratio of the mutant in the sorted population minus the enrichment ratio of the wild-type sequence. In this scoring system, a mutant with positive ES improves cNGF folding relative to the wild-type sequence.

After the first sort, bulk populations from all three libraries showed increased fluorescence associated with cNGF folding (**Figure 11b**). We made a number of observations based on these near-comprehensive datasets after the first sort. First, most mutations, including, premature stop codons, centered near an ES of 0 (for pro-cNGF: missense -0.12±0.52, nonsense -0.53±0.36; proΔ1,2-cNGF: missense -0.14±0.45, nonsense -0.03±0.44; mean values at 1 s.d.). This low signal

to noise for loss of function mutations is not surprising after the first sort because the starting constructs have a fluorescence intensity in the binding channel barely above background. Second, for each construct the correlation between each mAb was high: R^2=0.69 for proΔ1,2-cNGF and R^2=0.77 for pro-cNGF (**Fig 11c-d**). Because the observed reproducibility is similar to that seen in replicates using this deep sequencing methodology[75], we cannot differentiate between experimental noise inherent in the deep sequencing pipeline and true biological differences of folding probed by individual mAbs. Third, correlation between pro-cNGF and proΔ1,2-cNGF was

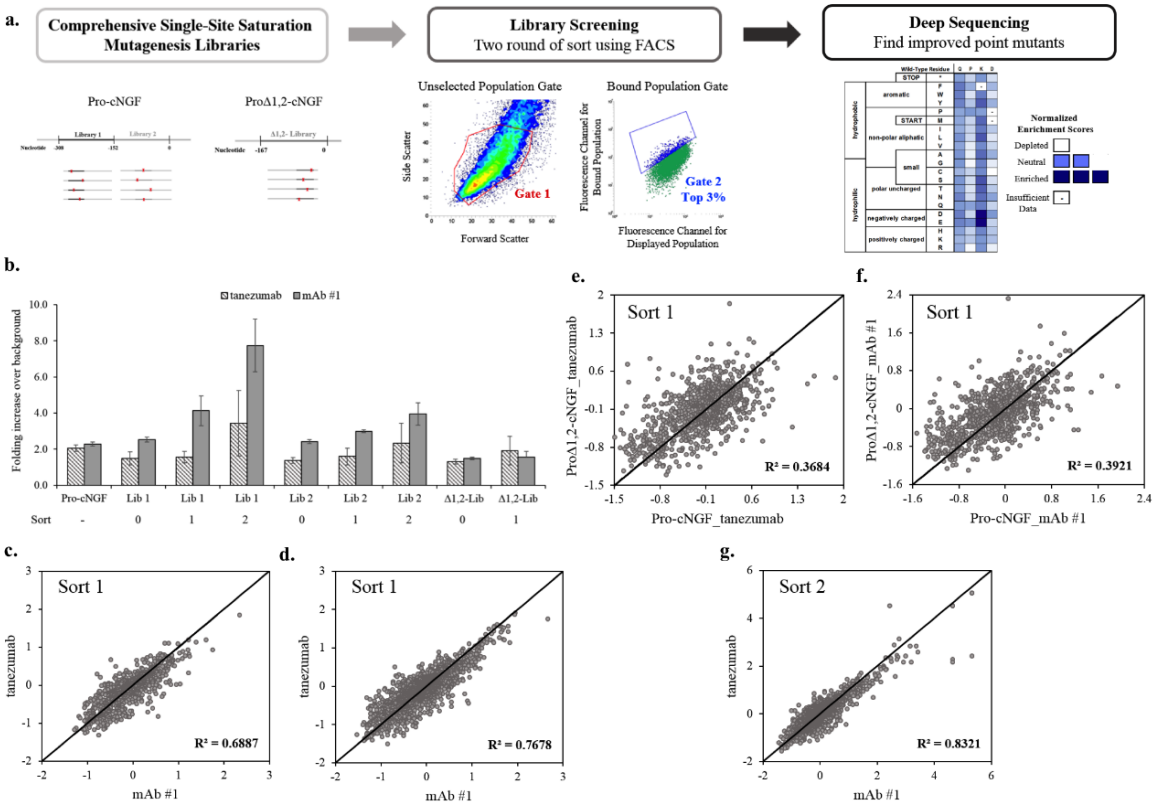


Figure 11. Identifying sequence-function relationships for pro region engineering using deep sequencing. (**a.**) Comprehensive site-saturation mutagenesis libraries were constructed for the pro region and sorted twice by FACS to collect the top 3% of cells using a diagonal gate set for fluorescent channels corresponding to mAb binding and surface display. Collected populations were deep sequenced, compared with the reference population and converted to an enrichment score. (**b.**) Increase in fluorescence channel associated with mAb binding for libraries before and after sorts (error bars, standard error of the mean, n≥3). (**c. – g.**) Correlation in enrichment scores for each mutation for different mAbs and initial constructs.

statistically significant (p-value $3.1x10^{-05}$ tailed paired t-test) but comparatively and unexpectedly lower (**Fig 11e-f**). Thus, mutations that confer differences in cNGF folding are, to a certain extent, context sensitive and suggest that higher order models would be needed to capture the chaperone function of the pro sequence. Fourth, there are many mutations that improve cNGF folding. Using an ES cut-off of three standard deviations above 0 we identified 49 (2.8%) pro-cNGF mutations that improve folding. Finally, because pro-cNGF had more beneficial mutations and a higher bulk population fluorescence than proΔ1,2-cNGF (**Figure 11b**), we moved ahead with pro-cNGF alone for the next round of sorting.

Whereas after the first sort libraries showed only a modest increase in cNGF folding, libraries after sort 2 obtained 3.4 and 7.7 increase in the signal:noise ratio for tanezumab and mAb#1, respectively (**Figure 11b**). The deep sequencing results were very similar between mAbs, with a R^2=0.82 for the entire dataset between conformational antibodies and all 27 tanezumab mutations with an ES above 2 matched in the mAb#1 dataset (**Figure 11g**). A full-length heatmap of the pro-region showing site-specific preferences probed by tanezumab binding is shown in **Figure 12** (full datasets for all constructs and sorts are listed in **Figures D7-D11**). Consistent with Suter et al. (1991), Box 3 is more conserved than the rest of the sequence (mean ES missense mutations -0.23 vs. 0.01; p-value $2.9x10^{-11}$ tailed paired t-test). In fact, for most positions outside of Box 3 a large majority of missense mutations are tolerated, highlighting the inessentiality of specific sequences for the majority of the pro region.

A number of intriguing sequence-function relationships were revealed by inspection of mutations enhancing cNGF folding. Of the 27 mutations with an ES score above 2, all were located either in Box 3 or close to the domain boundary in Box 2. Positive substitutions at Box 2 include substitutions at Ala -61, Thr -57, and Arg-53. Most substitutions at Ala -61 have positive ES, especially aliphatics and aromatics. We note that the nonsense codon has a slightly positive ES at this position: since a premature stop codon is unlikely to result in displayed cNGF, calculations for some slightly positive ES reported here may be within noise of the measurement. All substitutions at Thr -57 had positive ES, including major gains for mutations to Tyr, Met, Leu, Val, and the polar Gln. While Box 3 is more conserved than other pro regions, certain positions showed beneficial mutations. Most notably, most substitutions in the strongly conserved protease susceptible "KKRRLK" sequence had positive ES, including very high ES for charge reversal mutations at Lys-43Asp and Arg-40Asp.

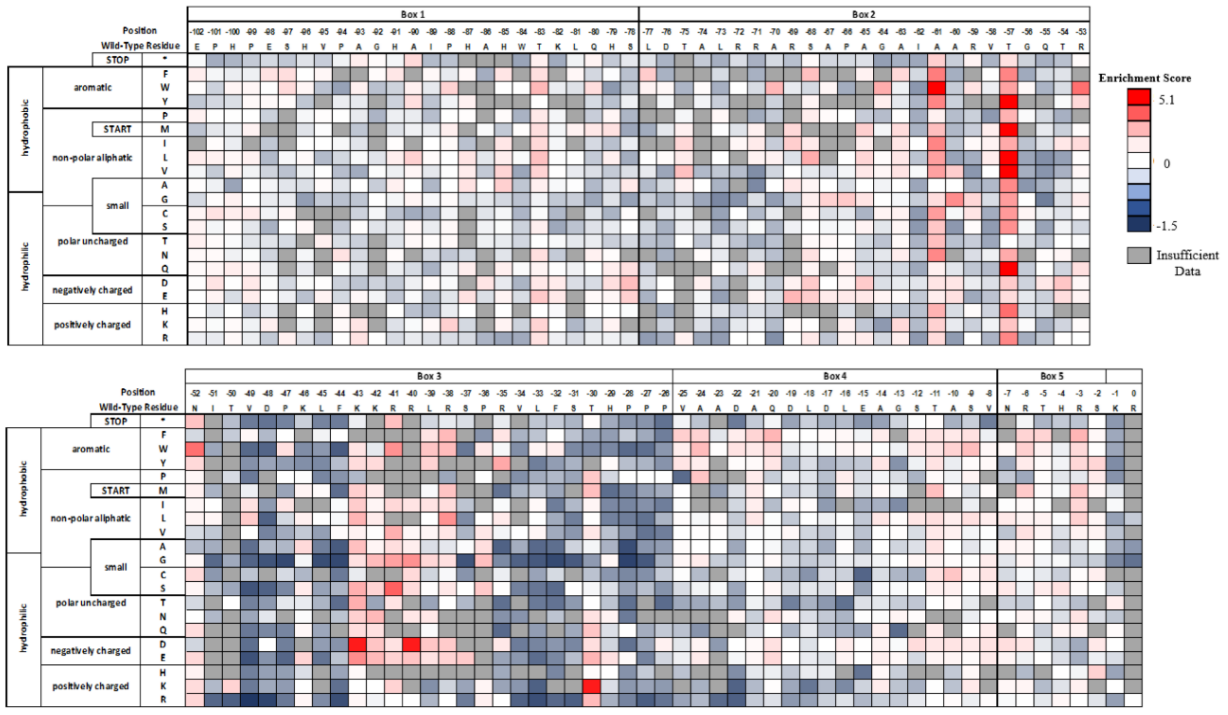


Figure 12. Per-position heatmap of enrichment scores for pro-cNGF mutants after 2 sorts with tanezumab.

3.4.3. Combining single mutants improves the amount of displayed, folded cNGF

To improve display of folded cNGF, we first made three isogenic constructs of the best mutations (ES values higher than 4.6) identified from the deep sequencing experiment (Box 2: A-61W, T-57Q; Box 3: K-43D) and tested their ability to recognize biotinylated tanezumab. In the yeast surface display context, all three chosen mutants showed improved MFI relative to pro-cNGF (**Figure D12**). Next, we reasoned that combining mutations would result in higher amounts of folded cNGF. Pro.v1-cNGF – Pro.v3-cNGF were double mutants, while Pro.v4-cNGF contained all three mutations (**Table 2)**. Each was tested for the expression of the folded protein by labeling with tanezumab, mAb #1, or two additional anti-cNGF mAbs (mAb #2, mAb #3). While all constructs demonstrated at least 12-fold increase in signal:noise ratio, Pro.v3-cNGF and Pro.v4-cNGF showed between a 19 and 23 increase in signal:noise ratio depending on the probe mAb (**Fig 13a-b**). Improved pro constructs gained at most a 1.9-fold increase in surface expression over pro-cNGF (**Figure 13c**), showing that most of the effect of mutations centered on improving the folding of displayed cNGF.

Table 2. Signal:noise ratio for Pro-cNGF constructs defined as the MFI of yeast cells labeled with specified mAb over MFI of unlabeled cells. Errors bars represent the standard error of the mean ($n \geq 3$).

Library Name	Mutations	tanezumab	mAb #1	mAb #2	mAb #3
Pro.v1-cNGF	T-57Q-K-43D	6.5 ± 1.5	7.6 ± 0.7	2.8 ± 1.6	5.3 ± 1.2
Pro.v2-cNGF	A-61W-T-57Q	10.1 ± 2.5	11.7 ± 1.3	4.3 ± 2.6	5.8 ± 1.0
Pro.v3-cNGF	A-61W-K-43D	14.3 ± 3.7	19.0 ± 2.8	6.3 ± 4.4	7.3 + 1.8
Pro.v4-cNGF	A-61W-T-57Q-K-43D	14.3 ± 3.4	20.5 ± 4.2	23.2 ± 2.6	8.7 ± 1.7

3.4.4. Small amount of pro-cNGF displays on the yeast surface

To determine whether pro-cNGF displays on the yeast surface or whether the pro sequence is processed, wild-type cNGF, pro-cNGF and pro.v1-v4-cNGF were labeled with fluorescence conjugated antibodies against the C-terminal c-myc epitope tag and the N-terminal FLAG epitope

tag. The percentage of cleavage was estimated by measuring the FLAG/c-myc ratios of pro-cNGF and pro.v1-v4-cNGF using the fluorescence mean values. As a control, these samples were compared with the ratio of $cNGF^{63}$ labeled exactly the same. In all cases, cells positive for c-myc were slightly positive for FLAG binding (**Figure 13d**), indicating that at least a portion of the pro sequence is proteolytically cleaved before display and suggesting that mature cNGF is displayed. All constructs, Pro.v1-v4-cNGF, showed from 88-96 % of cleavage, meaning that a low percentage of the full-length pro-region is displayed on the yeast surface.

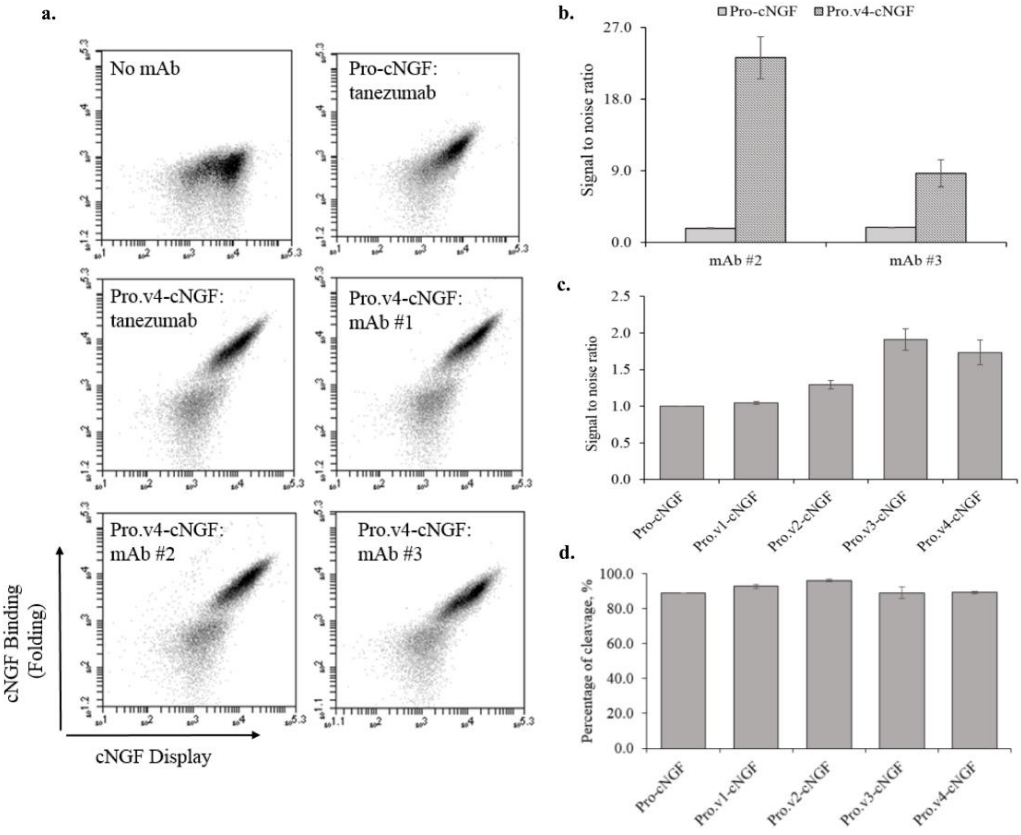


Figure 13. Pro-cNGF variants with multiple point mutations show improved cNGF folding. (**a.**) Flow cytograms for pro.v4-cNGF versus pro-cNGF. (**b.**) Increases in fluorescent channel probed by mAb#2 and mAb #3 binding for pro.v4-cNGF compared with pro-cNGF. (**c.**) Bar charts showing the increase of all pro-cNGF variants in surface expression probed by labeling with a fluorescently conjugated anti-cmyc mAb. (**d.**) Percentage of displayed, folded cNGF on the yeast surface (error bars, standard error of the mean, n≥3).

3.4.5. Conformational epitopes reveal similar profiles, but distinct epitopes, for tanezumab and all three mAbs

Using the pro.v4-cNGF yeast display construct we determined dissociation constants of tanezumab and mAbs #1-3 to cNGF. Yeast cells were incubated with varying amounts of mAb, washed, and labeled with secondary reagents prior to analysis by flow cytometry. Binding dissociation constants ranged from 143±44 pM for mAb #3 to 800±164 (1 s.d., n ≥ 3) for tanezumab (**Figure 14**, **Table E5**). Interestingly, best fits of the Hill coefficient for these mAbs are all significantly below 1 (p-value <0.0106 one tail t-test **Table E5**), indicating potential negative cooperativity between the dimeric displayed cNGF and dimeric mAb. Alternatively, Hill coefficients of less than 1 can arise for binding of non-equivalent binding sites. Since cNGF presumably exists in a range of folded conformations on the cell surface, both alternatives are plausible[76].

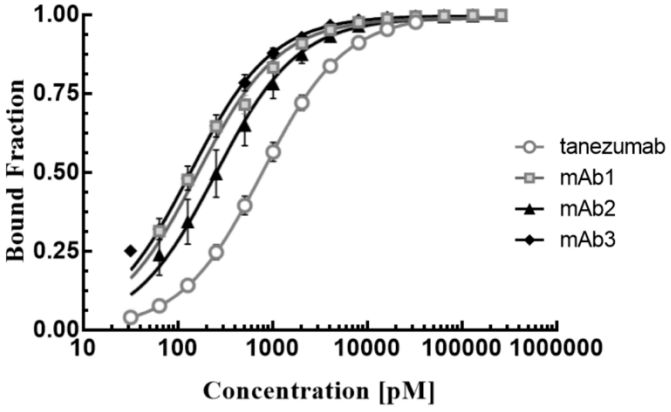


Figure 14. Binding titrations for surface-displayed pro.v4-cNGF for various mAbs. Each mAb was titrated at least three separate times on different days (error bars, 1.s.d. n≥3).

Next, we determined the fine conformational epitopes of tanezumab and the three anti-cNGF mAbs using a previously developed method involving yeast surface display, nicking mutagenesis, and deep sequencing[12,24]. The principle behind this method is that antigenic mutations that disrupt binding will map predominantly to the epitope positions recognized by the antibody. Compared with other epitope mapping strategies that collect the population that no

longer binds antibody[5,15], for our approach we collect the binding population. We then deep sequence the reference and binding populations. For each mutant, a relative binding term can be derived from the change in frequency of the bound population compared with a reference population. Shannon entropy (SE), a measure of sequence conservation, is then calculated. Positions with SE values less than or equal to the midpoint of the SE range are defined as belonging to the epitope[12].

In this experiment, two SSM libraries of cNGF were constructed by nicking mutagenesis, labeled with biotinylated mAb at half of the experimentally determined dissociation constant (**Table E5**), sorted by FACS, and deep sequenced. SSM libraries covered an average of 95.6% of all possible single nonsynonymous mutations in cNGF (1,115/1,200 for library 1 covering cNGF positions 1-60 and 1,178/1,200 for library 2 covering positions 61-120). Full statistics of library coverage is given in **Tables E8**, and per-position cNGF heatmaps for all mAb epitopes are given in **Figure 15a** and **Figures D13-D16**.

In our epitope mapping method, we typically start by removing positions that are structurally conserved from further analysis. We identify structurally conserved positions by collecting the population that surface displays the C-terminal c-myc epitope tag; structurally conserved positions are those that have a Shannon entropy less than or equal to the midpoint for the surface displayed population. However, our results show that mutations at most positions did not result in a change in the displayed protein (**Figure D17**). These results are roughly consistent with the initial results showing misfolded cNGF still displaying on the surface. Thus, knowing the mature NGF conformation, we reasoned that structurally conserved positions are mainly at the core or the homodimer interface and can be identified as positions with less than 25% accessible surface area (ASA) evaluated from solved NGF structures. This analysis excluded 48 out of 120

residues (**Figure D13**). The average Shannon entropy for the tanezumab binding population for these buried positions was significantly lower than for solvent accessible positions (1.75 vs. 2.31, p-value $<10^{-11}$ tailed paired t-test), indicating that conservation of the folded state of cNGF is essential to recognition by tanezumab.

As an initial control we compared our experimentally determined epitope with the previously published X-ray crystal structure of tanezumab Fab bound to hNGF[60]. 11/120 positions mapped to the tanezumab epitope (S19, W21, K32, G33, K34, F49, Y52, K88, A97, W99, and R100) while 38/120 positions were completely non-conserved (**Fig D13 and 15a**). As shown in **Figures 15b-h**, epitope positions form a contiguous patch that largely maps to the binding footprint of tanezumab previously described by La Porte et al. (2014)[60]. Of the eighteen NGF positions within 4Å of tanezumab and with Cα-Cβ vectors pointing towards the antibody, seven were identified as hits in our pipeline, while six were structurally conserved. Of the remainder, only one epitope position (R114) was identified as nonconserved; this Arg makes hydrogen bonding interactions with the main chain of tanezumab. Our deep sequencing approach identifies strong conservation of the cNGF loop centered around K32-K34 (**Figure 15e**) and positions central to the interface (S19, W21) (**Figure 15f**) that presumably make strong van der Waals contacts. The remaining epitope hits were second shell residues (e.g. W99, R100) buttressing these original contacts. However, many epitope positions at the homodimer interface where tanezumab binds to hNGF (e.g. F54, T56, T85, F86, T106, A107, and C108) are invisible to our method since they are structurally conserved (**Figure 15g**). Another shortcoming is that positions recognized by the antibody using main chain contacts (e.g. G10, E11, F12, and H84) are also invisible, as mutations to other amino acids will not disrupt antibody binding. Tanezumab binds to a ridge on NGF with side chains of residues R9-F12, V111, and R114 pointing directly away from the antibody (**Figure**

15h). All of these positions are completely non-conserved (**Figure 15a**).

Finally, we evaluated the binding interaction with three canine anti-cNGF mAbs. Shannon entropy results were very similar between tanezumab and each mAb, with a R^2=0.88, 0.89, and 0.77 for mAb #1, mAb #2, and mAb #3, respectively (**Figure 16**). All mAbs shared 5/11 conserved positions as the tanezumab epitope, but possessed different sequence-binding profiles and therefore unique epitopes (**Fig D14-D16**).

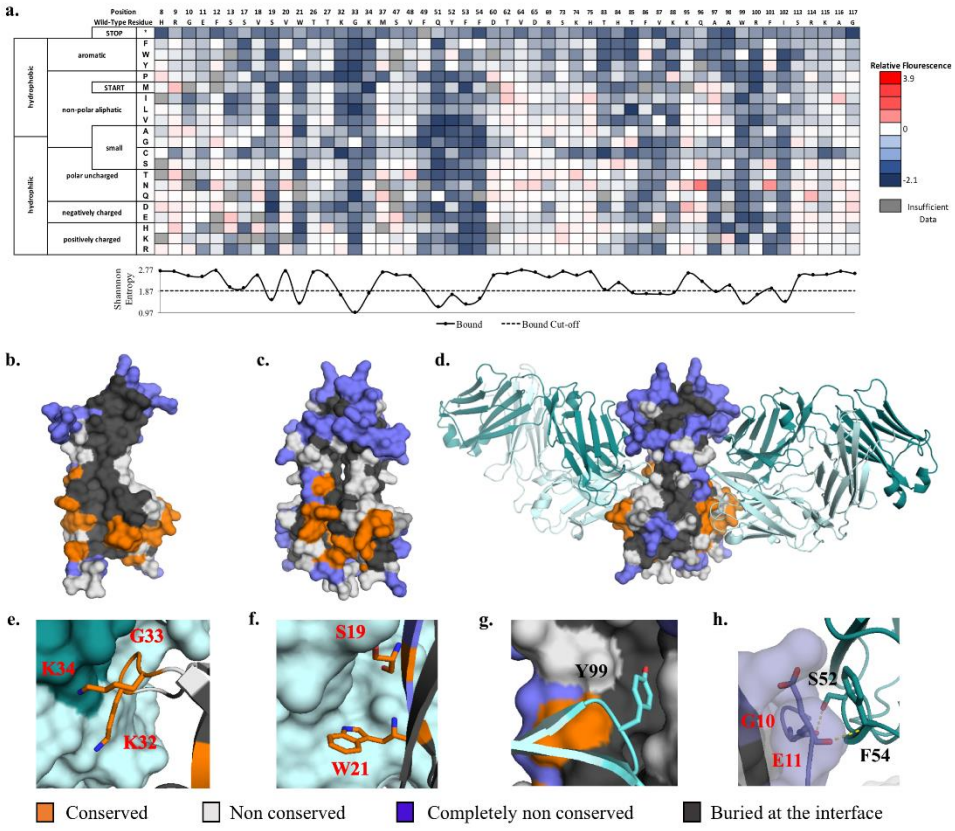


Figure 15. Determination of the cNGF:tanezumab conformational epitope using deep sequencing. (**a.**) A subset of the fitness metric per position heatmap of the top 7% bound population vs. the unselected population for tanezumab. Shannon entropy is plotted beneath with the midpoint cut-off shown as a dashed line. (**b.-c.**) Tanezumab epitope view at the interface core of each monomer (**b.**) and as a homodimer mature structure (**c.**). Positions colored in orange, blue, gray, and white are epitope, completely non-conserved, buried in the core, and non-conserved positions, respectively. (**d.-h.**) Close-up views of specific tanezumab:cNGF sidechain interactions. Molecular representation of the interaction was modeled on the solved tanezumab:hNGF structure (PDB ID: 4EDW).

Interesting, there were two mutations, V6M and M92F that show increased relative

fluorescence in the binding population for all four mAbs (**Fig D13-D16**). Our epitope mapping method cannot discriminate whether such mutations improve the binding affinity to a given mAb or whether the mutation increases the amount of folded cNGF. However, we speculate that the latter is the case as these two residues are surface exposed positions located distal to the tanezumab epitope, and all four mAbs bind in overlapping but atomically different epitopes.

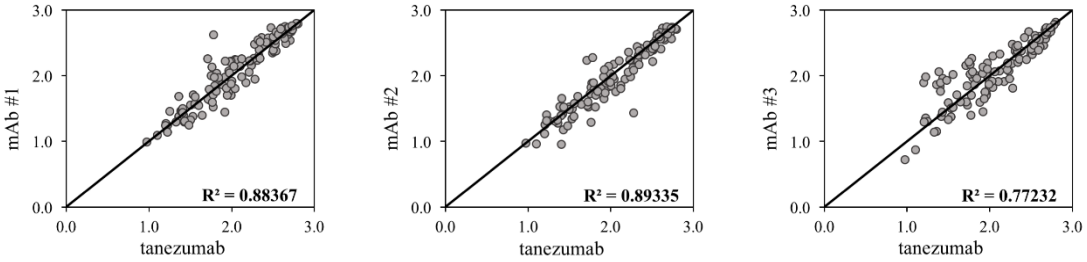


Figure 16. Correlation in Shannon entropy of the binding populations between tanezumab and other mAbs.

3.5. Discussion and Conclusion

In this study we engineered a pro-region to enhance folding of cNGF on the yeast surface. This yeast platform enabled our team to map conformational epitopes for tanezumab and a number of anti-cNGF mAbs. During the course of this research we showcased the power of deep sequencing to augment directed evolution workflows to improve protein properties[72], gained new insight into sequence-function relationships for neurotrophin pro regions, and identified fine conformational epitopes for anti-cNGF mAbs.

3.5.1. Pro-region engineering pipeline

cNGF and pro-cNGF displayed on the yeast surface but in mostly misfolded forms. This adds to a growing body of literature showing that while the quality control machinery in the *S. cerevisiae* endoplasmic reticulum[77] can impact the overall amount of protein displayed on the surface[8,75], for any given grossly misfolded protein some will still pass through quality control checkpoints in the secretory pathway and display on the yeast surface[8,12,74,78].

We used an engineering pipeline involving comprehensive site-saturation mutagenesis, FACS, and deep sequencing to improve surface displayed, folded cNGF. This engineering pipeline enabled us to evaluate the functional effect for over 2,000 individual point mutants to the pro-cNGF and proΔ1,2-cNGF constructs. We then created multi-site constructs by combining the best mutations cherry picked from the deep sequencing datasets. There are several advantages of this workflow compared with traditional directed evolution approaches. First, precise single-site mutagenesis libraries constructed from nicking mutagenesis allow us to assign a functional effect to a single mutation; libraries constructed by error-prone PCR often have multiple mutations per gene, which then need to be deconvoluted. Second, deep sequencing allows us to perform experiments in parallel: we performed duplicate sorts using different conformational mAbs, allowing one to identify mutations beneficial to both mAbs. Third, the small library size allows us to simplify steps for yeast transformation and sample handling, shorten FACS times, and allow us to complete our screening in 2 sorts compared 4-7 sorts for larger libraries ("go small, get them all"). One slight disadvantage of screening single point mutants is the chance of missing epistatic mutations but in the present work identified beneficial pro mutants could be combined additively.

Our best construct identified by deep sequencing greatly improved the folding of displayed cNGF and was sufficient for the conformational epitope mapping performed. However, we suspect that there is room for further optimization of the yeast display of folded cNGF as (i.) the max signal:noise ratio of pro.v4-cNGF is still lower than that seen for other antibody-antigen interactions probed by our lab; (ii.) two mutations on mature cNGF could presumably improve folding of the pro.v4-cNGF construct; and (iii.) the Hill coefficient for all antibodies was less than one, which is consistent with cNGF existing in a range of folded and partially misfolded configurations on the yeast surface. Further optimization could be done by selecting similar

beneficial point mutants identified from the deep sequencing datasets (**Figure 12**). We speculate that the combination of an additional subset of these point mutations would facilitate production of mature cNGF from yeast[63], although this is beyond the scope of the present work.

3.5.2. Pro-region sequence-function relationships

Although the classic work from Suter *et al*. (1991)[70] showed that Box 3 and 5 of the pro region were sufficient to produce active mouse NGF, mutations for the truncated variant proΔ1,2-cNGF did not improve folding as noticeably as for pro-cNGF. Given that many of the strongest beneficial mutations were in Box2, we speculate that Box2 encodes some chaperone activity necessary for complete cNGF folding. However, further studies should be performed to understand domain boundaries and functional differences between constructs.

Conserved and beneficial mutations cluster around Box 3 as was suggested for mouse NGF[70]. Notable lack of conservation in Box 3 centered around a cleavage site for proprotein convertases at Lys-43 – Arg-40. These four positions demonstrated improved profiles for almost all the mutations. Indeed, Lys-43Asp was one of the mutations introduced in our engineered pro-cNGF that in turn, disrupt recognition of this KR dipeptide by endogenous Kex2 protease in *S. cerevisiae*. Interestingly, both pro-cNGF and prov.4-cNGF display a mixture of full-length pro-cNGF and at least partially cleaved cNGF, suggesting that enhanced folding conferred by the Lys-43Asp mutation does not result from differential protease cleavage.

A previous study utilizing hydrogen-deuterium exchange with the pro region and mature NGF determined that Trp-83 – Ala-63 in the pro sequence (Trp-84 – Gly-64 canine numbering) was involved in the intramolecular chaperone-like interactions with mature human NGF[64]. Inconsistent with this experiment, we observed all positions in this region to be non-conserved, with most mutations centered around an ES of 0. Our deep sequencing experiments probe folding

of cNGF but do not directly assess the chaperone function of specific pro sequence variants, so further work is necessary to reconcile this apparent discrepancy.

3.5.3. Mapping epitopes targeted by anti-cNGF mAbs

We were able to determine conformational epitopes for tanezumab and three potential canine mAbs. Our experimentally determined tanezumab epitope largely overlapped with the previously published tanezumab-mNGF structure[60], although many epitope positions (i.) at the dimerization interface, and (ii.) not participating in side chain contacts with antibody were invisible to our deep sequencing method. These shortcomings will be shared with all mutational-based epitope mapping methods. The three anti-cNGF mAbs had different sequence-binding profiles at the epitope. We conclude from these results that our yeast display platform for cNGF is able to map fine conformational epitopes for candidate anti-NGF mAbs.

Although our epitope mapping workflow cannot directly measure if cNGF is displayed as a monomer or dimer on the yeast surface, the results strongly suggests that our best constructs display as mature dimer cNGF. Mutations at the homodimer interface core away from the epitope will disrupt the folded state of cNGF necessary for mAb recognition, and indeed these are depleted in our experiments. These results are consistent with our lab's previous determination of the conformational epitope tumor necrosis factor (TNF)-infliximab[12], where mutations at the TNF homo-trimeric interface disrupted antibody binding. However, we note that the solution-based measurements of cNGF-mAb binding are high fM-low pM, whereas measurements from the yeast surface are mid- to high-pM. While Gai and Wittrup have shown the rough equivalence in affinity between yeast- and solution measurements[79], their dataset included only 1 fM binder and thus we do not necessarily expect equivalence for these high affinity binders.

In conclusion, we used a deep sequencing pipeline to develop a yeast display platform for folded cNGF. This contribution highlights the power of deep sequencing to identify nearly all beneficial point mutants in a protein in a simplified workflow. Since a major limitation of yeast display is proper folding of complicated mammalian proteins, this work complements a recent directed evolution study [63] to show that effective strategies exist to overcome such limitations. Our pipeline could be a promising platform to increase the production of highly active titers of the other members of the neurotrophin family, to determine the specificity and affinity to their respective receptors, and to enable the epitope mapping for therapeutics against neuronal diseases such as the brain disorders caused by BDNF.

CHAPTER 4

4. User-defined single pot mutagenesis using unpurified oligo pools

Portions of this chapter were adapted from "User-defined single pot mutagenesis using unpurified oligo pools" in *PEDS* (under review)

4.1. Abstract

User-defined mutagenic libraries are fundamental for applied protein engineering workflows. Here we show that unpurified oligo pools can be used to prepare libraries from plasmid DNA with near-complete coverage of desired mutations and few off-target mutations. We find that oligo pools yield higher quality libraries when compared to individually synthesized degenerate oligos. We also show that multiple libraries can be multiplexed into a single oligo pool, making preparation of multiple libraries less expensive and more convenient.

4.2. Introduction

Directed mutagenesis is foundational for synthetic biology and protein engineering. Recent methods support the creation of large libraries of user-defined mutations in a single reaction [23,80,81]. Such protocols rely on annealing a short oligonucleotide to a parental template, wherein the oligo encodes a mutation by template mismatch. The complementary strand encoding the desired mutation is synthesized, after which the parental template strand is specifically destroyed. For large libraries the mismatch is encoded using degenerate nucleotides, such that a single oligo can make up to 63 different mutations per codon substitution. However, degenerate oligonucleotides require hand-mixing to avoid overrepresentation of nucleobases[80] and are often unable to encode a desired subset of amino acids. Microarray-synthesized oligo pool technology has recently found use in synthetic biology[82], with multiple vendors offering relatively long oligos at moderate error rates. Clever techniques have emerged to use these pools for gene synthesis[83], but the low femtomolar concentrations of individual oligos usually necessitates amplification and further processing from pools, limiting usability.

We recently described the Nicking Mutagenesis (NM)[23] method to construct user-defined libraries in one pot using routinely prepared plasmid DNA. Because NM uses very low oligo

concentrations, with the template in large excess to the mutational primer, we hypothesized that unpurified single stranded oligonucleotides from microarrays could be used directly in the reaction. To test whether unpurified oligo pools are compatible with NM, we synthesized a single custom oligonucleotide pool (Agilent technologies) comprising oligos encoding all missense and nonsense mutations from positions 1-69 of the bacterial aliphatic amidase AmiE[84] (1449 oligos of length 33-60 nts), all possible single nucleotide polymorphisms covering residues 15-114 for the anti-Influenza human antibody variable heavy gene UCA9[85] (1000 57-nt oligos), and targeted mutagenic oligos for the *Arabidopsis thaliana* abscisic acid receptor PYR1[86] (185 51-nt oligos printed with 6 replicates). Full oligo sequences are given in Medina-Cucurella et al (under review, *PEDS*).

The lyophilized oligo pool was solubilized in 40 μL TE at a total concentration of 200 nM, phosphorylated, and diluted directly so as to contain 1.9 nM AmiE-specific oligos. This dilution was used directly in a standard nicking saturation mutagenesis protocol with an *AmiE*-encoding plasmid as a template (see **Supporting Note C4**). Two replicates were performed. We also performed a control reaction using degenerate 'NNN' oligos that were individually synthesized (IDT) covering the same stretch of the gene. Libraries were sequenced on an Illumina MiSeq in 250 bp paired end mode and processed using PACT[87]. 100% (1380/1380) of the desired mutations were incorporated for each replicate (see **Table E10** for full library statistics of all sequences; processed datasets are shown as heatmaps in **Figure D18**). The frequency of specific mutants had a correlation of 0.88 between replicates, demonstrating the repeatability of the mutagenesis protocol (**Figure 17a**). Importantly, the oligo pool library had a much more even representation of all 20 amino acids compared with the degenerate oligos, with a mean coefficient of variation (CV) of 0.12 compared with 0.59, respectively (**Fig 18a and Fig D19**). However, the cumulative

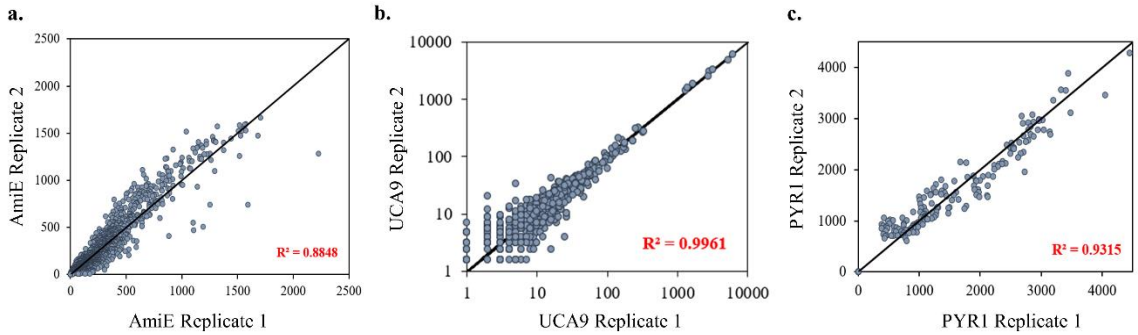


Figure 17.Correlation between the frequency of (a.) AmiE, (b.) PYR, and (c.) UCA9 mutants between replicates.

distributions of libraries - normalized to 200-fold coverage – were broadly similar (**Figure 18b**). To demonstrate that multiple libraries can be prepared from the same oligo pool, we sought to construct a library of all single nucleotide polymorphisms on the majority of the UCA9 by performing NM with an *UCA9*-encoding plasmid as a template with the unpurified oligo pool. Sequencing confirmed an average of 98% (594/612 and 606/612 nonsynonymous mutations from replicate 1 and replicate 2, respectively) of coverage of the desired mutations (**Fig 18c, Fig D20**) with a correlation of 0.99 and similar cumulative distributions between replicates (**Fig 17b**). Notably, mutations were specifically programmed, as there were on average only 5.8% (86/1,488) off-target mutations observed in the read window (**Figure 18c**).

Oligo pools offer user-defined mutations relieved from the constraints of degenerate codon compatibility. We next sought to construct a library of 185 designed mutations at 17 positions in the PYR1[86] receptor by performing NM using the unpurified oligo pool with a *PYR1*-encoding plasmid as a template. All 185 mutations were encoded in the library (**Figure 18d**) and with a correlation of 0.93 between replicates (**Fig 17c**). Specifically programmed mutations were found on average at 91-fold higher frequencies than the other 1556 potential single non-synonymous mutations in the 261-nt Illumina sequencing window (median 1038 counts vs 0; mean 1198 counts

vs. 13 for encoded vs. non-encoded positions) (**Table E10 and Fig D21)**.

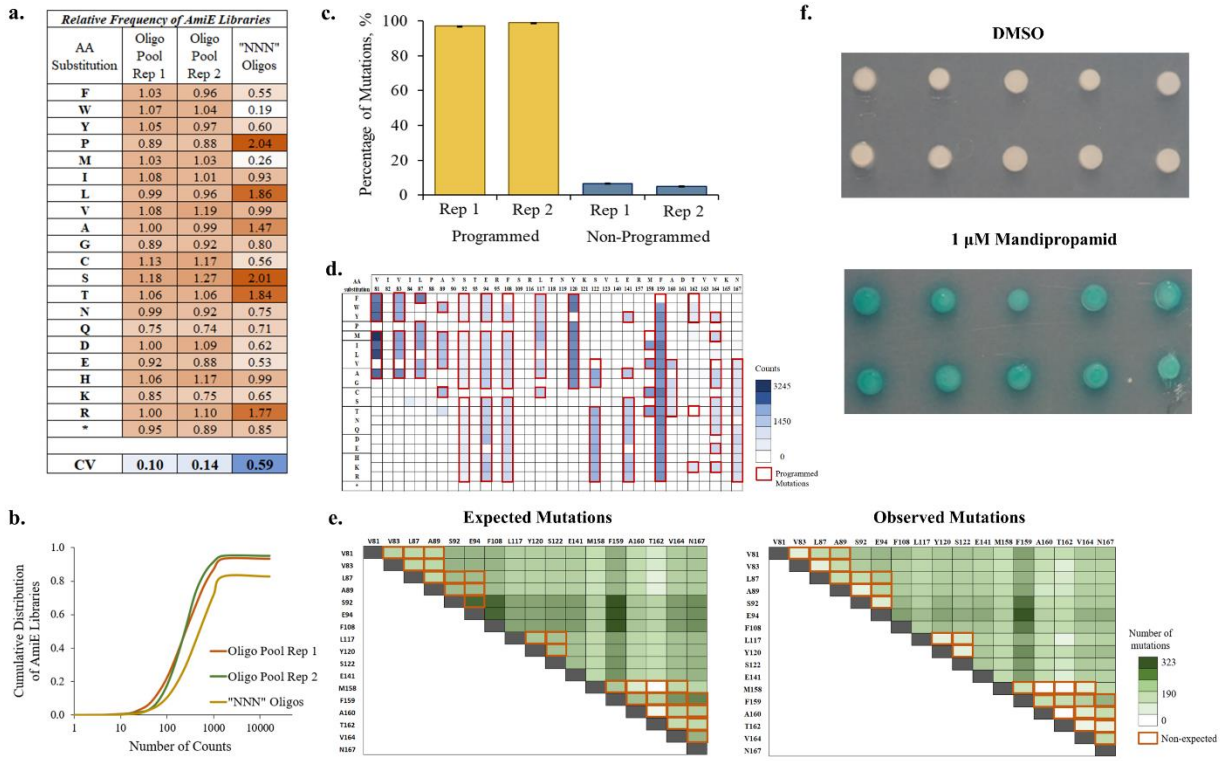


Figure 18. A single unpurified oligo pool combined with nicking mutagenesis shows a near-complete coverage of all programmed mutations for user-defined single and double mutagenic libraries. (**a.**) The relative amino acid substitution frequencies for AmiE libraries prepared using unpurified oligo pools and "NNN" oligos (CV = coefficient of variation; rep = replicate). (**b.**) The cumulative distribution function of AmiE libraries as a function of the number of counts normalized to a 200-fold depth of coverage. (**c.**) Bar charts showing the percentage of programmed versus non-programmed mutations for UCA9 libraries for the two different replicates. (**d.**) A subset of the per position heatmap showing the number of counts per mutations encoded in PYR1 single site libraries. Boxes framed in red represent the programmed mutations. (**e.**) Per position heatmaps comparing the expected versus observed programmed mutations in the PYR1 double site mutagenic library. Boxes framed in orange represent the non-expected mutations due to nucleotide mismatches between oligos and DNA template. (**f.**) X-Gal staining of yeast colonies for PYR1 double mutants that interact with HAB1 in the presence of 1 µM Mandipropamid using an established yeast-two hybrid system. DMSO is the negative control.

It is also possible to sequentially perform NM with the same oligo pool, creating mutagenic libraries with two mutations per gene. We tested this synthesis by performing NM again using the PYR1 single mutagenesis library plasmid DNA as a template. We expected 10,845 out of the 1,462,000 possible double point mutants in the Illumina sequencing read window. Deep

sequencing recovered 13,904 double mutants with 2 or more read counts, of which 8,316 were specifically programmed (**Fig 18e**). Double mutations were depleted at near-adjacent positions (**Fig 18e**), presumably because a second oligonucleotide containing mismatches would either not anneal or overwrite the mutation encoded from the first oligonucleotide. To demonstrate utility, this library was screened against the non-native ligand mandipropamid using a previously established yeast 2 hybrid screen[86]. We uncovered PYR1 mutants specific to and responsive at 1 μM mandipropamid; sequencing of 10 constructs showed all 10 had the same specifically programmed F108A F159M double mutant, which outcompeted single mutants F10A and F159M present in the library (**Fig 18f**).

In summary, we have shown that oligo pool synthesis technology can be integrated with nicking mutagenesis to construct user-defined single and double mutagenesis libraries. We anticipate its incorporation into standard directed evolution experiments and its utility for more thorough evaluation of local protein fitness landscapes.

4.3. Materials and Methods

4.3.1. Strains

The *Escherichia coli* strain used in this study was XL1-Blue (Agilent, Santa Clara, CA*) endA1 supE44 thi-1 hsdR17 recA1 gyrA96 relA1 lac* [F' *proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^{r})]. The *Saccharomyces cerevisiae* strain used in this study was MaV99, *MAT*a SPAL10::URA3.

4.3.2. Plasmid Constructs

The pEDA3_AmiE plasmid was created as described in Wrenbeck et al. 2016. Plasmid pBD_PYR1_BbvCI was constructed from pBD_PYR1[86] by inserting a BbvCI restriction site using standard cloning. pETcon_UCA9 was created by inserting a codon-optimized V_H gene encoding

Met1 – Ser122 (Integrated DNA Technologies, Coralville, IA) into yeast display vector pETcon (Addgene plasmid #41522) using standard restriction cloning. Sequences were verified by Sanger sequencing (Genewiz, South Plainfield, NJ) and listed in **Note C3**.

4.3.3. Degenerate Oligos and Oligo Pool Design

All degenerate "NNN" mutagenesis oligos were designed using Quick Change Primer Design Program (www.agilent.com) and were ordered from Integrated DNA Technologies. A single 7,118-member oligonucleotide library pool was custom synthesized by Agilent Technologies (sequences listed in Medina-Cucurella et al., under review, *PEDS*).

4.3.4. Preparation of Mutagenesis Libraries

Single and double-site saturation mutagenesis libraries were constructed using nicking mutagenesis as described in Wrenbeck et al. 2016[88] with the following changes. To conserve the 20:1 template to oligonucleotide ratio, the volume and concentration of oligos are determined using **Supplementary Note C4** for the AmiE and PYR1 libraries. The UCA9 libraries were prepared using an additional 1000:1 dilution of the oligo pool. The AmiE library covered residues Met1 – Pro69, the PYR1 targeted library covered Val81 – Arg167, and the UCA9 library covered Pro15 – Gln114.

4.3.5. Deep Sequencing Preparation and Data Analysis

The mutagenesis plasmids were prepared for deep sequencing exactly as described in Kowalsky et al. 2015, following 'method B'. Then, libraries were pooled and sequenced on an Illumina MiSeq using 2 x 250bp pair-end reads by the BioFrontiers Sequencing Core at the University of Colorado, Boulder. Primers used for deep sequencing are listed in **Table E9** and a summary of statistical results are in **Table E10**. The software package PACT[87], freely available at GitHub (https://github.com/JKlesmith/PACT/), was used to calculate the sequencing counts

obtained from raw FASTQ files. Raw sequencing reads for this work have been deposited in the SRA (SAMN10992661 – SAMN10992668).

4.3.6. Yeast two-hybrid screening

The PYR1 double mutant library was transformed into yeast two-hybrid reporter strain MaV99 pACT-HABI and tested for responsiveness to 1 μM mandipropamid (Sigma-Aldrich, St. Louis, MO) as previously described by Park et al. 2015[86].

CHAPTER 5

5. Summary and Future Approaches

5.1. Summary

As demonstrated in this dissertation, next-generation sequencing confers the ability to assign the functional effect of thousand mutations in one protein partner before and after a high-throughput selection for function. During the course of all these studies, we presented a standardized pipeline of comprehensive mutagenesis, yeast surface display, FACS, and deep sequencing to address fundamental aims and limitations in the field of protein engineering. The rich informational datasets allow to map conformational epitopes of potential mAbs, to determine binding sites of target antigen receptors, and to engineer antigens to express them in conformationally active forms. This technique can be extended to evaluate multiple protein complexes, enzyme engineering, antibody affinity, paratope mapping, and plant ligand-receptor modules, among others.

In **Chapter 2** we aimed to understand the interactions between interleukin 31, a cytokine involved in chronic skin inflammations, and its receptors. Such information can lead to the development of antagonist anti-IL-31 mAbs to inhibit the downstream signaling pathway. Our binding sites largely overlapped with the previous sites described by Le Saux et al.[20] and also revealed a new overlapping site between both receptors not described before. Furthermore, the mapped epitope of a candidate mAb suggested its efficacy by antagonizing IL-31 signaling pathway by binding mostly to the overlapping site between receptors.

Although our method is suitable for studying sequence-function relationships between PPIs, there were some limitations. **Chapter 3** described a deep sequencing-guided protein engineering workflow developed to increase the production of displayed proteins in a conformation recognizable by the binder partner. In this case, we aimed to increase the amount of folded canine nerve growth factor, a neurotrophin involved in multiple chronic pain conditions[58].

Mutational libraries created using this engineering pipeline revealed new insight into the role of the neurotrophin pro region. A combination of beneficial mutations within the pro-region of NGF allowed us to (i.) enhance the display of mature NGF on the yeast surface by a 23-fold above background and (ii.) determine the conformation epitopes of multiples anti-NGF mAbs[51]. This research proves that yeast surface display platforms can be engineered even for complicated mammalian proteins.

To continue improving deep mutational scanning, in **Chapter 4** we presented the integration of oligo pool synthesis technology with nicking mutagenesis to prepare precise and focused mutagenic libraries for multiple proteins which was easily extended to double mutant libraries. This technology avoids the need of hand-mixing oligos which improved the even representation of nucleobases with a near-complete coverage of targeted mutations. We speculate that oligo pools can be used for future protein science studies. All these improved throughput technologies will definitively contribute to the design of new strategies for the development of safe and efficient therapeutics and vaccines.

5.2. Future Approaches

Although we addressed multiple limitations from deep mutational scanning pipelines, there still some room for cost-effective and time-consuming improvements. As described by Kowalsky et al.[12], conformational epitope maps for new antibody-antigen interactions can be obtained in at least 3-5 weeks which is mostly limited by the design and preparation mutagenesis libraries, and screening through FACS. To address such concerns, the computational modeling software Rosetta is useful to predict protein structures and folding mechanisms, and to model multiple protein-ligand interactions[89,90]. Given two partner proteins and a random initial orientation, the Rosetta Docking[91] algorithm predicts bound state structures and binding pockets of the interacting proteins

with the lowest energy. It uses a score function composes of multiples energetic terms involved in common non-covalent interactions. Unfortunately, these computational models might not represent an optimal time-consuming improvement as users obtain thousands of docked models after a single run. Thus, researchers will need to evaluate each model to select the best ones based on individual assumptions. Nevertheless, these new designs allow us to design strategic point mutants for either additional modeling or experiments. For example, given an accurate three-dimensional structure of the partner protein displayed on the yeast surface, we could design mutagenic oligos for only surface residues. In addition, and by having a predicted model for the bound state of the protein:protein complex, oligos could also be design only for positions that are expected to disrupt the interaction.

As we showed in all these projects, mutational datasets allow us to the epitope positions for multiple mAbs. Our lab has applied similar tools for other antibody applications. First, we mapped the site of an antibody that binds to an antigen, the paratope, by displaying scFvs in the surface of the yeast. Second, we used our workflow to engineer antibody specificity. For a candidate mAb with affinity to two related targets, these workflows can identify mutations that improve the specificity of a mAb for one target while decreasing the binding affinity for the second target. Consequently, we are confident that similar techniques can also be applied for the evaluation of other antibody forms including bispecific and polyclonal antibodies.

Although mAbs have become the lead molecules for therapeutics and vaccines applications, there are some potential clinical and commercial limitations including the selectivity and high costs. One prominent approach to address these limitations is by integrating therapeutic Abs into bispecific formats. Bispecific antibodies (bsAbs) were designed to simultaneously bind to two different antigens or different epitopes within the same antigen. Some improvements could

be implemented to deep mutational scanning to screen multiple antigen constructs in a single reaction. One option is the incorporation of unique barcodes next to all genes to be displayed on the yeast surface. Then, epitopes for each displayed antigen is determined by following the same screening steps through FACS and deep sequencing. However, new custom scripts will need to be developed to determine the contribution of individual assemblies based on barcode sequences.

Similar high-throughput approaches can be used to delineate the bulk human antibody response upon a vaccination or an infection. Some viral diseases have no approved vaccines with efficacy for all age-groups. One example is Dengue virus (DENV), a worldwide mosquito-borne viral disease[92,93]. The main challenge for development of Dengue vaccines is the antibody-dependent enhancement[94] (ADE) of the infection where the pre-existing, poorly cross-reactive antibodies developed by the human immune response after primary infection tend to increase the secondary heterotypic dengue infection[95,96]. Thus, new throughput technologies are need to deconvolute complex immune responses against this kind of infection. New experiments could give us valuable information (i.) to trace the behavior of the complex polyclonal response upon vaccination or infection, (ii.) to understand the distribution of epitopes targeted by individual antibodies and (iii.) to predict the immunodominant and subdominant responses for any given complex polyclonal antibody mixture. To that end, we expect that all these future applications will have an enormous impact in the biomedical field by offering innovative epitope binning and mapping methods.

APPENDICES

APPENDIX A:

Characterizing protein-protein interactions using deep sequencing coupled to yeast surface display

Portions of this appendix were adapted from "Characterizing protein-protein interactions using deep sequencing coupled to yeast surface display" in *Methods in Molecular Biology* (**2018**) 1764:101-121

A.1. Introduction

In this appendix, we provide a detailed protocol to determine relative binding affinities and conformational epitope maps for PPIs (overview in **Figure 1**). We cover creation of single site saturation mutagenesis (SSM) libraries using nicking mutagenesis[23], transformation of libraries into yeast by the method of Gietz and Woods[97], screening of the SSM YSD library using FACS, DNA preparation for sequencing on an Illumina platform, and data analysis to determine a relative binding score and conformational epitope map. Relative binding calculations and estimated errors are carried out according to methods described in Kowalsky *et al.*[12]. Note: we assume the end-user has (i.) one PPI partner successfully induced and displayed in a YSD format with the other partner biotinylated; with (ii.) reproducible measurement of the apparent dissociation constant using protocols as described in Chao *et al.*[17].

A.2. Materials

A.2.1. Yeast and Bacteria Strains and Plasmid

1. **Yeast Strain:** *Saccharomyces cerevisiae* strain EBY100 is available at American Type Culture Collection and prepared to be chemically competent according to Gietz and Woods[97]. (see a shortened protocol in ***Section A.3.1.4***)
2. **Bacterial Strain:** *Escherichia coli* strain XL1-Blue high-efficiency electrocompetent cells are available through Agilent Technologies. Other competent cells with at least $1x10^9$ transformants per μg of plasmid can be used.
3. **Yeast display vectors**: The YSD vector used, pETconNK, is freely available on Addgene (plasmid #81169)[75]. The gene of interest is inserted between NdeI and XhoI restriction sites.

A.2.2. Nicking Site-Saturation Mutagenesis (SSM) Library Preparation

All enzymes and buffers for SSM library preparation with Nicking Mutagenesis are from New England BioLabs Inc. (NEB) unless noted otherwise.

A.2.2.1. Reagents, Media and Plates:

1. pETconNK plasmid containing gene of interest (freshly prepared from a dam^+ bacterial strain)
2. Nuclease-free water (NFH_2O, Integrated DNA Technologies)
3. Custom mutagenic primers (see Section ***A.3.1.1***)
4. SEC_Rev primer: 5' – CAAGTCCTCTTCAGAAATAAGCTTTTGTTC – 3'
5. T4 Polynucleotide Kinase Buffer
6. 10X CutSmart Buffer
7. 5X Phusion HF Buffer
8. 10 mM ATP
9. 50 mM DTT
10. 50 mM NAD^+
11. 10 mM dNTPs
12. 50% v/v sterile glycerol solution using deionized H_2O
13. **TB media**: 4.76% w/v of TB powder (pre-mixed), and 0.8% v/v of glycerol. Sterilize by autoclaving.
14. **LB agar plates**: 2.5% w/v of LB powder (pre-mixed), and 1.5% w/v of agar. Sterilize by autoclaving.

* Add kanamycin to a final concentration of 30 μg/ml when preparing the small plates to calculate the transformation efficiency and the large bioassay dishes for SSM Libraries (see Section ***A. 3.2.1***).

A.2.2.2. Enzymes:

1. 10 U/μl T4 Polynucleotide Kinase
2. 10 U/μl Nt.BbvCI
3. 10 U/μl Nb.BbvCI
4. 100 U/μl Exonuclease III
5. 20 U/μl Exonuclease I
6. 2 U/μl Phusion High-Fidelity DNA Polymerase
7. 40 U/μl Taq DNA Ligase
8. 20 U/μl DpnI

*Diluent for all enzymes required for Section **A.3.2.1** is 1X NEB CutSmart Buffer.

A.2.2.3. Equipment and Materials:

1. Zymo Clean & Concentrator-5 kit (Zymo Research)
2. Corning square bioassay dishes, 245 mm x 245 mm x 25 mm (Sigma-Aldrich)

A.2.3. Chemically Competent Library Yeast Transformation

A.2.3.1. Yeast Solutions and Plates:

1. **Growth Media:** Synthetic Dextrose medium supplemented with Casamino acids (SDCAA): 2% w/v dextrose (D-glucose), 0.67% w/v yeast nitrogen base without amino acids (Sigma-Aldrich), 0.5% w/v bacto casamino acids technical (BD Biosciences), 0.54% w/v Na_2HPO_4, and 0.856% w/v $Na_2HPO_4 \bullet H_2O$. Filter sterilize. Add 1% v/v of 10,000 U/ml Penicillin-Streptomycin immediately prior to growth to prevent bacterial contamination.

2. **Induction Media:** Synthetic Galactose medium supplemented with Casamino acids (SGCAA): prepare like SDCAA but with 2% w/v of galactose instead of dextrose.
3. **SDCAA agar plate:** 0.54% w/v Na_2HPO_4, 0.856% w/v $Na_2HPO_4 \bullet H_2O$, 18.2% w/v sorbitol, and 1.5% w/v agar. Sterilize by autoclaving. 2% w/v dextrose (D-glucose), 0.67% w/v yeast nitrogen base without amino acids, 0.5% w/v bacto casamino acids technical. Sterilize by filtrating. Add the filter sterilized solution into the cool autoclaved mix (approximately below 50°C) at 1:10 ratio. Store for up to 6 months at 4°C.
4. **Yeast storage buffer:** 20% w/v glycerol, 20 mM HEPES, and 150 mM NaCl pH 7.5. Filter sterilize.

A.2.3.2. Reagents:

1. 10 mg/ml Salmon Sperm DNA (Invitrogen)
2. 50% w/v Polyethylene Glycol, PEG Filter sterilize.
3. 1 M Lithium Acetate, LiOAc

A.2.4. Library Screening

A.2.4.1. Buffers and Reagents:

1. Phosphate buffered saline, PBS, at pH 7.4: 0.8 w/v NaCl, 0.02% w/v KCl, 0.144% w/v Na_2HPO_4, and 0.024% w/v KH_2PO4. Sterilize by filtrating.
2. Phosphate buffered saline with bovine serum albumin (PBS-BSA) at pH 7.4: prepare as PBS and supplemented with 0.1% w/v bovine serum albumin (BSA). Sterilize by filtrating.
3. Anti-c-myc-FITC antibody, FITC (Miltenyi Biotec)
4. Streptavidin, R-Phycoerythrin Conjugate, SAPE (Thermo Fisher)
5. Biotinylated PPI partner protein (see ***Note A1***).

A.2.5. Deep Sequencing Preparation of Yeast DNA

A.2.5.1. Buffers and Reagents*:*

1. **TE media**: 10 mM Tris-HCl at pH 8.0 and 0.1 mM EDTA
2. 5 U/µl Zymolyase (Zymo Research)
3. 10X Lamba Nuclease buffer (NEB)
4. SYBR- Gold Nucleic Acid Gel Stain (Thermo Fisher)
5. Agencourt AMPure XP (Beckman Coutler)
6. Quant-it Pico Green dsDNA Assay kit (Life Technologies)
7. 70% v/v Ethanol

A.2.5.2. Enzymes:

1. 5,000 U/ml Lambda Nuclease (NEB)

A.2.5.3. Equipment:

1. Zymo Research Yeast Plasmid Miniprep II kit
2. Qiagen mini-prep kit
3. 96-well magnetic plate

A.3. Methods

A.3.1. Library Preparation: Site-Saturation Mutagenesis (SSM)

Because a protein of 250 amino acids is encoded by a 750-bp gene, separate SSM libraries are prepared for the gene of interest (**Figure A1a**) to allow compatibility with 250 bp paired end (PE) Illumina MiSeq sequencing reads (see ***Note A2*** for considerations for library preparation).

A.3.1.1. Design of mutagenic oligonucleotides

SSM libraries are created using degenerative oligonucleotides containing a "*NNK*" codon to

cover all possible point mutations, where N represents any of the A/T/G/C, and K represents T/G. Mutagenic oligos are designed to be complementary to the wild-type template sequence as determined by the orientation of the BbvCI restriction site on the pETconNK vector (**Figure A1b-c;** see ***Note A3***).

1. Design your mutagenic oligos using Quick Change Primer Design Program (www.agilent.com). Use a degenerate "***NNK"*** codon to cover all possible 20 amino acids at each codon position.
2. Order the mutagenesis oligos on a 500 picomole DNA Plate Oligo from Integrated DNA Technologies and resuspend to 10 µM in TE, pH 8.

A.3.1.2. Preparation of SSM Libraries by Nicking Mutagenesis

This protocol is exactly as described in Wrenbeck *et al*.[88]. All reactions should be prepared on ice unless otherwise stated.

1. To phosphorylate the oligos, make a mixture comprising 5 µl of each NNK mutagenic primer.
2. Into a PCR tube, add 20 µl of the 10 µM mutagenic primer mixture, 2.4 µl of T4 Polynucleotide Kinase buffer, 1µl of 10 mM ATP, and 1 µl of T4 Polynucleotide Kinase. Incubate the reaction mixture at 37°C for 1 hour.
3. At the same time and in a separate PCR tube, add 18 µl of NFH_2O, 3 µl of T4 Polynucleotide Kinase buffer, 7 µl of 100 µM SEC_Rev primer, 1µl of 10 mM ATP, and 1 µl of T4 Polynucleotide Kinase. Incubate the reaction mixture at 37°C for 1 hour.
4. Store phosphorylated oligos at -20°C.
5. The day of mutagenesis, dilute phosphorylated mutagenic primers 1:1000 and SEC_Rev primer 1:20 using NFH_2O (*see* ***Note A4***).

6. For the preparation of ssDNA template strand, in a PCR tube, add 0.76 pmol of dsDNA plasmid (approximately 2-3 µg), 2 µl of 10X CutSmart buffer, 1µl of 1:10 diluted Exonuclease III (final concentration of 10 U/µl), 1 µl of Nt.BbvCI, 1 µl of Exonuclease I, and NFH_2O to 20 µl final volume.
7. Place the tube in a preheated (37°C) thermal cycle with the following program: 60 minutes at 37°C, 20 minutes at 80°C, and hold at 4°C.

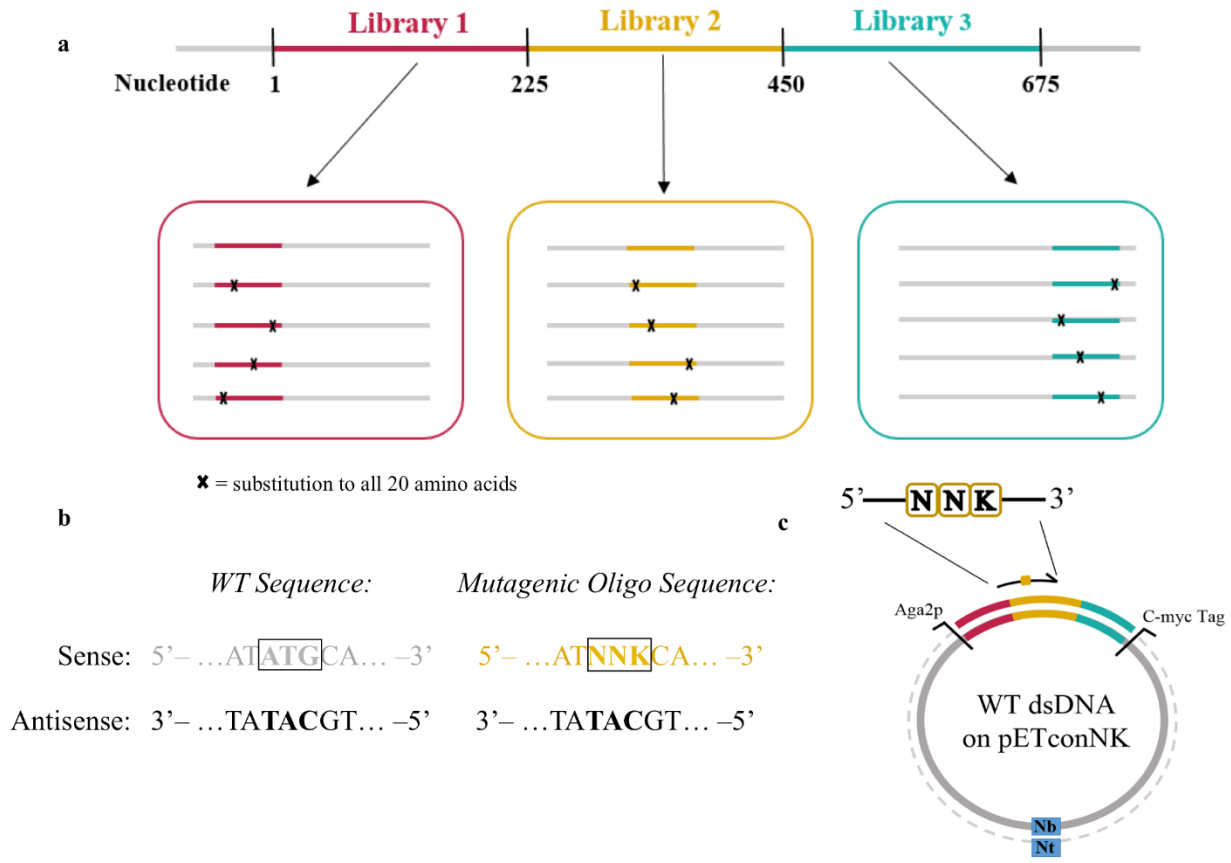


Figure A 1. Essential considerations needed for preparing Site Saturation Mutagenesis (SSM) Libraries. (**a.**) The gene of interest is segmented in multiple libraries containing contiguous sections of 200-250 bp. Here, sections of 225 bp are shown for compatibility with 250 bp PE Illumina MiSeq sequencing. (**b.**) Each mutagenic oligo contains an "NNK" codon to cover all possible 20 amino acids. (**c.**) An Nt.BbvCI restriction enzyme (Nt) is used to create a nick on the sense strand. Mutagenic oligos are designed to be complementary to the antisense ssDNA template

8. To proceed with the comprehensive codon mutagenesis on the first strand, in each PCR tube, add 26.7 µl NFH_2O, 20 µl of 5X Phusion HF buffer, 4.3 µl 1:1000 diluted

phosphorylated mutagenic oligos, 20 µl of 50 mM DTT, 1 µl of 50 mM NAD^+, 2 µl of 10 mM dNTPs, 5 µl of *Taq* DNA Ligase, and 1 µl of Phusion High-Fidelity DNA Polymerase. Mix the tube content briefly.

9. Place the tube into a preheated (98°C) thermal cycler with the following program: 2 minutes at 98°C, 15x cycles of 30 seconds at 98°C, 45 seconds at 55°C and 7 minutes at 72°C, followed by a final incubation at 45°C for 20 minutes, and hold at 4°C. Add additional 4.3 µl oligo at the beginning of cycles 6 and 11.
10. Purify each reaction using a Zymo Clean and Concentrate kit to a final volume of 15 µl using NFH_2O according to the manufacturer's instructions.
11. To degrade the template strand, transfer 14 µl of the purified DNA product to a new PCR tube and add 2 µl of 10X CutSmart buffer, 2 µl of 1:50 diluted Exonuclease III (final concentration of 2 U/µl), 1 µl of 1:10 Nb.BbvCI (final concentration of 1 U/µl), and 1 µl of Exonuclease I.
12. Place the reaction tube in a preheated (37°C) thermal cycle with the following program: 60 minutes at 37°C, 20 minutes at 80°C, and hold at 4°C.
13. To synthesize the 2^{nd} mutagenic strand, add 27.7 µl NFH_2O, 20 µl of 5X Phusion HF buffer, 3.3 µl of 1:20 diluted phosphorylated SEC_REV primer, 20 µl of 50 mM DTT, 1 µl of 50 mM NAD^+, 2 µl of 10 mM dNTPs, 5 µl of *Taq* DNA Ligase, and 1 µl of Phusion High-Fidelity DNA Polymerase to the same reaction mixture. Mix the tube content briefly.
14. Place the tube in a preheated (98°C) thermal cycler with the following program: 30 seconds at 98°C, 45 seconds at 55°C, 10 minutes at 72°C, 20 minutes at 45°C, and hold at 4°C.

15. Add 2 µl of DpnI into each reaction tube and incubate the reaction for 60 minutes at 37°C to degrade methylated and hemi-methylated wild-type DNA.
16. Purify each reaction using a Zymo Clean and Concentrate kit to a final volume of 6 µl using NFH_2O according to the manufacturer's instructions.
17. Transform the entire 6 µl reaction product into E. coli XL1-Blue following standard electrocompetent transformation protocol[98].
18. After recovery, bring the final volume of the transformation to 2-2.5 mL with additional sterile media (TB media).
19. Prepare six 10-fold serial dilutions and plate 10 µl of each. To calculate the transformation efficiency, the next day count the section that contains between 10-100 colonies. It is important to obtain at least 99.9% coverage of the theoretical diversity of the library (*see **Note A5***).
20. Spread the remaining cells onto the prepared large BioAssay dishes.
21. Place in a 37°C humidity-controlled incubator overnight when BioAssay dishes have dried.

A.3.1.3. Extraction of dsDNA SSM Library Plasmid

1. On the next day, scrape the large plates using between 5-10 mL TB media and collect the cells in a 50 ml centrifuge tube.
2. Vortex the cell suspension and extract the library plasmid DNA of a 1 ml aliquot of the cell suspension using a Qiagen mini-prep kit. Additional mini-preps can be done if large amounts of library DNA are required.

3. Store the rest of the cells at -80°C by resuspending the pellet in 3 ml of 50% v/v glycerol.

A.3.1.4. Chemically Competent Library Yeast Transformation

Competent yeast can be prepared up to six months ahead of time.

1. Grow the EBY100 cells in 500 mL YPD to an OD_{600} of 1.2 and are harvested at 4000xg for 5 minutes.
2. Resuspend the pelleted cells in 250 ml sterile H_2O, and repellet.
3. Resuspend the pelleted cells in 10 ml of 100 mM LiOAc and repellet.
4. Resuspend in 3.5 ml of 100 mM LiOAc and then, add 1.5 ml of 50% v/v glycerol and the mixture vortexed.
5. Prepare aliquots of 210 µl of cells to a tube and store at -80ºC. Do not snap-freeze cells.
6. Boil 30 µl of Salmon Sperm DNA at 97°C for 10 minutes.
7. Add 720 µl of 50% PEG, 108 µl of 1 M LiOAc, and 30 µl of boiled Salmon Sperm DNA to 210 µl of chemically competent EBY100 cells.
8. Vortex hard until there is a uniform mixture.
9. Add 5 µg of library plasmid to the mixture and vortex briefly.
10. Incubate the mixture at 30°C for 30 minutes.
11. Heat shock the cells by incubating at 42°C for 20 minutes.
12. Pellet the cells by spinning at 14000 rpm for 30 seconds.
13. Resuspend the cells pellet in 1 ml of SDCAA media and let stand for 5 minutes.
14. Prepare six 10-fold serial dilutions from the suspension and plate on SDCAA plates using 10 µl of each. Incubate for 2-3 days at 30°C to calculate transformation efficiency (*see **Note A5***).
15. Add the remaining culture into 100 ml of SDCAA media. Grow for 30 hours at 30°C and 250 rpm.

16. On the next day, resuspend the cell culture at OD_{600}=1 in 50 ml of SDCAA media.
17. Grow overnight at 30°C and 250 rpm.
18. Prepare multiple cells stocks by pelleting, resuspending in yeast storage buffer to an OD_{600}=1, and storing in 1 mL aliquots (approximately $1x10^7$ cells) at -80°C. Do not snap-freeze cells (*see **Note A6***).

A.3.2. Library Screening

A.3.2.1. Preparation of Labeling Reactions

1. For each PPI partner to analyze, thaw a 1 ml aliquot as prepared on previous section, spin down at 2500g for 3 minutes, and remove the supernatant.
2. Resuspend the pellet in 1 ml SDCAA media and grow for 4-6 hours at 30°C and 250 rpm.
3. Spin down the cells at 2500g for 3 minutes and re-inoculate at OD_{600} = 1.0 in 1 ml of SGCAA media. Induce overnight using the predetermined induction conditions (*see **Note A7***).
4. Spin down the cells at 2500g for 3 minutes, wash with 1 ml of ice-cold PBS-BSA, and spin down again.
5. Resuspend the cells in ice-cold PBS-BSA at an OD_{600} = 2.0.
6. In PBS-BSA, label 1 ml ($2x10^7$) cells with the biotinylated protein at half of the apparent dissociation constant and incubate at room temperature for 30 minutes using a table top mixer. Vary the total reaction volume to ensure that the number of biotinylated protein is at least 10-fold higher than the PPI partner that is displayed on the yeast cell surface. For example, assuming a 10:1 partner:displayed protein ratio at a typical PPI apparent dissociation constant of 10 nM, $2x10^7$ cells (1 ml) should be labeled with 5 nM biotinylated partner protein (half of the apparent dissociation constant). The total reaction volume is

calculated following equation A.3.1. Thus, label 1 ml of cells 2,305 µl of PBS-BSA with 16.6 µl of 1µM partner stock solution in.

$$Total\ Reaction\ Volume =$$

$$\frac{10^6 \mu l}{5 nmol\ partner} x \frac{10\ nmol\ partner}{1\ nmol\ displayed\ protein} x \frac{1\ nmol\ displayed\ protein}{6.02x10^{14} protein} x \frac{5x10^4 protein}{cells} x\ 2x10^7 cells =$$

$$3{,}322\ \mu l\ of\ total\ reaction\ volume$$

A.3.1

7. Spin down at 2500g for 5 minutes, wash the pellet with 5 ml of PBS-BSA and spin down and remove supernatant again. In this and subsequent steps, PBS-BSA should be ice-cold, the tabletop centrifuge should be refrigerated, and all tubes should be kept on ice and protected from light.
8. Label cells with 60 µl of FITC, 50 µl of SAPE and 1.89 ml of PBS-BSA, vortex briefly and incubate the labeled cells on ice for 10 minutes.
9. Repeat step 7.
10. Leave the cell pellet on ice until ready to sort.

A.3.2.2 Sorting Conditions Set-up

1. Set Gate1, Gate2, and Gate3 on your cell sorter as shown in **Figure A2**.
2. Add 4 ml of ice-cold PBS-BSA to the cell pellet, mix by vortexing, and transfer to a FACS-compatible tube.
3. Obtain the reference population by sorting 240,000 cells (*see **Note A8***) using Gate 1^+ (**Figure A2a**).
4. Obtain the displayed population by sorting 240,000 cells using Gate1^+/Gate2^+ (**Figure A2b**).

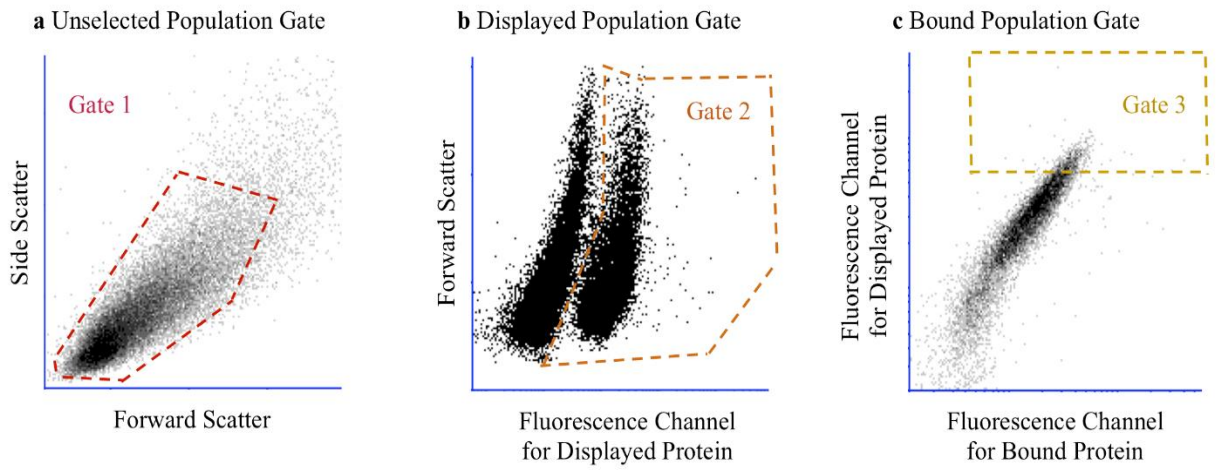


Figure A 2. Sorting gates used for library screening. Yeast SSM libraries are labeled with biotinylated complementary protein at half of the apparent dissociation constant. Next, SSM libraries are sorted using three different gates as shown: (**a.**) Gate 1 set with the light scatter parameters for yeast, forward scatter/side scatter; (**b.**) Gate 2 set on the forward scatter and the fluorescence channel for displayed protein (FITC); and (**c.**) Gate 3 set on the fluorescence channel for displayed protein and fluorescence channel for bound protein. Gate 3 is configured to collect the top 5-10% of the bound population.

5. Obtain the bound population for each PPI by sorting 240,000 cells using Gate1^{+}/Gate2^{+}/Gate3^{+} (**Figure A2c**).
6. Recover the collected cells in 5 ml of SDCAA media for approximately 30 hours at 30°C and 250 rpm.
7. Prepare cells stocks by storing 1 ml of OD_{600}=4 cell stocks in yeast storage buffer and at -80°C.

A.3.3. Deep Sequencing Preparation of Yeast DNA

A.3.3.1. Primer Design and Library Amplification Test

Yeast DNA is prepared for deep sequencing using a 2-step PCR amplification: the first step is with a gene-specific primer set (“inner” primers), while the second step uses an invariant set of “outer” primers (**Figure A3**). Inner primers are designed to be complementary to adjacent 5’ and 3’ ends of each library followed by an Illumina universal primer sequence (**Figure A3a**). The

following rules needs to be considered to determine these regions:

1. The length of the segment section plus the library should not be longer than 250 base pairs.
2. Design the segment region to have a melting temperature of 53-56°C using the NEB Phusion melting point calculator using Phusion High-Fidelity Polymerase.
3. Once the gene-specific sequence is designed, append the conserved primer sequence as shown in ***Table A1***.
4. Upon receiving the inner primers, we recommend performing a PCR verification with wild-type plasmid as a template to confirm a single band of the expected size.

Further steps for yeast DNA deep sequencing preparation requires the addition of universal primers to add the Illumina adapters and barcodes. Universal primers are designed using the TruSeq Small RNA Oligo Sequences. The forward primer is the same for all preparations while the reverse primer contains an indexing barcode that allows multiplexing of samples on an Illumina lane (**Figure A3b;** full sequences shown in **Table A1**).

A.3.3.2. Yeast plasmid DNA preparation for Deep Sequencing

1. Thaw an aliquot of the stored yeast library by hand, spin down at 2500g for 3 minutes, and remove the supernatant.
2. Resuspend the pellet cells in 200 µl of Solution 1 and add 5µl of 5U/µl of Zymolyase.
3. Incubate at 37°C for 4 hours and mix once per hour.
4. Perform one freeze-thaw cycle in dry ice/EtOH bath and 42°C incubation.
5. Add 200 µl of Solution 2, mix briefly and incubate for 3-5 minutes at room temperature.
6. Add 400 µl of Solution 3, mix well and centrifuge at 17,000g for 5 minutes.
7. Transfer the supernatant to a Qiagen mini-prep column and spin down for 1 minute at 17,000g.

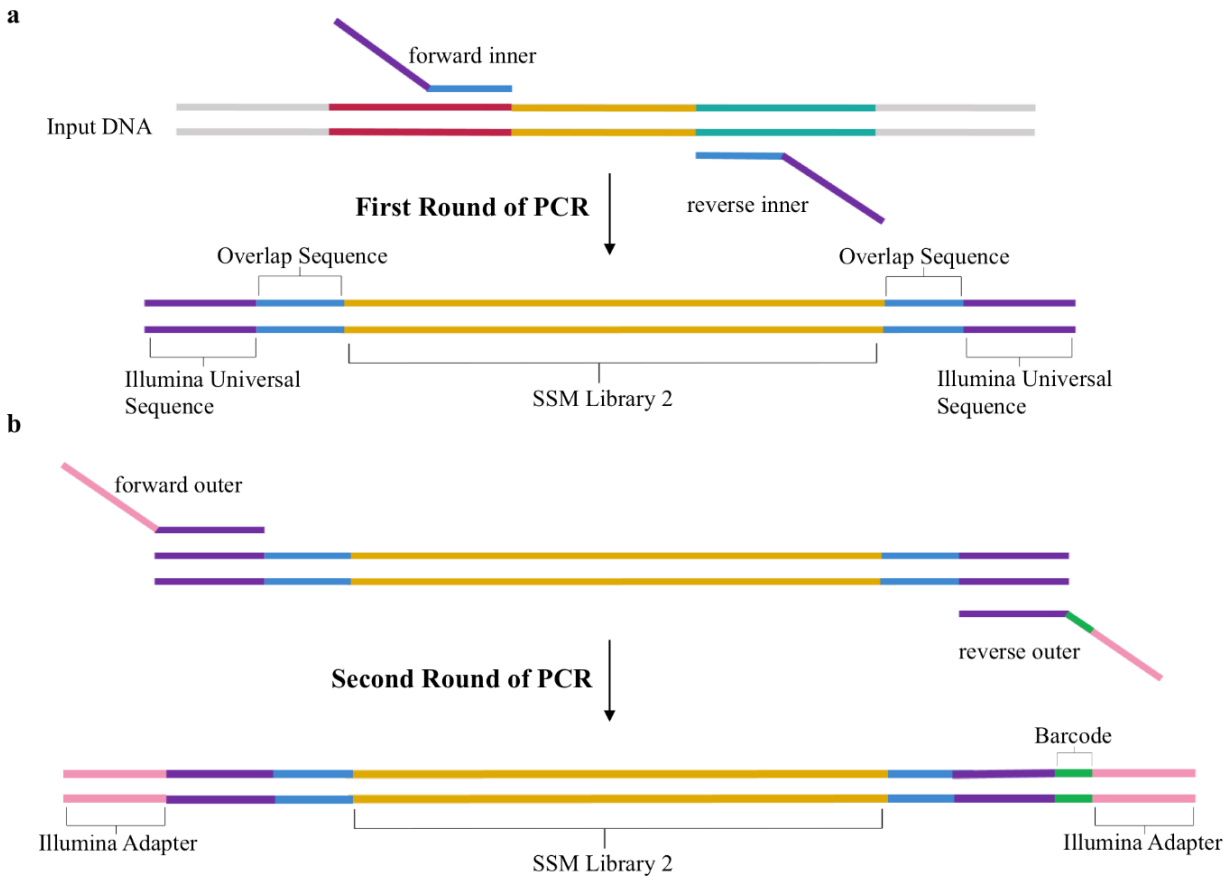


Figure A 3. PCR steps performed for deep sequencing preparation of SSM libraries. Sequential PCR reactions to amplify the genes of interest and attach the Illumina adapters are shown for SSM library 2 (gold). (**a.**) After extracting the plasmid DNA from yeast cells, SSM libraries are amplified by PCR using a set of inner primers containing a segment that overlaps with the gene of interest (light blue) and the Illumina universal sequence (purple). (**b.**) A second round of PCR is performed to attach the Illumina adapter sequence using a set of outer primers which contain an overlapping region to the Illumina universal sequence (purple), a unique barcode on the reverse primer (green), and Illumina adapter sequences (yellow).

8. Add 700 µl of PB buffer and spin down for 30 seconds at 17,000g.

9. Add 700 µl of PE buffer and spin down for 30 seconds at 17,000g.

10. Repeat step 9.

11. Take out supernatant and spin down again at 17,000g for 1 minute to dry the column.

12. Transfer the column to a new clean 1.5 ml microfuge tube, add 30 µl of elution buffer and spin down for 1 minute at 17,000g.

13. Reload the column with the eluate and spin down again. Store 15 µl of eluate and proceed with the remaining 15 µl.
14. For the purification of plasmid from the yeast preparation, in a PCR tube, add 15 µl of dsDNA plasmid, 2 µl of Exonuclease I, 1 µl of Lambda Nuclease, and 2 µl of 10X Lambda buffer.
15. Place the mixture in a preheated (30°C) thermocycler with the following cycle: 90 minutes at 30°C, 20 minutes at 80°C, and hold at 4°C.
16. Clean the PCR product following the standard procedure from Qiagen mini-prep PCR cleanup and elute in 30 µl of TE buffer.
17. Store 15 µl of eluate and proceed with the remaining 15 µl.
18. For the gene library amplification, in a PCR tube, add 10 µl of 5X Phusion HF buffer, 18.5 µl of NFH_2O, 1 µl of 10 mM dNTPs, 2.5 µl of 10 µM of forward inner primer, 2.5 µl of 10 µM of reverse inner primer, 15 µl of dsDNA template, and 0.5 µl of Phusion High-Fidelity Polymerase.
19. Place the tube in a preheated (98°C) thermocycler with the following cycle: 30 seconds at 98°C, 14x cycles of 5 seconds at 98°C, 15 seconds at 53°C and 15 seconds at 72°C, follow by a final incubation for 10 minutes at 72°C, and a hold at 4°C.
20. Add 1.87 µl of 1:10 diluted Exonuclease I.
21. Place the tube back in the thermocycler with the following cycle: 30 minutes at 37°C, 5 minutes at 95°C, and a hold at 4°C.
22. In a new PCR tube, add 10 µl of 5X Phusion HF buffer, 32.5 µl of NFH_2O, 1 µl of 10 mM dNTPs, 2.5 µl of 10 µM of forward outer primer, 2.5 µl of 10 µM of reverse outer primer,

1 µl of dsDNA template from previous PCR amplification, and 0.5 µl of Phusion High-Fidelity Polymerase.

23. Repeat the same PCR cycle used for the inner primers.
24. Run 5 µl of PCR product on 2% agarose gel and visualize with SYBR-Gold. It is important to verify that you have single clear band before proceeding (*see* ***Note A9***).
25. Purify and clean the PCR product using Agencourt AMPure XP following the manufacturer's instructions for the 96-well format procedure.
26. Measure the concentration of each sample.
27. Stored the purify product at -20°C.

Table A 1.Gene amplification and Illumina adapter primers to prepare samples for deep sequencing.

Inner Primers for Library Amplification	
Primer Name	***Sequence***
Inner_FWD	5'- gttcagagttctacagtccgacgatc<segment that overlaps to sense strand> - 3'
Inner_REV	5'- ccttggcacccgagaattcca<segment that overlaps to antisense strand> - 3'
Outer Primers to add the Illumina adapters and barcodes	
Illumina_FWD	5'- aatgatacggcgaccaccgagatctacacgttcagagttctacagtccgacgatc - 3'
Illumina_REV	5'- caagcagaagacggcatacgagat**nnnnnn**gtgactggagttccttggcacccgagaattcca - 3'
Magenta: Illumina Adapter, **nnnnnn**: Indexing Barcode (see Kowalsky *et al.*[25] for complete set), and Purple: Illumina Universal Sequence	

A.3.3.3. dsDNA Quantification using Quant-iT Pico Green

At this point, samples are ready for deep sequencing. Follow the instructions for the Illumina MiSeq 2x250 bp Submission from your Sequencing Facility. Usually, each Illumina MiSeq sequencing holds between 10-15 million reads per lane. Based on the read depth and library size, calculate the amount of reads necessary for each sample – our group uses approximately 500,000 reads per sample and multiplexes 20-30 samples per lane. Individual samples are quantified and mixed together in a single vial. The following procedure was adopted from the Invitrogen MP 07581 manual. The final yield should be about 1-4 ng in 40 µl.

1. Allow the Quant-iT reagent to warm to room temperature while covered in foil.
2. Prepare a 200-fold dilute solution of Quant-iT into TE buffer using a foil covered culture tube. (Example: 25 µl of Pico Green reagent into 4.975 ml of TE). This solution should be prepared and used the day of the experiment.
3. Beginning with a 50 ng/ml stock of a kit-supplied Lambda DNA standard, prepare a blank and a 1:2 standard curve (0, 1.56, 3.12,…,25 ng/ml) using the first column of a 96-well black plate.
4. In a black 96 well plate add 2.5 µl of each sample to 97.5 µl of TE in wells.
5. Carry out extra dilutions as necessary if the concentration is too high.
6. Add 100 µl of diluted Pico Green solution to DNA samples and standard samples, mix briefly, and incubate for 5 minutes at room temperature covered with foil to protect from light.
7. Measure the fluorescence of the samples (excitation ~480 nm, emission ~520 nm).
8. Subtract the fluorescence value of the reagent blank from that of each of the samples.
9. Use the corrected data to generate a standard curve of fluorescence versus DNA standard concentrations and calculate the concentrations of each sample. In our hands the final concentration is between 5-40 ng/µl.
10. Mix equivalent mass amounts of samples in a single 1.5 ml Eppendorf tube and send to your sequencing facility.

A.3.4. Data Analysis

Custom scripts used in the data analysis are available at Github (user: JKlesmith). Sample command lines and instructions are provided at the same source.

1. Use the modified version of Enrich 0.2 software as describe in Kowalsky *et al.* [25] to compute the enrichment ratios of individual mutants for the DNA sequencing results from Illumina MiSeq run (*see* **Figure A4** and ***Note A10***). Enrich 0.2[16] documentation is available at http://depts.washington.edu/sfields/software/enrich/docs/0.2/enrich.html. The output from Enrich 0.2 is required as input for the remaining steps. The wild-type protein sequence is also required as input for the following steps.
2. The relative binding of each variant on the displayed and bound population is calculated using a custom Python script called QuickNormalize.py (*see* ***Note A11***). The output from this script is a .csv file that can be read by multiple programs. In our lab we use Microsoft Excel to visualize the data as heatmaps and to carry out the data analysis (*see* ***Note A12***).
3. Calculate the Shannon Entropy for each variant on the displayed and bound population using a custom script called FACSEntropy.py - the output file is a .csv. The entropy values are used to discriminate those residues that participate in the protein-protein interaction and to determine the conformational epitope following the cut-off analysis flowchart as shown in **Figure A5** (*also see* ***Note A13***).
4. Calculate the reportable statistics using QuickStat.py script. Statistics will report the reads passing through enrich, the percentage of possible codon substitutions observed, the percent of reads with none, one, and multiple nonsynonymous mutations, and the coverage of possible single nonsynonymous mutations.

A.4. Notes

1. The PPI partner protein is chemically biotinylated following the instructions for EZ-Link NHS-Biotin Reagents (Thermo-Fisher). We prefer chemical biotinylation to genetically encoded biotinylation (e.g. avi-tag) as the former has a higher fluorescence signal. If

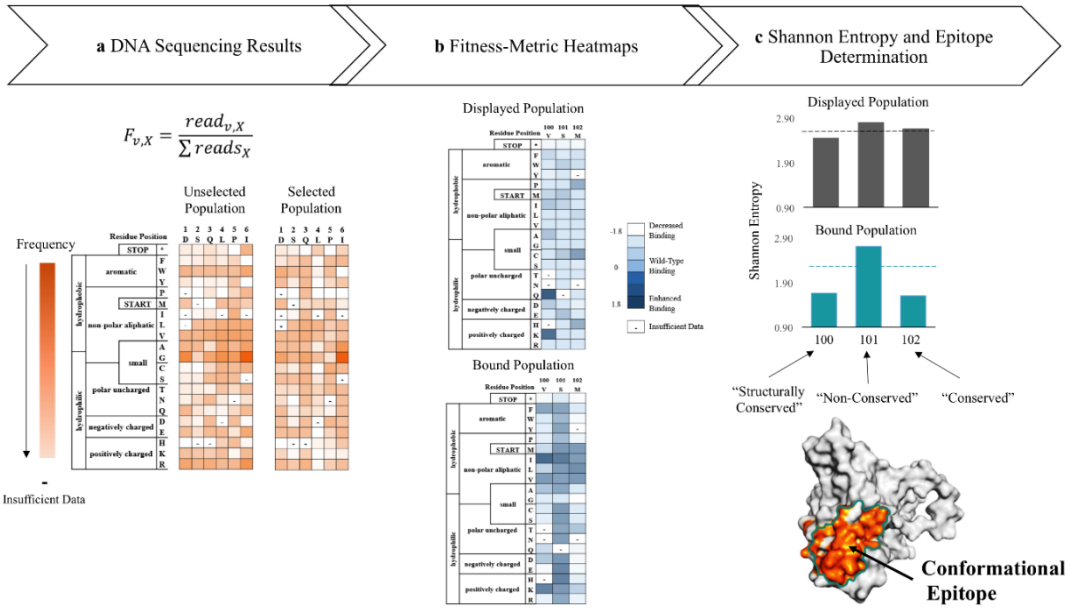


Figure A 4. Deep sequencing results and data analysis used to determine the conformational epitope. (**a.**) DNA sequencing results are processed using Enrich 0.2 software[99] to calculate the frequency, $F_{v,X}$, of each point mutant, v, for each position, x, in the primary sequence. (**b.**) The frequency data of each variant from different populations is transformed into heatmaps comparing the relative fluorescence of each variant in the displayed population (top) and the bound population (bottom) against the unselected population. (**c.**) Heatmaps are used to calculate the Shannon entropy for each residue on the displayed (black) and bound populations (turquoise). Next, the entropy is used to determine the conserved and non-conserved positions which allow to identify the conformational epitope.

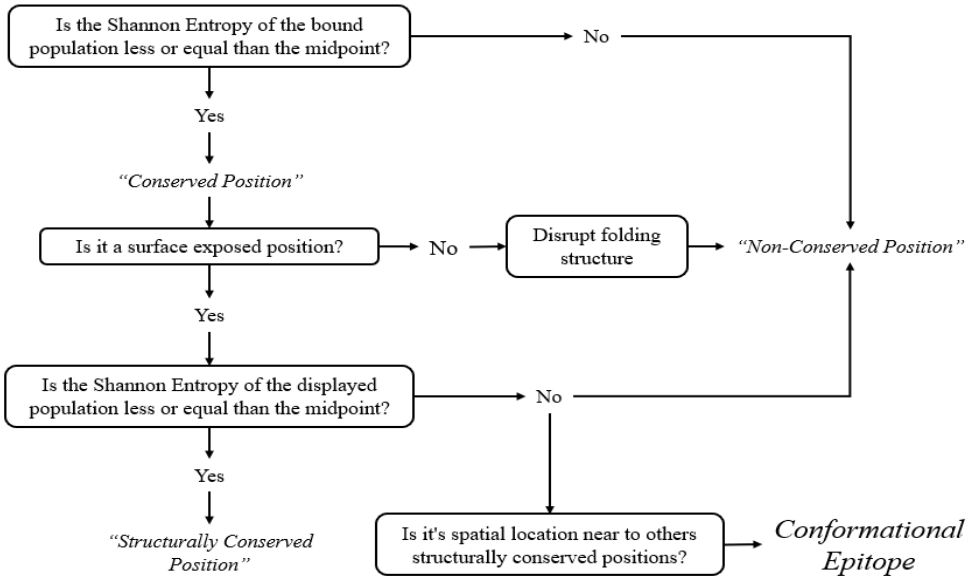


Figure A 5. Flow-chart of analysis used to determine the conformational epitope.

proteins are small, covalent labeling with multiple biotins may disrupt the structure; in such a case we recommend genetically fusing the PPI partner to a carrier like maltose binding protein or an IgG Fc. Anecdotally, we have noticed cleaner results with PPI partners with monovalent interactions, and for that reason recommend creating a Fab if the PPI partner is a mAb.

2. The following rules apply for preparing separate mutagenesis libraries: (i.) The length of each library should be divisible by three to avoid splitting a codon; (ii.) the gene should be segmented into libraries with a maximum length of 225 base pairs for Illumina 250 bp paired-end sequencing (273 base pairs for 300 bp paired-end sequencing); and (iii.) libraries should be similar in length (+/- three nucleotides).
3. In some cases, the gene sequence of interest also contains a BbvCI restriction site. If the site is in the same orientation as the site on the pETconNK plasmid, continue the protocol as usual. If the BbvCI site is in the opposite direction as the site on the pETconNK plasmid, use the YSD plasmid pETCON (Addgene # 41522) as this plasmid does not contain a nicking site. The orientation of the BbvCI may not be in the same way as exists on the pETconNK plasmid. For example, if the nicking site is in the opposite direction from **Figure A1b**, Nb.BbvCI (not Nt.BbvCI) should be used first to create the ssDNA wild-type template; otherwise follow the protocol as described in Section **A.3.1.2**.
4. We recommend preparing phosphorylated oligos no earlier than the day before the nicking mutagenesis procedure. Avoid repeated freeze/thaw cycles.
5. For a library with ***NNK*** SSM at 75 amino acids the theoretical library size is 2,400 nucleotide variants. The percentage theoretical coverage is described by the following

equation: $\%\ coverage = \left(1 - e^{\frac{-number\ of\ transformants}{theoretical\ library\ size}}\right) * 100$. In the above case, 16,500 transformants will give 99.9% coverage.

6. At this point, cells could be inoculated in fresh SGCAA media for Library Screening Preparation (Section **A**.**3.2**) or frozen aliquots can be prepared for long-term storage. We often re-inoculate the cells in SGCAA media to an OD_{600}= 1 and induce at 22°C to confirm that the mutagenesis libraries display on the yeast surface and binds the PPI partner. We prepare between 20-48 aliquots for long-term storage.
7. Induction temperature should be the same as used to prepare the PPI partner in a YSD format according to Chao *et al.*[17]. For each new YSD protein our lab tests induction of surface display at 18, 20, 22, and 30°C.
8. It is important that the number of collected cells should be at least 100-fold higher than the theoretical library size to avoid complexity bottlenecks. For example, at least 240,000 cells should be collected for each sorted population for a library with ***NNK*** SSM at 75 amino acids with a theoretical library size of 2,400 nucleotide variants.
9. If the correct band size was not obtained from the second PCR product, we recommend troubleshooting by running 5 µl of the first PCR product on a 2% agarose gel and visualized on SYBR-Gold to identify which PCR amplification did not work. We recommend staining the gel on SYBR-Gold for at least 1 hour to resolve low intensity bands. Fewer or more cycles of each PCR could be used to improve the product.
10. Our group routinely analyzes the quality of the Illumina sequencing data using FastQC available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Poor quality reads can hinder the data analysis using Enrich 0.2. The quality of the Illumina sequencing data is highest for the forward read and the first 150 bp. For issues where

quality is poor on the reverse read, perform Enrich only for the forward read. We have also performed Enrich for short segments of the reads where the quality is highest.

11. The relative binding (ζ_i) for variant *i* is defined as,

$$\zeta_i = log_2\left(\frac{\bar{F}_i}{\bar{F}_{wt}}\right), \quad \text{A.4.1}$$

where $\bar{F}_i$ is the mean fluorescence of variant i and $\bar{F}_{wt}$ is the mean fluorescence of wild type. There are a number of assumptions used to calculate relative binding – see Kowalsky et al.[12] for further details.

12. Positions with insufficient data at more than 10 substitutions should be excluded from analysis.

13. In the current experimental set-up, discriminating mutations that disrupt the interface and maintain the overall fold between those that destabilize the structure is difficult to determine, as unfolded mutants still predominantly display on the yeast surface [8,74]. However, a recent study confirms that destabilizing mutations display with fewer copies on the yeast surface than stabilizing mutations, at least for proteins with >200 residues [75] – for small proteins destabilizing mutants appear to display at the same rate as stable mutants (T.A.W. and A.M.C., unpublished data). To further identify mutations that stabilize larger proteins, a FACS protocol is used with a sort gate set to collect the top 5% of the displaying population. For library screening, $2x10^6$ yeast cells per ml, in PBS-BSA, are labeled with 1 µl of anti-c-myc-FITC per $2x10^5$ yeast cells. The population is sorted using a gate that collects the top 5% of the displaying population. Shannon entropy obtained from this study is used to identify structurally conserved positions.

APPENDIX B:

Preferential Identification of Agonistic OX40 Antibodies by Using Cell Lysate to Pan Natively Paired, Humanized Mouse-Derived Yeast Surface Display Libraries

Portions of this appendix were adapted from "Preferential Identification of Agonistic OX40 Antibodies by Using Cell Lysate to Pan Natively Paired, Humanized Mouse-Derived Yeast Surface Display Libraries" in *Antibodies* (**2019**) 8: 17

*Supplementary Materials are available online at http://www.mdpi.com/2073-4468/8/1/17/s1

B.1. Abstract

To discover therapeutically relevant antibody candidates, many groups use mouse immunization followed by hybridoma generation or B cell screening. One modern approach is to screen B cells by generating natively paired single chain variable fragment (scFv) display libraries in yeast. Such methods typically rely on soluble antigens for scFv library screening. However, many therapeutically relevant cell-surface targets are difficult to express in a soluble protein format, complicating discovery. In this study, we developed methods to screen humanized mouse-derived yeast scFv libraries using recombinant OX40 protein in cell lysate. We used deep sequencing to compare screening with cell lysate to screening with soluble OX40 protein, in the context of mouse immunizations using either soluble OX40 or OX40-expressing cells and OX40-encoding DNA vector. We found that all tested methods produce a unique diversity of scFv binders. However, when we reformatted forty-one of these scFv as full-length monoclonal antibodies (mAbs), we observed that mAbs identified using soluble antigen immunization with cell lysate sorting always bound cell surface OX40, whereas other methods had significant false positive rates. Antibodies identified using soluble antigen immunization and cell lysate sorting were also significantly more likely to activate OX40 in a cellular assay. Our data suggest that sorting with OX40 protein in cell lysate is more likely than other methods to retain the epitopes required for antibody-mediated OX40 agonism.

B.2. Introduction

Many antibody drugs bind to disease targets expressed on cell surfaces. For example, antibodies may bind to the surface of tumor cells and induce antibody-dependent cellular cytotoxicity (ADCC). Conventionally, antibody drug discovery groups use either hybridomas[100] or phage display[101] to discover antibody drugs. Hybridomas are typically screened for cell surface

binders using enzyme-linked immunosorbent assays (ELISAs) in 96-well plates[102]. Hybridoma methods, therefore, require expensive robotics to screen thousands of antibody candidates. Phage display has a much higher throughput, because billions-diverse phage libraries can be panned against cells affixed to well plates[103]. However, most therapeutic antibodies have been discovered in mice[26], perhaps due to difficulties with developability of artificial antibodies, such as low solubility binders discovered in phage display[104,105].

Recently, we invented a novel method for screening millions-diverse antibody repertoires using microfluidics, yeast display, and deep sequencing[28–30]. Our method leverages the developability advantages of naturally paired antibodies with the massively parallel throughput of display technologies. Other groups later further validated our work with similar methods[106,107]. However, our previously published methods required soluble antigen for both mouse immunization and fluorescence-activated cell sorting (FACS). This limitation excluded the possibility of using the method to identify antibodies against multi-pass transmembrane proteins, such as G-protein coupled receptors. Additionally, the requirement for soluble protein may lead to antibodies directed against spurious epitopes not present in the native conformation on the surface of target cells.

OX40, or tumor necrosis factor receptor superfamily member 4 (TNFRSF4), is a costimulatory immune receptor transiently expressed on T cells which upregulates T cell activity upon binding to its ligand, OX40L. Therapeutic agonism of OX40 may increase T cell differentiation and tumor killing functions[108]. Agonism requires a ligand binding to OX40 in a way that generates complexes of crosslinked OX40 molecules on cell surfaces[109]. Although the crystal structure of OX40 binding to OX40L has been resolved[110], the specific epitopes required for agonism are not well understood. Development of novel therapeutic antibodies would benefit from

a method that generates large panels of antibodies directed against a variety of OX40 epitopes that are bioavailable at the cell surface.

To improve OX40 antibody discovery, we adapted our previously published methods[28–30] to test different immunization methods (cells versus soluble antigen) and different antibody selection methods (cell lysate versus soluble antigen). The cell lysate selection method was adapted from prior work[111,112], specifically by using a peptide tag rather than biotin to label the cell lysate. We synthesized forty-one monoclonal antibodies (mAbs) from the various methods and found that soluble OX40 soluble antigen immunization followed by sorting with cell lysate was most likely to identify antibodies that bind cell surface antigen and yielded more antibodies that activate OX40 in cellular assays.

B.3. Materials and Methods

B.3.1. Mouse Immunization and Sample Preparation

All mouse work was performed at Antibody Solutions (Sunnyvale, CA, USA) and overseen by a licensed veterinarian. All experiments were performed using mice from Trianni (San Francisco, CA, USA), which are C57BL/6 that transgenically express a complete repertoire of fully human immunoglobulin gamma (IgG) and immunoglobulin kappa (IgK) V(D)J genes, but retain mouse promoters, introns, and constant domains.

For the soluble OX40 immunizations, five Trianni mice were immunized with recombinant His-tagged human OX40 extracellular domain (Acro OX40-H5224, Newark, DE, USA), using ALD/MDP (alhydrogel/muramyl dipeptide) as an adjuvant. 10 μg of OX40 protein with adjuvant was injected into the footpad twice per week for three weeks. We assessed titer at Day 21 with ELISA, using a dilution series of antigen, ranging from 1000 ng/mL to 1 ng/mL and goat anti-

mouse IgG-HRP (Jackson ImmunoResearch 115-035-071, West Grove, PA, USA) (**Supplementary Figure S1**). After assessing serum titer, two more footpad boosts of 10 μg without adjuvant were administered to each animal before sacrifice.

For the cells/DNA OX40 immunizations, we first transfected Flp-In 3T3 cells (Thermo Fisher Scientific, Waltham, MA, USA) with a vector encoding un-tagged, full-length human OX40 (**Supplementary Figure S2**). A pool of OX40-positive cells was selected using Hygromycin B (Gemini Bio 400123, West Sacramento, CA, USA) for 2 weeks. Cells were treated with Mitomycin C before cryopreservation. One to two million cells were injected per mouse. A footpad injection was performed with three Trianni mice on day 0 with cells, then days 3, 7, and 10 with 20 μg DNA plasmid encoding full-length, untagged human OX40, then day 14 with cells, day 17 and 21 with DNA, and final boosts on Days 24 and 27 with cells prior to tissue harvest. Before the final boosts, mouse serum titer was assessed with flow cytometry, using a dilution series of each animal's serum, starting at 1:200 and ending at 1:145,000 (**Supplementary Figure S3**). Briefly, the same 3T3 cells stably expressing OX40 were incubated with the serum dilution, washed, and then stained with goat anti-mouse IgG-rPE (Jackson ImmunoResearch, 115-116-071, West Grove, PA, USA). The final library was generated from two of the three mice, as the third mouse died prior to tissue harvest.

We surgically removed lymph nodes (popliteal, inguinal, axillary, and mesenteric) and spleens from the sacrificed animals. Single cell suspensions for spleen and lymph nodes were made by manual disruption followed by passage through a 70 μm filter. We used the EasySep™ Mouse Pan-B Cell Isolation (Stemcell Technologies, Vancouver, Canada) negative selection kit to isolate B cells from the single cell suspensions. Cells were stained for viability using Trypan blue and

then quantified with a C-Chip hemocytometer (Incyto, Chungnam-do, Korea). We then diluted the cells to 6000 cells/μL in PBS with 12% OptiPrep™ Density Gradient Medium (Sigma, St. Louis, MO, USA). The purified cell populations were used for microfluidic encapsulation as described below.

B.3.2. Generating Paired Heavy and Light Chain Libraries

As described previously[28–30], the generation of libraries comprised of three steps: (i) poly(A)+ mRNA capture, (ii) multiplexed overlap extension reverse transcriptase polymerase chain reaction (OE-RT-PCR), and (iii) nested PCR to remove artifacts and add adapter sequences for deep sequencing or yeast display libraries.

Briefly, we isolated 1.6–1.9 million B cells into fluorocarbon oil (Dolomite, Royston, UK) emulsion microdroplets (**Supplementary Table S1**) with a lysis buffer (20 mM Tris pH 7.5, 0.5 M NaCl, 1 mM EDTA, 0.5% Tween-20, and 20 mM DTT) and oligo(dT) beads (New England BioLabs, Ipswich, MA, USA), using an emulsion droplet microfluidic chip[28–30]. We purified beads from the droplets using Pico-Break solution (Dolomite, Royston, UK).

We then performed multiplex OE-RT-PCR in emulsions, using purified RNA-bound beads as a template, as described elsewhere[28–30]. The OE-RT-PCR product was gel purified and PCR was performed to add adapters for Illumina sequencing or yeast display; for sequencing, a randomer of seven nucleotides was added to increase base calling accuracy in subsequent next generation sequencing steps. Nested PCR is performed with 2× NEBNext High-Fidelity amplification mix (New England BioLabs, Ipswich, MA, USA) with either Illumina adapter containing primers or primers for cloning into the yeast expression vector.

B.3.3. Yeast Library Screening

Saccharomyces cerevisiae EBY100 cells (ATCC, Manassass, VA, USA) were electroporated (Bio-Rad Gene Pulser II; 0.54 kV, 25 uF, resistance set to infinity) with gel-purified nested PCR product and linearized pYD vector[28–30] for homologous recombination in vivo. Transformed cells were expanded and induced with galactose to generate yeast scFv display libraries.

For the soluble OX40 FACS experiments, human OX40-His (described above) protein was biotinylated using the EZ-Link Micro Sulfo-NHS-LC-Biotinylation kit (Thermo Fisher Scientific, Waltham, MA, USA). The biotinylation reagent was resuspended to 9 mM and added to the protein at a 50-fold molar excess. The reaction was incubated on ice for 2 hours, and then the biotinylation reagent was removed using Zeba desalting columns (Thermo Fisher Scientific, Waltham, MA, USA). The final protein concentration was calculated with a Bradford assay. The scFv libraries were then stained with anti-c-Myc (Thermo Fisher Scientific A21281, Waltham, MA, USA) and an AF488-conjugated secondary antibody (Thermo Fisher Scientific A11039). Biotinylated OX40 was added to the yeast culture (250 nM final concentration) and stained with APC-streptavidin (Thermo Fisher Scientific, Waltham, MA, USA). Approximately two million cells were then flow sorted on a FACSMelody (BD, San Jose, CA, USA) for double positive cells (AF488+/APC+). Populations of binder scFv clones were recovered, expanded, and then subjected to a second and third round of FACS with the same antigen at 250 nM final concentration. A fourth round of FACS was additionally performed on select samples (**Supplementary Figure S4**).

For the cells/DNA OX40 FACS experiments, we engineered an expression vector that expresses full-length human OX40 fused to a FLAG peptide at the N-terminus (**Supplementary Figure S2**). This vector was used to stably transfect CHO cells via targeted genome integration.

Approximately 12.5×10^6 OX40-positive transfected cells encoding full-length human OX40 were used to prepare the cell lysate for each staining condition. First, cells were harvested and washed twice with 10 mL of ice-cold PBS. Second, cells were resuspended in a lysis buffer (PBS, 1% Triton X-100, 2 mM EDTA, and 1 × protease inhibitor cocktail) to a final concentration of 5×10^7 cells/mL and were incubated, rotating for 30 mins at 4 °C[111]. Finally, cells were harvested and the supernatant (the detergent-solubilized cell lysate) was removed to a fresh tube and stored at 4 °C until use. The final total protein concentration in the lysate was calculated using a Bradford assay. The scFv yeast libraries were labeled with 250 µL of cell lysate and incubated, rotating, overnight at 4 °C. The next day, labeled yeast cells were stained with anti-c-Myc, an AF488-conjugated secondary antibody, and APC anti-FLAG (clone L5, BioLegend 637308, San Diego, CA, USA). Approximately, four million cells were flow sorted on a FACSMelody. As described above, the collected populations of binder scFv clones were recovered, expanded, and subjected to two additional rounds of FACS using the same cell lysate concentration.

B.3.4. Sequence Analysis

Libraries were sequenced on a MiSeq (Illumina, San Diego, CA, USA) using a 500 cycle MiSeq Reagent Kit v2, as described previously[28–30]. Sequencing was performed in two different runs. In the first run, we directly sequenced the scFv libraries to obtain a forward read of 357 cycles for the light chain complementarity-determining region (CDR)3 and V-gene, and a reverse sequence read of 162 cycles across the heavy chain CDR3 and part of the heavy chain V-gene. In the second run, we first used the scFv library as a template for PCR to independently amplify heavy and light chain V-genes. We then obtained a forward read of 255 cycles and a reverse read of 255 cycles for the heavy and light chain Ig separately. The second run yields overlapping reads, which is useful for sequencing error correction.

We used previously published methods for error correction, reading frame identification, and FR/CDR junction calls[28–30,113]. We discard reads with E > 1 (E is the expected number of errors), retaining sequences for which the most probable number of base call errors is zero. We also discard singleton nucleotide reads to further improve confidence in antibody sequences. In order to identify V and J gene families and calculate percent identity to germline, we aligned antibody nucleotide sequences with the IMGT database[114].

We define "clones" conservatively, with an emphasis on sequence accuracy. First, we concatenated the CDR3K and CDR3H amino acid sequences from each scFv sequence into a single contiguous amino acid sequence. Next, we used USEARCH[115] to compute the total number of amino acid differences in all pairwise alignments between each concatenated sequence in each data set. Groups of sequences with ≤2 amino acid differences in the concatenated CDR3s were counted as a single clone. Finally, we used the majority amino acid identity at each residue position to generate the consensus amino acid sequence of the clone from sequences of the members of the group.

To generate clonal cluster plots, we first used USEARCH[115] to generate all pairwise alignments across the complete set of FACS-sorted IgH and IgK scFv sequences (**Supplementary Tables S2-S9**). We then computed the total number of amino acid differences between each scFv sequence. We then generated clustering plots using the igraph R package[116], using the "layout_with_graphopt" option. Antibody clones are represented by "nodes" in the plots. The size of the nodes corresponds to the frequency of the antibody clone in the FACS-sorted population: small (<2% frequency), medium (2-12% frequency), and large (>12% frequency). An "edge" (a line linking nodes) was drawn between any sequences with ≤9 amino acid differences in the concatenated CDR3s.

B.3.5. Monoclonal Antibody Expression and Characterization

We synthesized mAbs by cloning antibody sequences into a variant of the pCDNA5/FRT mammalian expression vector (Thermo Fisher Scientific, Waltham, MA, USA), as described previously[28–30]. Expression constructs were prepared using a BioXP™ robotic workstation (SGI DNA, La Jolla, CA, USA). Human IgHG1 isotype was used for all constant domains. MAb plasmids were then transiently transfection into ExpiCHO cells (Thermo Fisher Scientific, Waltham, MA, USA). Transfected cells were cultured in ExpiCHO medium for 7–9 days. An IgG ELISA kit (Abcam, Cambridge, UK) was used to quantify the concentration of antibody in the supernatants.

To measure cell surface binding, we first generated stable human OX40-expressing Flp-In Chinese hamster ovary (CHO) cells (Thermo Fisher Scientific, Waltham, MA, USA). One million cells (1:1 mix of OX40 and irrelevant PD-1-expressing negative control cells) were stained with 1 μg of anti-OX40 mAb in 100μL MACS Buffer (DPBS with 0.5% BSA and 2mM EDTA) for 30 min at 4 °C. Cells were then co-stained with anti-human IgG Fc-PE (BioLegend clone M1310G05, San Diego, CA, USA) and anti-human PD-1-APC (BioLegend clone EH12.2H7, San Diego, CA, USA) antibodies for 30 min at 4 °C. We then used a FACSMelody (BD, San Jose, CA, USA) quantify binding. We used FlowJo to determine the intensity of the OX40-expressing cells versus the irrelevant negative controls (**Supplementary Figure S5**).

For measurement of the kinetics of binding to soluble OX40, 5 μg/mL antibodies were loaded onto a Protein A biosensor using the Octet Red96 system (ForteBio, Fremont, CA, USA) by a contract research organization (Bionova, Fremont, CA, USA). Loaded biosensors were dipped into His-tagged OX40 extracellular domain (Acro OX40-H5224, Newark, DE, USA) at 200 nM,

100 nM, and 50 nM, or 1600 nM, 800 nM, and 400 nM, depending on the strength of the response to the OX40 antigen binding the mAb. Kinetic analysis was performed using a 1:1 binding model and global fitting (**Supplementary Figure S6**).

To determine the ability of each mAb to activate OX40 in vitro, we used a kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. We performed OX40 activation assay in the presence of cells expressing FcγRIIB, which simulates the putative in vivo mechanism of OX40 cross-linking[109]. On the day prior to the assay, FcγRIIB/CHO-K1 cells were thawed into 95% RPMI 1640/5% FBS and plated into 96-well plates. After incubating for 5–7 hours at 37 °C, 5% CO_2, OX40-expressing Jurkat cells were thawed and added to the wells containing FcγRIIB/CHO-K1 cells. After incubating the cell mixtures overnight, antibodies were diluted in 95% RPMI 1640/5% FBS. The antibody dilutions were then added to the wells containing the cells. The cell/antibody mixtures were incubated at 37 °C, 5% CO_2 for 5 h, after which we added Bio-Glo Reagent. Luminescence was read using a Spectramax i3x plate reader (Molecular Devices, San Jose, CA, USA). IC50 was calculated by plotting RLU (relative luminescence units) vs concentration using SoftMax Pro (Molecular Devices, San Jose, CA, USA) (**Supplementary Figure S7**). In-house produced pogalizumab was used as a positive control, and an antibody binding to an irrelevant antigen was used as a negative control.

B.4. Results

B.4.1. Overview of the Experimental Approach

First, we stably expressed full-length human OX40 protein in mouse 3T3 cells. Next, we immunized transgenic humanized Trianni mice with either OX40-expressing 3T3 cells or soluble OX40 extracellular domain using a rapid immunization protocol. The cohort of mice immunized with OX40-expressing 3T3 cells was additionally boosted with a DNA vector driving expression

of full-length OX40 protein. All mice were checked for anti-OX40 serum titer (**Supplementary Figures S1, S3**) and sacrificed after approximately four weeks. Spleen and lymph nodes were disaggregated into single cell solutions, tissues from replicate animals were pooled, and B cells were isolated from the single cell solutions and cryopreserved.

We then used droplet microfluidics[28–30] to isolate millions of single cells from each experimental arm (**Supplementary Table S1**) into aqueous-in-oil picoliter droplets. Cells were lysed inside the droplets, and mRNA from the single cells was bound to oligo(dT) beads. The oligo(dT) beads were then injected into a second emulsion with multiplex primers that amplify heavy and light chain Ig. The primers are designed with overlapping linker sequences that physically link heavy and light chain Ig into scFv expression constructs. The linked Ig libraries are subjected to deep sequencing to quantify clonal antibody diversity (**Supplementary Table S1**). Each library was then electroporated into yeast for scFv display (**Figure B1**).

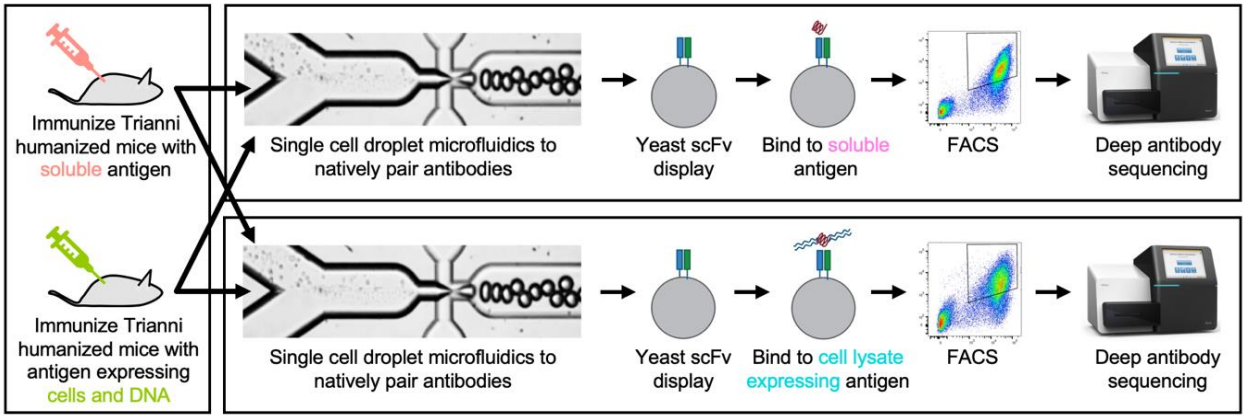


Figure B 1. Overview of the generation and screening of scFv libraries derivates from B cells from humanized mice with either soluble OX40 or cells and DNA expressing OX40. B cells are isolated from spleen and lymph nodes. Next, B cells are encapsulated into droplets with oligo-dT beads and a lysis solution to generate DNA amplicon that encodes the scFv libraries with native pairing heavy and light Ig. The scFv libraries are then transfected into yeast cells and labeled with either soluble OX40 or lysate from cells expressing OX40. Next, FACS is used to collect scFv with the highest FACS signal. Finally, deep sequencing is used to identify all clones in the pre- and post-sort populations.

Next, each yeast scFv library was subjected to FACS using either lysate from FLAG-tagged OX40-expressing CHO cells or soluble His-tagged OX40 extracellular domain (**Figure B2**). Cell lysate is prepared by lysing recombinant cells in a buffer containing a surfactant (1% Triton X-100) and quantified using a Bradford assay. The cell lysate was then incubated with each yeast scFv library, stained with an anti-FLAG secondary antibody and anti-c-Myc staining to quantify expression of scFv, and then subjected to FACS to pan for antigen-positive, c-Myc-positive binders. FACS with soluble OX40 was performed as described previously[28–30]. We performed either three or four rounds of FACS panning and deep sequenced the binders.

Finally, to compare the functional characteristics of antibodies identified with each experimental method, we synthesized forty-one monoclonal antibodies from the panning experiments. We chose a sample of antibodies from distinct putative clonal lineages, with an emphasis on the most common clones from each experimental method (**Supplementary Tables S2–S9**). We then used the full-length mAbs to perform kinetics measurements, cell surface binding assays, and in vitro cellular activation assays.

B.4.2. Analysis of Serum Titers

Soluble OX40 antigen yielded consistently high anti-OX40 serum titers in five replicate mice (**Supplementary Figure S1**). We pooled the splenocytes or lymph nodes from these five animals to produce one yeast scFv library for each tissue type, for a total of two soluble immunization OX40 libraries. Immunizations using OX40 cells/DNA were less consistent across three replicate mice, generating a non-responder, a medium responder, and a high responder (**Supplementary Figure S3**). Splenocytes or lymph nodes from two of three animals were pooled to produce a single natively paired yeast scFv library for each tissue type (the medium responder

animal died prior to tissue harvest). Thus, we generated a total of four natively paired yeast scFv libraries (2 tissues × 2 immunization methods = 4 libraries; **Supplementary Table S1**).

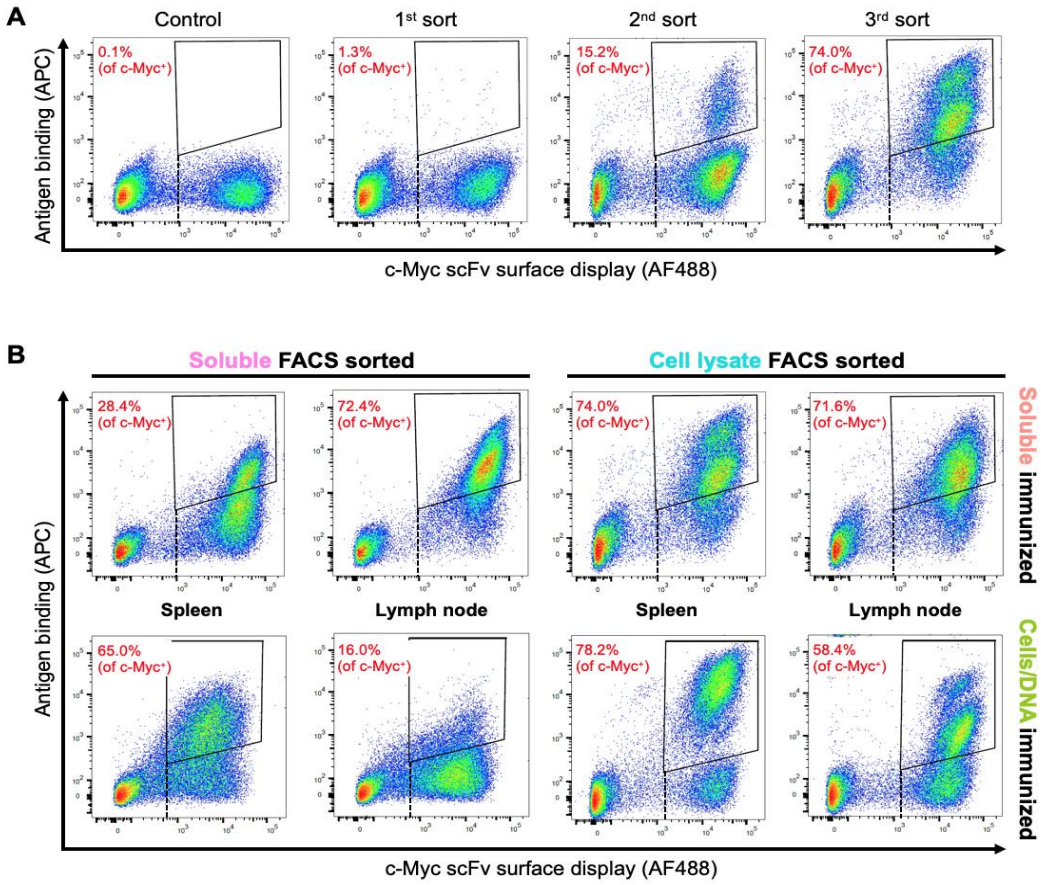


Figure B 2. scFv libraries from immunized mice subjected to FACS selection for OX40. An anti-c-Myc (AF488) staining is used to identify yeast displaying scFv on the cell surface (x-axis). APC-streptavidin is used to identify yeast cells with biotinylated soluble OX40 bound, and anti-FLAG (APC) staining is used to identify yeast cells bound to lysate from cells expressing OX40-FLAG (y-axis). A negative control is used to set a quadrangle gate for the FACS selection (upper right corner). The percentage in each quadrangle represents the proportion of c-Myc positive yeast cells that fell within the gate. (**a.**) An example of the soluble immunized, cell lysate sorted, spleen library subjected to three rounds of sorting to enrich the cells for positive antigen binding and scFv display on yeast surface. (**b.**) FACS plots showing the percentage of enriched antibodies after the three rounds of sorting under each condition. Note that the image used for soluble immunized, cell lysate sorted, spleen is the same image as the 3rd sort from A.

B.4.3. Selection of OX40 scFv Binders with FACS

Prior publications have described protocols for panning yeast surface scFv display libraries with biotinylated, detergent-solubilized cell lysate[111,112]. We reasoned that biotinylation was

suboptimal because the method labels all proteins in the cell lysate rather than only the target protein, leading to a loss of specificity in sorting and additional labor-intensive steps for every panning experiment. Therefore, we developed an approach based on an OX40 protein fused to a FLAG peptide tag. Briefly, we engineered an expression vector that expresses full-length human OX40 fused to a FLAG peptide at the N-terminus (**Supplementary Figure S2**). This vector was used to stably transfect CHO cells via targeted genome integration. Fresh cell lysate was prepared for each panning experiment by lysing recombinant cells in a buffer containing a surfactant (1% Triton X-100).

In a typical experiment, we obtained around 7-8 mg/mL of total protein concentration per 5.0×10^7 cells/mL. Prior work has shown that the sensitivity and specificity of scFv binder discovery are functions of the molarity of a soluble target used during panning[28–30]. We, therefore, tested panning with four different concentrations of cell lysate (**Supplementary Figure S8**). After two rounds of panning, the fraction of scFv binders was as low as 6.4% (0.475 mg/mL) and as high as 31% (3.8 mg/mL). The FACS plots at 1.9 mg/mL of cell lysate were qualitatively and quantitatively similar to our prior panning experiments using various soluble antigens at 7–70 nM[28–30]. Therefore, we used approximately 2 mg/mL of cell lysate for all subsequent cell lysate panning experiments.

Although all tissues, from both immunization methods, and with both FACS methods yielded scFv binders, there were qualitative and quantitative differences in the FACS plots (**Figure B2**). After three rounds of panning, the fraction of scFv binders was as low as 16.0% (cells/DNA immunization, soluble antigen FACS) and as high as 78.2% (cells/DNA immunization, cell lysate FACS). On average, the soluble immunogen yielded a higher fraction of scFv binders than the cells/DNA immunization (61.6% versus 54.4%, respectively), and the cell lysate FACS yielded a

higher fraction of scFv binders than the soluble antigen FACS (70.6% versus 45.5%, respectively). Because the cells/DNA immunogen followed by soluble FACS yielded a lower fraction of scFv binders, we performed a fourth round of panning on these libraries (**Supplementary Figure S3**), which improved the fraction of scFv binders by as much as 58.3% (from 16.0% to 74.3%), suggesting an increase in specificity. In general, cell lysate FACS produced a more significant shift in FLAG-APC fluorescence (antigen binding) than soluble antigen FACS.

B.4.4. Sequence Characteristics of OX40 scFv Binders

We deep sequenced the yeast scFv libraries before and after FACS (**Supplementary Table S1**), as described previously[28–30]. Note that we use extremely conservative error processing, which favors clone sequence quality over capturing the "long tail" of clonal diversity. Before FACS, the scFv libraries contained between 16,491 and 19,509 clones. After FACS, the scFv libraries were much more oligoclonal, containing between 61 and 238 clones.

We analyzed the most common (≥0.1% frequency) scFv sequences to determine pre- versus post-FACS clonal enrichments achieved by each method. We did not observe statistically significant differences between the mean clone counts of soluble versus cells/DNA immunizations, or between soluble versus cell lysate sorts ($p > 0.01$, *t*-test). The average pre-sort scFv clone abundance was 0.032%, with a range from 0% (not detected) to 0.71%. Sequences present in the post-sort libraries were not detected in the pre-sort libraries for 54/268 (20.1%) of clones, suggesting that many candidate binder clones were extremely rare in the mouse repertoires. The average enrichment between pre- and post-sort clone counts was 5056-fold, with a range from 2.2-fold to 500,000-fold. We note that prior work on Balb/c, SJL, and Medarex HuMAb mice[28,30] yielded similar levels of enrichment and clonal diversity both before and after FACS.

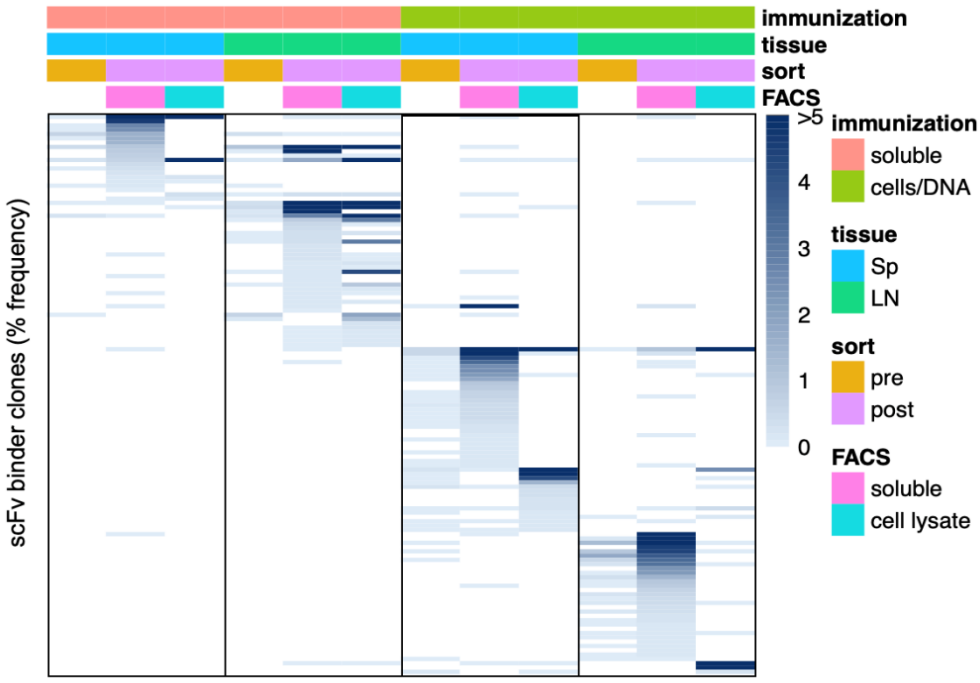


Figure B 3. Overlapping clones in the pre- and post-sort populations obtained from each experimental parameter. Clone frequencies are represented by the blue heatmap, with unique clones aligned into rows across the repertoires. We only show clones that are present with frequencies of 0.1% or higher in at least one of the post-FACS repertoires. We organize the repertoires by pre- vs. post-FACS, tissue of origin (lymph nodes vs. spleen), immunization method (soluble antigen vs. cells/DNA), and panning condition (soluble antigen vs. cell lysate). Done by GigaGen.

Next, to determine whether different methods discover the same scFv binders, we analyzed the most common (≥0.1% frequency) scFv sequences for overlap between each post-FACS library (**Supplementary Tables S2–S9**). Only 6.8% of enriched clones (16/235 non-redundant, unique clones) were shared between at least two of the eight series of scFv panning series (2 tissues × 2 FACS methods × 2 immunization methods) (**Figure B3**), suggesting different immunization methods and different FACS methods typically capture different sequences. Notably, 81.3% (13/16) of the shared clones were generated using the soluble immunogen and identified with both soluble and cell lysate FACS methods. The only three scFv clones that were shared between lymph node and spleen were generated using the cells/DNA immunization method, suggesting that cells/DNA induces a more systemic antigen response than soluble immunization.

A clonal cluster plot of full-length IgH and IgK sequences from scFv binder clones highlights similarities and differences among scFv binder sequences (**Figure B4**). An "edge" (a line linking nodes) was drawn between any sequences with ≤9 amino acid differences in the concatenated CDR3s. However, most clonal clusters comprised only a single clone, i.e., no related sequences were detected. Only 14 clonal clusters were comprised of five or more scFv clones (putative clonal lineages). Of those, 100% (14/14) comprised clones derived from only the soluble immunogen (using either FACS method) or only cells/DNA immunogen (using either FACS method). In general, sequence analysis suggests that each immunization method, FACS method, and tissue produces mostly unique clones. However, where there is overlap, FACS with different methods is more likely to generate similar clones than different tissues, and different immunization methods are least likely to generate similar clones.

We did not find any significant differences in the sequence characteristics of the most common (≥0.1% frequency) scFv clones between each post-FACS library. Sequence identity to germline (%ID) was high across all methods, averaging 98.4% for IgKV and 97.6% for IgHV. This suggests low levels of affinity maturation in vivo. Variable (V)-gene diversity was low across all methods (Supplementary Figures S9–S10), for example, 172/268 (64.2%) of clones were some allele of IgHV1, and 72.4% of clones were some allele of IgKV1. In general, scFv clone binders previously identified in Balb/c, SJL, and Medarex HuMAb mice[28,30] showed similarly high levels of germline %ID and similarly low levels of V-gene diversity. Though each immunization and panning method yielded distinct clones (Figure B4), there was a strong bias toward a limited variety of V-Joining (J) combinations, with between-method Pearson correlation coefficients ranging from 0.59 to 0.81 (Supplementary Figure S11), all of which were significant ($p < 0.001$). We postulate that such similarities arise due to the limited V-J diversity in the pre-sort repertoires

(Supplementary Figures S9–S10), though we cannot rule out the possibility that OX40 binding is more likely given particular V-J pairs.

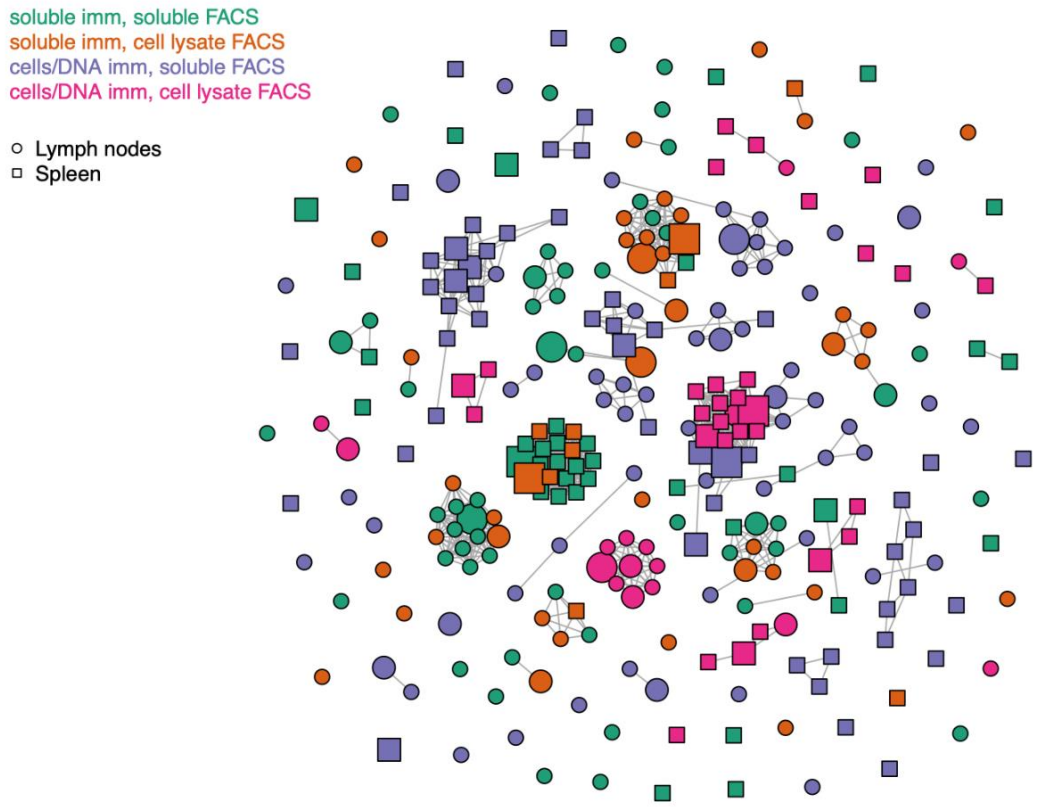


Figure B 4. Clonal cluster plot of anti-OX40 clones with frequencies higher than 0.1% in the post-sorted populations. scFv isolated from lymph nodes and spleen are indicated with circles and squares, respectively. scFv isolated from soluble immunization and sorted with either soluble OX40 or cell lysate expressing OX40 are colored in green and orange, respectively. scFv isolated from cell and DNA immunization and sorted with either soluble OX40 or cell lysate expressing OX40 are colored in purple and magenta, respectively. Antibody clones are represented by "nodes" in the plots. The size of the nodes corresponds to the frequency of the antibody clone in the FACS-sorted population: small (<2% frequency), medium (2–12% frequency), and large (>12% frequency). An "edge" (a line linking nodes) was drawn between any sequences with ≤9 amino acid differences in the concatenated scFvs. Done by GigaGen

B.4.5. Functional Characteristics of Monoclonal Antibody Binders

Next, to investigate the therapeutic potential of scFv binders, we synthesized forty-one putative binder scFv as full-length mAbs (**Table B1**), using methods described elsewhere[30]. We

chose at least the top two (range: 2–7 scFv) most common scFv enriched in each tissue, from each immunization and FACS method combination, along with several other strong and weak enriched scFv. First, we used a FACS assay to assess the ability of each mAb to bind OX40 recombinantly expressed on cell surfaces. In total, 39% (16/41) of the mAbs bound cell surface antigen, for a 61% false positive rate (**Supplementary Figure S5; Supplementary Table S10**). Notably, immunizing with soluble OX40 antigen followed by sorting with lysate from OX40-expressing cells yielded a 0% false positive rate, i.e., all mAbs identified with this method bound cell surface OX40. The other methods yielded significantly higher false positive rates (*z*-test for proportions, $p < 0.01$), ranging from 35% to 67%. We also observed two distinct positive binder peaks in the histograms for 68.8% of cell surface binding scFv (11/16), for unknown reasons.

Table B 1. Functional characteristics of 41 scFv binders converted into full-length IgG1 mAbs.

mAb ID	Enriched? Soluble immunized, soluble sorted	Enriched? Soluble immunized, lysate sorted	Enriched? Cells/DNA immunized, soluble sorted	Enriched? Cells/DNA immunized, lysate sorted	Binds cells?	Promega in vitro assay EC50 (ug/mL) (agonist)	KD (nM) [Octet, global fit]
tOX40.2	Yes	No	No	No	No	not tested	not tested
tOX40.4	Yes	Yes	No	No	Yes (2 peaks)	0.044	9
tOX40.15	Yes	No	No	No	Yes (1 peak)	1.276	6
tOX40.19	Yes	Yes	No	No	Yes (1 peak)	0.811	7.7
tOX40.20	Yes	No	No	No	No	not tested	not tested
tOX40.21	Yes	Yes	No	No	Yes (1 peak)	does not agonize	7
tOX40.22	Yes	Yes	No	No	Yes (2 peaks)	0.087	7.3
tOX40.23	Yes	Yes	No	No	Yes (2 peaks)	0.419	5.2
tOX40.24	Yes	Yes	No	No	Yes (2 peaks)	0.024	22.9
tOX40.28	Yes	No	No	No	No	not tested	not tested
tOX40.31	Yes	Yes	No	No	Yes (2 peaks)	0.043	22.4
tOX40.33	Yes	No	No	No	No	not tested	not tested

Table B1 (cont'd)

tOX40.34	Yes	No	No	No	No	not tested	not tested
tOX40.35	Yes	No	No	No	Yes (1 peak)	4.316	2.7
tOX40.36	Yes	No	No	No	No	not tested	not tested
tOX40.37	Yes	Yes	No	No	Yes (2 peaks)	0.195	12.5
tOX40.38	Yes	No	No	No	No	not tested	not tested
tOX40.39	Yes	Yes	No	No	Yes (2 peaks)	0.199	90
tOX40.40	Yes	Yes	No	No	Yes (2 peaks)	0.426	151
tOX40.41	No	Yes	No	No	Yes (2 peaks)	0.068	58.4
tOX40.42	No	No	Yes	No	Yes (1 peak)	does not agonize	no binding
tOX40.43	No	No	Yes	No	No	not tested	not tested
tOX40.44	No	No	Yes	No	No	not tested	not tested
tOX40.45	No	No	Yes	No	No	not tested	not tested
tOX40.46	No	No	Yes	No	No	not tested	not tested
tOX40.47	No	No	Yes	No	No	not tested	not tested
tOX40.48	No	No	Yes	No	No	not tested	not tested
tOX40.49	No	No	Yes	No	No	not tested	not tested
tOX40.50	No	No	Yes	Yes	Yes (2 peaks)	0.091	29.8
tOX40.51	No	No	No	Yes	No	not tested	not tested
tOX40.52	No	No	No	Yes	No	not tested	not tested
tOX40.54	No	No	No	Yes	Yes (2 peaks)	0.321	184
tOX40.55	No	No	No	Yes	No	not tested	not tested
tOX40.56	No	No	Yes	No	No	not tested	not tested
tOX40.57	No	No	Yes	No	No	not tested	not tested
tOX40.58	No	No	Yes	No	No	not tested	not tested
tOX40.59	No	No	Yes	No	No	not tested	not tested
tOX40.60	No	No	Yes	No	No	not tested	not tested
tOX40.61	No	No	No	Yes	No	not tested	not tested
tOX40.62	No	No	No	Yes	No	not tested	not tested
tOX40.63	No	No	No	Yes	No	not tested	not tested

We then tested the sixteen cell-surface binding mAbs for in vitro activation in a cellular assay. The average EC50 was 0.59 μg/μL, with a range from 0.024 to 4.3 μg/μL. Cell surface binding was a good predictor of in vitro agonism, with 87.5% (14/16) of cell surface binders demonstrating agonism (**Table B1; Supplementary Figure S7; Supplementary Table S10**), for

a 77% (27/41) false positive rate overall. Again, immunizing with soluble OX40 antigen followed by sorting with OX40-embedded cell lysate yielded the lowest false positive rate, at 9.1% (1/11). The other methods yielded significantly higher false positive rates (*z*-test for proportions, $p < 0.01$), ranging from 42.1% to 92.9%. The number of peaks in the cell surface binding assay was associated with the strength of agonism: the 1-peak mAbs have an average EC50 of 2.1 (with 2 mAbs not showing any agonist activity), whereas the 2-peak mAbs have an average EC50 of 0.17 (with all mAbs having agonist activity). Note that the positive control benchmark (pogalizumab) is a 2-peak binder with a strong EC50 (0.039 μg/μL). We speculate that mAbs in the 2-peak group comprise a different epitope bin than the mAbs in the 1-peak group. The two mAbs that failed to agonize (tOX40.21 and tOX40.42) also showed the weakest fluorescence shift in the flow cytometry cell surface binding experiments (**Supplementary Figure S5**). We did not observe significant differences among the protocols in in vitro agonism EC50 ($p > 0.01$, *t*-test).

We also tested the cell surface binding mAbs for affinity using Octet. Of the mAbs that bound cell surface antigen, 93.8% (15/16) also bound soluble antigen (**Table B1; Supplementary Figure S6; Supplementary Table S10**). The average K_D was 41.1 nM, with a range from 2.7 to 184 nM. The pogalizumab positive control yielded a K_D of 1.9 nM. One of the antibodies that bound cell surface antigen (tOX40.42) failed to bind antigen by Octet and also failed in vitro agonism. This mAb was discovered using the cells/DNA immunization and soluble antigen sorting method. Another mAb (tOX40.21) did not agonize OX40 in vitro but did bind soluble antigen (K_D = 7nM). This antibody was among the weakest binders in the cell surface flow cytometry assay and was discovered using the soluble antigen immunization with both the soluble sort and cell lysate sort methods. We speculate that this mAb binds non-specifically, resulting in high affinity

but weak agonism and cell surface binding. We did not observe significant differences among the protocols in K_D ($p > 0.01$, *t*-test).

B.5. Discussion

In this study, we adapted previously published methods[28–30] to test whether different immunization methods (cells/DNA versus soluble antigen) and different selection methods (cell lysate versus soluble antigen) yielded mAbs with higher potential as therapeutic OX40 agonists. Though all methods successfully identified anti-OX40 mAbs, using cell lysate for selection generally yielded mAbs that were more likely to bind to cell surface antigen and activate OX40 in cellular assays. We speculate that cell lysate contained OX40 trimers, whereas soluble antigen comprised OX40 monomers, perhaps leading to the identification of more physiologically relevant binders. Using massively parallel microfluidics and deep sequencing allowed us to rigorously characterize mouse responses to different types of immunogens. The large scFv repertoires generated from the animals also facilitated robust testing of FACS methods. Other methods, such as hybridomas, would have required significantly more effort to generate such a comprehensive data set.

In this study, immunization with cells/DNA was inconsistent and yielded a low proportion of agonist mAbs. In future experiments, we could establish a titer cutoff and only make yeast libraries from animals with titers exceeding that cutoff. Still, there are many ways that cells/DNA immunization could be optimized in the future. For example, we could test different concentrations of cells in the mouse immunizations, different adjuvants, different DNA vectors, or alternate cell lines. We could also ensure high levels of cell surface antigen expression by using FACS to isolate populations of cells with the highest antigen expression, as described elsewhere[117]. A more aggressive cells/DNA immunization schedule could increase titer and reproducibility, for example,

through daily injections of cells for the first few days of the immunization protocol[117], biweekly immunizations with cells for ten weeks[118], or biweekly immunizations with DNA for eight weeks[119].

Though the cell lysate sorting method yielded the highest proportion of agonist mAbs in this study, there are many opportunities for further improvement. We tested several different cell lysate concentrations, but a more rigorous optimization would require a more thorough analysis of the impact of lysate concentration on FACS sensitivity and specificity. We only tested the cell lysate sorting methods with OX40, whereas other targets may yield different results. For example, certain targets may unfold in our lysis buffer, yielding antibodies less likely to bind to a properly folded protein target. Additionally, we might find that peptide tags other than FLAG (for example, His tag) might yield better results, or that C-terminal tags are preferential for certain targets. Finally, further work might compare screening yeast scFv libraries with cell lysate versus screening phage scFv display libraries against cells affixed to plates.

To our knowledge, no other group has published in-depth studies of the antibody repertoire response of Trianni humanized mice to immunogens. Our work yielded a low diversity of light chain V-genes, for example, >70% of scFv binder clones were IgKV1. This level of light chain Ig diversity after immunization and FACS selection is similar to results obtained in wild type Balb/c and SJL mice[28] and humanized Medarex HuMAb mice[30]. Additionally, the V sequences of scFv binders from both libraries were ~98% identical to germline V sequences, suggesting little if any affinity maturation in vivo. Prior work on repertoires of mice administered various immunogens found only 2-5 amino acid substitutions per V-gene[28,30,120–122]. Future work should investigate whether Trianni mice generate similar responses with other immunogens.

Our methods open up exciting directions for mAb discovery and development. For example, we could use cell lysates to select for mAbs that bind to a specific epitope, or do not bind to a specific epitope. In one scenario, cells could be engineered that express OX40 protein with mutations in the amino acids required for binding OX40L. Then, scFv libraries could be sorted using the mutated OX40, perhaps identifying antibodies that bind outside the OX40:OX40L binding domain. Another intriguing approach would be to immunize mice with tumor cells, and then pan for scFv that bind to lysates from tumors but not to lysates from normal tissue. This approach could be used to find mAbs directed against novel tumor-specific targets.

B.6. Patents

The OX40 full-length antibody sequences and the OX40-enriched yeast scFv libraries described in this article are patent-pending subject matter in USPTO provisional patent application number 62/788687, priority date 4 January 2019.

APPENDIX C:

Supplementary Notes

Note C 1. Sequence for IL-31 construct

C1.1. Feline IL-31 DNA Sequence

ATGTCTCACATGGCTCCAGCACATAGATTACAACCATCAGATATTAGAAAGATCATC
TTGGAATTAAGACCAATGTCTAAAGGTTTGTTGCAAGATTATTTGAAGAAAGAAATC
GGTTTACCAGAATCAAACCATTCTTCATTGCCATGTTTATCTTCAGATTCTCAATTGC
CACATATCAACGGTTCAGCAATCTTGCCATACTTTAGAGCTATTAGACCATTGTCAG
ATAAGAACACAATCGATAAGATCATCGAACAACTAGACAAGTTGAAGTTTCAAAGA
GAACCAGAAGCAAAAGTTTCTATGCCAGCTGATAACTTCGAAAGAAAGAATTTCAT
CTTGGCAGTTTTACAACAATTTTCAGCTTGTTTGGAACATGTTTTGCAATCTTTGAAT
TCAGGTCCACAA

Note C 2. Sequences for Pro-NGF constructs

C2.1. cNGF DNA Sequence

ATGTCTTCATCTCATCCAGTTTTTCATAGAGGTGAATTTTCTGTTTGTGATTCAGTTTC
TGTTTGGGTTGGTGACAAGACTACAGCTACAGATATCAAGGGTAAAGAAGTTATGG
TTTTGGGTGAAGTTAACATCAACAACTCAGTTTTCAAGCAGTATTTCTTTGAAACAA
AATGTAGAGATCCAACTCCAGTTGATTCTGGTTGTAGAGGTATCGATTCAAAGCATT
GGAACTCTTACTGTACTACAACTCATACATTCGTTAAGGCATTGACTATGGATGGTA
AACAAGCTGCATGGAGATTCATTAGAATTGATACTGCTTGTGTTTGTGTTTTATCTAG
AAAAGCAGGTAGAAGAGCT

C2.2. Pro-cNGF DNA Sequence

GAACCACATCCAGAATCTCATGTTCCAGCAGGTCATGCTATTCCACATGCTCATTGG
ACAAAGTTGCAACATTCATTGGATACTGCATTGAGAAGAGCTAGATCTGCTCCAGCA
GGTGCTATTGCTGCAAGAGTTACAGGTCAAACTAGAAACATCACAGTTGATCCAAA
GTTGTTTAAGAAAAGAAGATTGAGATCACCAAGAGTTTTATTTTCTACTCATCCACC
ACCAGTTGCTGCAGATGCACAAGATTTGGATTTGGAAGCAGGTTCAACAGCTTCTGT
TAACAGAACTCATAGATCAAAGAGATCTTCATCTCATCCAGTTTTTCATAGAGGTGA
ATTTTCTGTTTGTGATTCAGTTTCTGTTTGGGTTGGTGACAAGACTACAGCTACAGAT
ATCAAGGGTAAAGAAGTTATGGTTTTGGGTGAAGTTAACATCAACAACTCAGTTTTC
AAACAATACTTTTTTGAAACAAAATGTAGAGATCCAACTCCAGTTGATTCTGGTTGT
AGAGGTATCGATTCAAAGCATTGGAACTCTTACTGTACTACAACTCATACATTCGTT
AAGGCATTGACTATGGATGGTAAACAAGCTGCATGGAGATTCATTAGAATTGATACT
GCTTGTGTTTGTGTTTTATCTAGAAAAGCAGGTAGAAGAGCT

C2.3. ProΔ1,2-cNGF DNA Sequence

CAAACTAGAAACATCACAGTTGATCCAAAGTTGTTTAAGAAAAGAAGATTGAGATC
ACCAAGAGTTTTATTTTCTACTCATCCACCACCAGTTGCTGCAGATGCACAAGATTT
GGATTTGGAAGCAGGTTCAACAGCTTCTGTTAACAGAACTCATAGATCAAAGAGAT
CTTCATCTCATCCAGTTTTTCATAGAGGTGAATTTTCTGTTTGTGATTCAGTTTCTGTT
TGGGTTGGTGACAAGACTACAGCTACAGATATCAAGGGTAAAGAAGTTATGGTTTT
GGGTGAAGTTAACATCAACAACTCAGTTTTCAAACAATACTTTTTTGAAACAAAATG
TAGAGATCCAACTCCAGTTGATTCTGGTTGTAGAGGTATCGATTCAAAGCATTGGAA
CTCTTACTGTACTACAACTCATACATTCGTTAAGGCATTGACTATGGATGGTAAACA
AGCTGCATGGAGATTCATTAGAATTGATACTGCTTGTGTTTGTGTTTTATCTAGAAAA
GCAGGTAGAAGAGCT

Note C 3. Sequence for AmiE, UCA9, and PYR1 constructs

C3.1. AmiE DNA Sequence

ATGAGACATGGCGATATTAGCTCGTCAAATGATACCGTAGGCGTAGCCGTGGTGAA
TTACAAGATGCCGCGTTTACATACTGCTGCTGAAGTCCTGGATAATGCCCGCAAAAT
TGCGGAAATGATCGTTGGTATGAAGCAAGGTCTGCCGGGCATGGATCTGGTTGTGTT
TCCTGAATATTCTTTACAGGGTATTATGTACGACCCTGCTGAAATGATGGAAACAGC
CGTGGCGATTCCAGGCGAAGAAACGGAAATCTTTAGCCGTGCTTGTAGAAAAGCAA
ATGTTTGGGGTGTGTTCTCCCTGACCGGCGAACGTCATGAAGAACACCCTAGAAAGG
CACCATACAACACTCTGGTCTTGATCGATAACAACGGTGAAATCGTACAAAAGTAC
AGAAAGATCATCCCATGGTGTCCGATTGAAGGCTGGTATCCAGGTGGCCAGACATA
CGTCTCTGAAGGTCCGAAAGGCATGAAGATCTCATTAATTATCTGCGATGACGGTAA
TTATCCGGAAATTTGGAGAGATTGTGCCATGAAGGGTGCGGAATTGATCGTTCGCTG
CCAAGGCTATATGTACCCTGCTAAAGACCAACAAGTTATGATGGCTAAGGCAATGG
CCTGGGCGAATAACTGTTATGTCGCTGTAGCAAACGCTGCAGGTTTTGATGGCGTTT
ATAGCTACTTCGGTCATAGTGCCATTATCGGTTTTGACGGCCGTACTCTGGGTGAAT
GCGGCGAAGAAGAAATGGGCATTCAATACGCGCAGTTGTCTCTGTCACAAATCCGC
GATGCCCGTGCGAATGACCAAAGTCAGAACCATTTGTTTAAAATCTTGCACAGAGGT
TACTCCGGTTTGCAGGCTTCGGGCGATGGCGACCGTGGTCTGGCAGAATGCCCATTT
GAATTCTACCGTACCTGGGTTACTGATGCTGAAAAGGCAAGAGAAAACGTGGAACG
CCTGACTCGCTCCACAACAGGTGTCGCCCAATGCCCAGTAGGTCGTCTGCCGTATGA
AGGCCTCGAG

C3.2. UCA9 DNA Sequence

ATGCAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCTCGGT
GAAGGTCTCCTGCAAGGCTTCTGGAGGCACCTTCAGCAGCTATGCTATCAGCTGGGT
GCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAGGGATCATCCCTATCTTTG
GTACAGCAAACTACGCACAGAAGTTCCAGGGCAGAGTCACGATTACCGCGGACGAA
TCCACGAGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGACACGGCCGT
GTATTACTGTGCGAAAGGTAGTGGTTATCATGTCCGCGATTACTTTGACTACTGGGG
CCAAGGAACCCTGGTCACCGTCTCCTCA

C3.3. PYR1 DNA Sequence

ATGCCTTCGGAGTTAACACCAGAAGAACGATCGGAACTAAAAAACTCAATCGCCGA
GTTCCACACATACCAACTCGATCCAGGAAGCTGTTCATCACTCCACGCGCAACGAAT
CCACGCGCCTCCGGAACTCGTCTGGTCAATCGTACGACGATTCGACAAACCACAAA
CATACAAACACTTCATCAAATCCTGCTCCGTCGAACAAAACTTCGAGATGCGCGTCG
GATGCACGCGCGACGTGATCGTCATCAGTGGATTACCGGCGAACACATCAACGGAA
AGACTCGATATACTCGACGACGAACGGAGAGTTACCGGATTCAGTATCATCGGAGG
CGAACATAGGCTGACGAATTACAAATCCGTTACGACGGTGCATCGGTTCGAGAAAG
AGAATCGGATCTGGACGGTGGTTTTGGAATCTTACGTCGTTGATATGCCGGAAGGTA
ACTCGGAGGATGATACTCGTATGTTTGCTGATACGGTTGTGAAGCTTAATTTGCAGA

AACTCGCGACGGTTGCTGAAGCTATGGCTCGTAACTCCGGTGACGGAAGTGGTTCTC
AGGTGACG

Note C 4. A protocol to determine the concentration of pooled mutagenic oligos for gene X used in the Nicking Saturation Mutagenesis (NSM) experiments

Overview: These steps will allow the user to achieve a 20:1 molar ratio of dsDNA template: pooled oligo compatible with NSM experiments. We **assume** that the end user has a lyophilized oligo pool that has been resuspended to a total oligonucleotide concentration of 200 nM ($0.2\mu M$), with a certain number of oligos in the pool that are specific to gene *X*. The steps are listed below:

a. Calculate the molar fraction of mutagenic oligos for gene *X* in the oligo pool
b. Calculate the effective concentration of mutagenic oligos for gene *X*
c. Calculate the concentration of the phosphorylated oligos using 20 µL of the lyophilized oligo pool (step 1 from NSM, https://www.nature.com/protocolexchange/protocols/5125)
d. Calculate the moles of mutagenic oligos needed to achieve a 20:1 molar ratio of dsDNA template: pooled oligo
e. Calculate the dilution of phosphorylated oligos needed after NSM step 1

Example:

Assuming i) a total of 0.2µM pooled oligos and ii) 1,449 out of 7,188 pooled oligos are specific for gene *X*.

a. Molar fraction of protein *X* specific oligos:

$$\frac{1{,}449}{7{,}118} = 0.204 \frac{gene\ X - specific\ oligos}{total\ oligos}$$

b. Effective concentration of mutagenic oligos:

$$(0.2\mu M\ total\ oligos) * \left(0.204\ \frac{gene\ X - specific\ oligos}{total\ oligos}\right) = 0.0408\ \mu M\ gene\ X - specific\ oligos$$

c. Concentration of phosphorylated oligos (see step 1 from NSM):

$$\frac{(0.0408\mu M) * (20\ \mu L\ oligo\ pool\ volume)}{24.4\ \mu L\ total\ reaction\ volume} = 0.0334\ \mu M$$

d. Moles of oligos to achieve a 20:1 molar ratio:

Based on NSM experiments, we need 0.76 pmoles of dsDNA plasmid to prepare the ssDNA template strand (step 2). For a total reaction volume of 20 µL,

$$\frac{0.76\ pmols}{2.0E^{-}05\ L} = 3.8E04\ pM = 0.038\ \mu M$$

$$\frac{20}{1}\ molar\ ratio = \frac{0.038\ \mu M}{X}$$

$$X = 0.0019\ \mu M$$

e. Dilution of phosphorylated oligos:

$$\frac{0.0334\ \mu M}{0.0019\ \mu M} = 17.5$$

So here, we would add

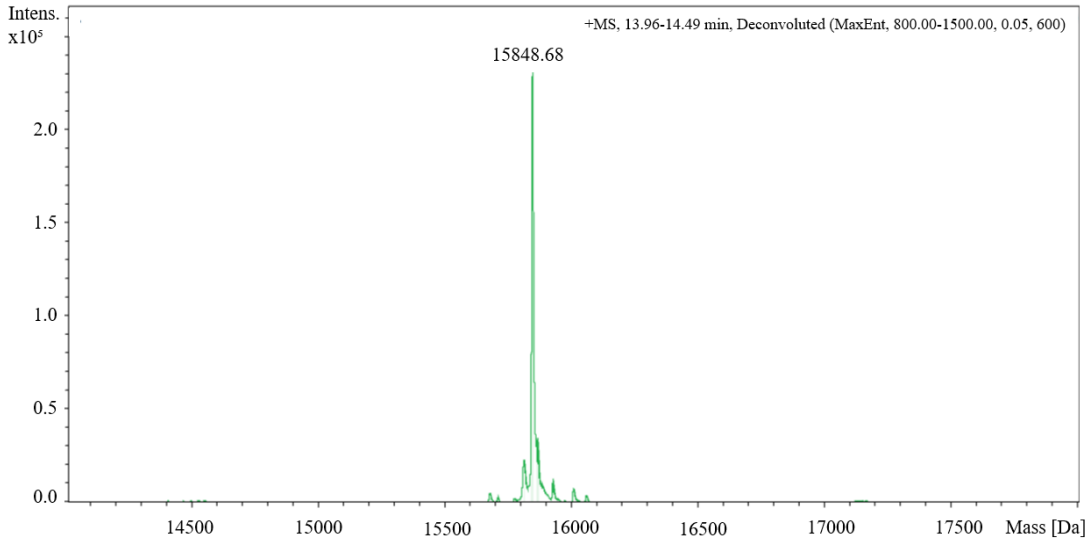
$$\frac{20\ \mu L}{17.5} = 1.14\ \mu L$$

of the phosphorylated oligos to the 20 µL nicking mutagenesis protocol.

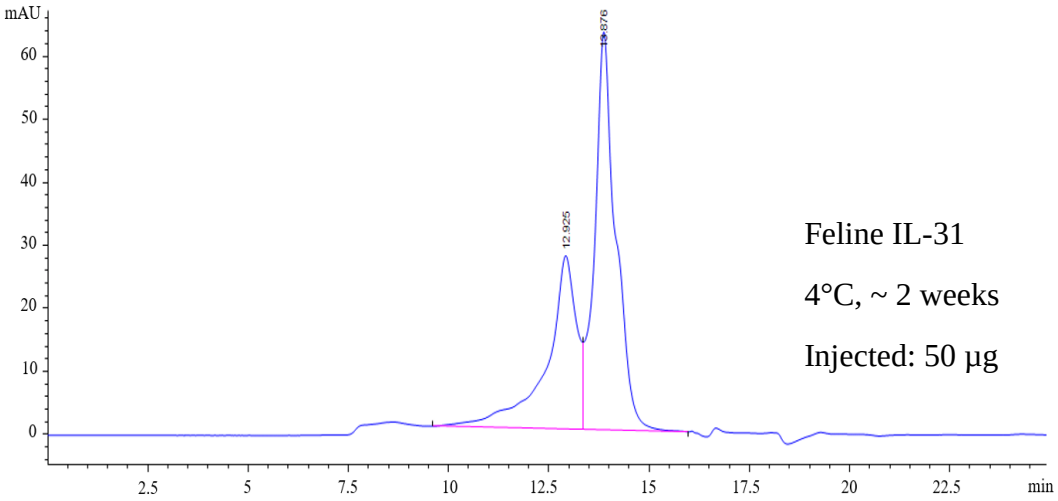
APPENDIX D:

Supplementary Figures

a.



b.



#	Time	Area	Height	Width	Area %	Symmetry
1	12.925	1458.8	27.5	0.7101	38.590	1.991
2	13.876	2321.4	63.2	0.5064	61.410	0.724

Figure D 1. Mass spectrometry and analytical SEC for feline IL-31 construct. (**a.**) Mass spec analysis of IL-31 after PNGase treatment. Conditions: 50mM Tris, pH 8.0. (**b.**) Chromatographic conditions: TSK SuperSW3000, 4.6 x 30 mm, mobile phase 200 mM sodium phosphate, pH 7.2, flow rate 0.25 mL/min for 25 minutes, injection amount: 50 µg. Done by Zoetis.

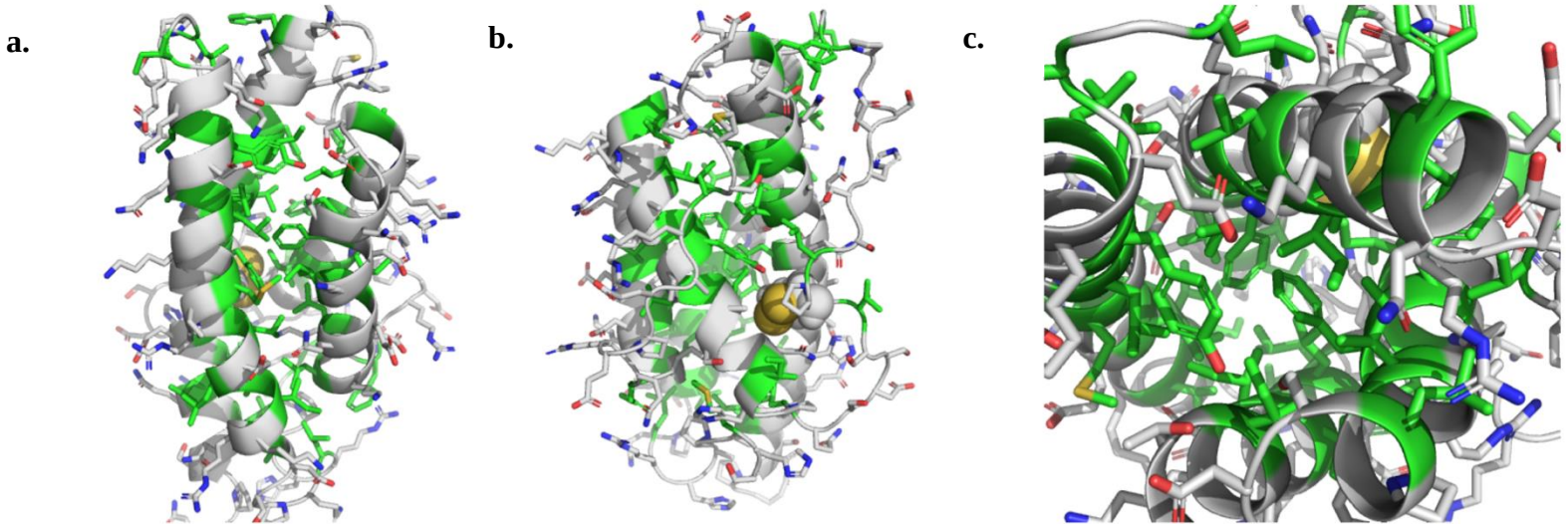


Figure D 2. Structural model of fIL-31. (**a.-b.**) Two different views showing packing of hydrophobic core. Aliphatic residues Phe, Tyr, Leu, Ile, Val, and Met are colored in green sticks, and the remaining residues are shown as white sticks. The Cys49-Cys132 disulfide bond is shown as spheres. (**c.**) A closer view of the hydrophobic core of the structural model.

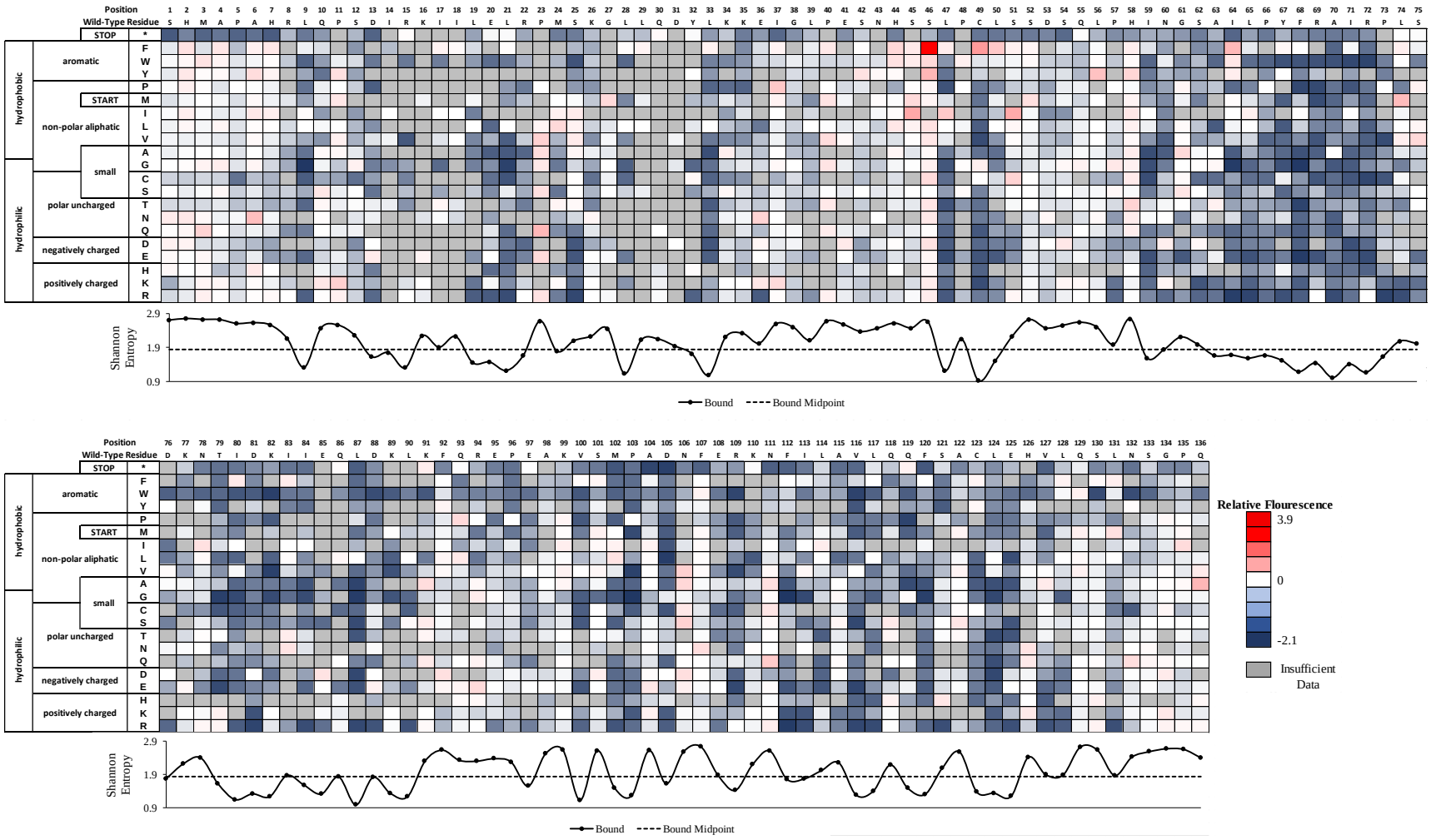


Figure D 3. Determination of the binding sites for fIL-31/fIL31RA-1FNIII interaction. Fitness Metric heatmap of the top 7% bound population vs the reference population. The Shannon entropy is plotted below with its respective cut-off midpoint (dashed line).

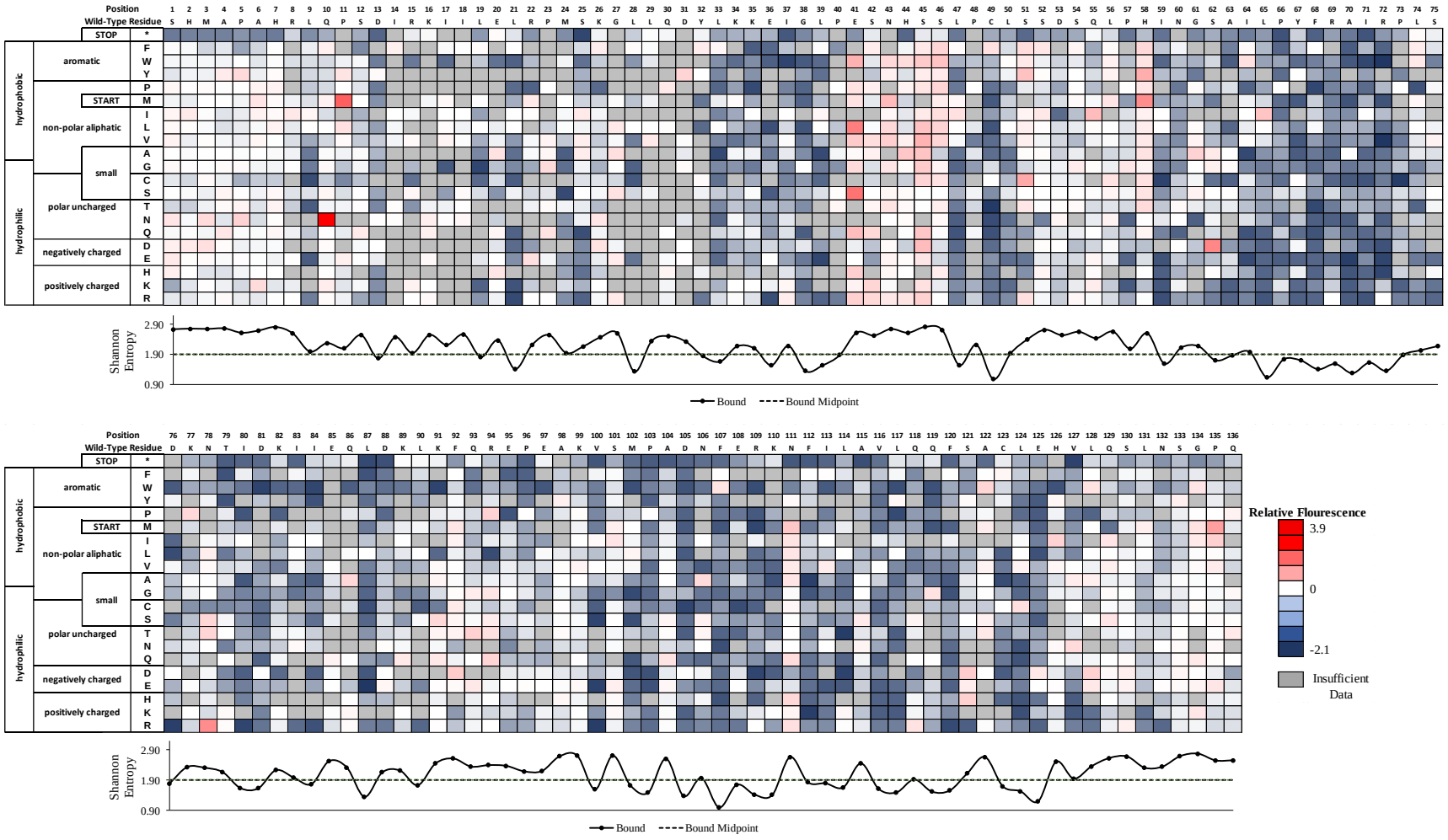


Figure D 4. Determination of the binding sites for fIL-31-fOSMR-ECD interaction. Fitness Metric heatmap of the top 7% bound population vs the reference population. The Shannon entropy is plotted below with its respective cut-off midpoint (dashed line).

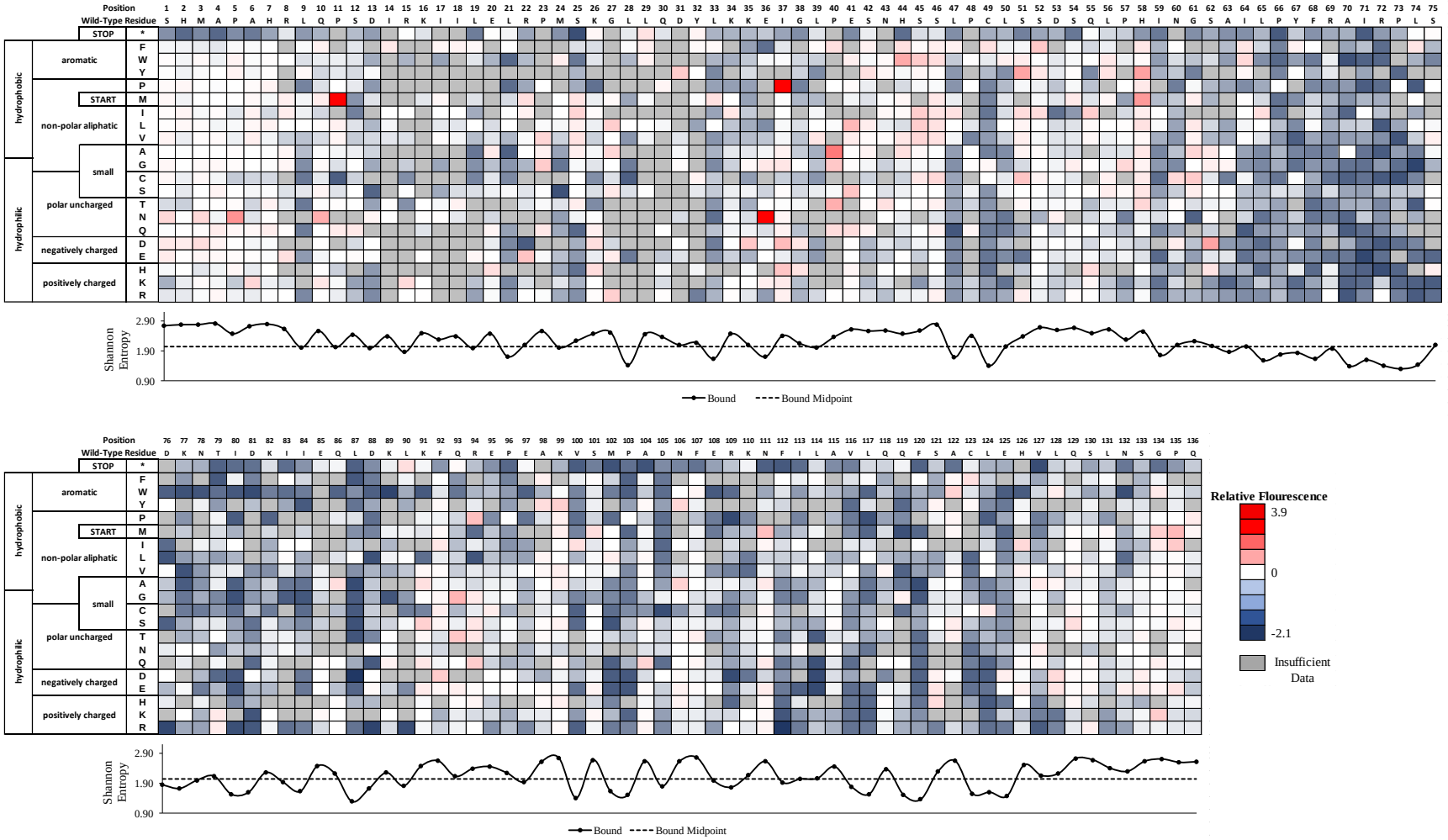


Figure D 5. Determination of the conformational epitope for fIL-31/mAb#1 interaction. Fitness Metric heatmap of the top 7% bound population vs the reference population. The Shannon entropy is plotted below with its respective cut-off midpoint (dashed line).

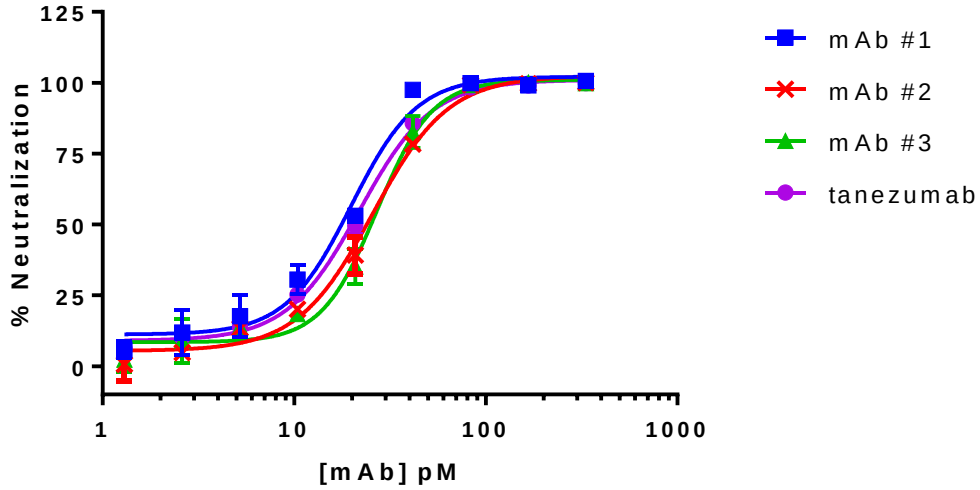


Figure D 6. Effects of anti-NGF mAbs on canine β-NGF Induced Proliferation of TF-1 Cells (representative curves). Done by Zoetis. (error bars, standard deviation, n=2)

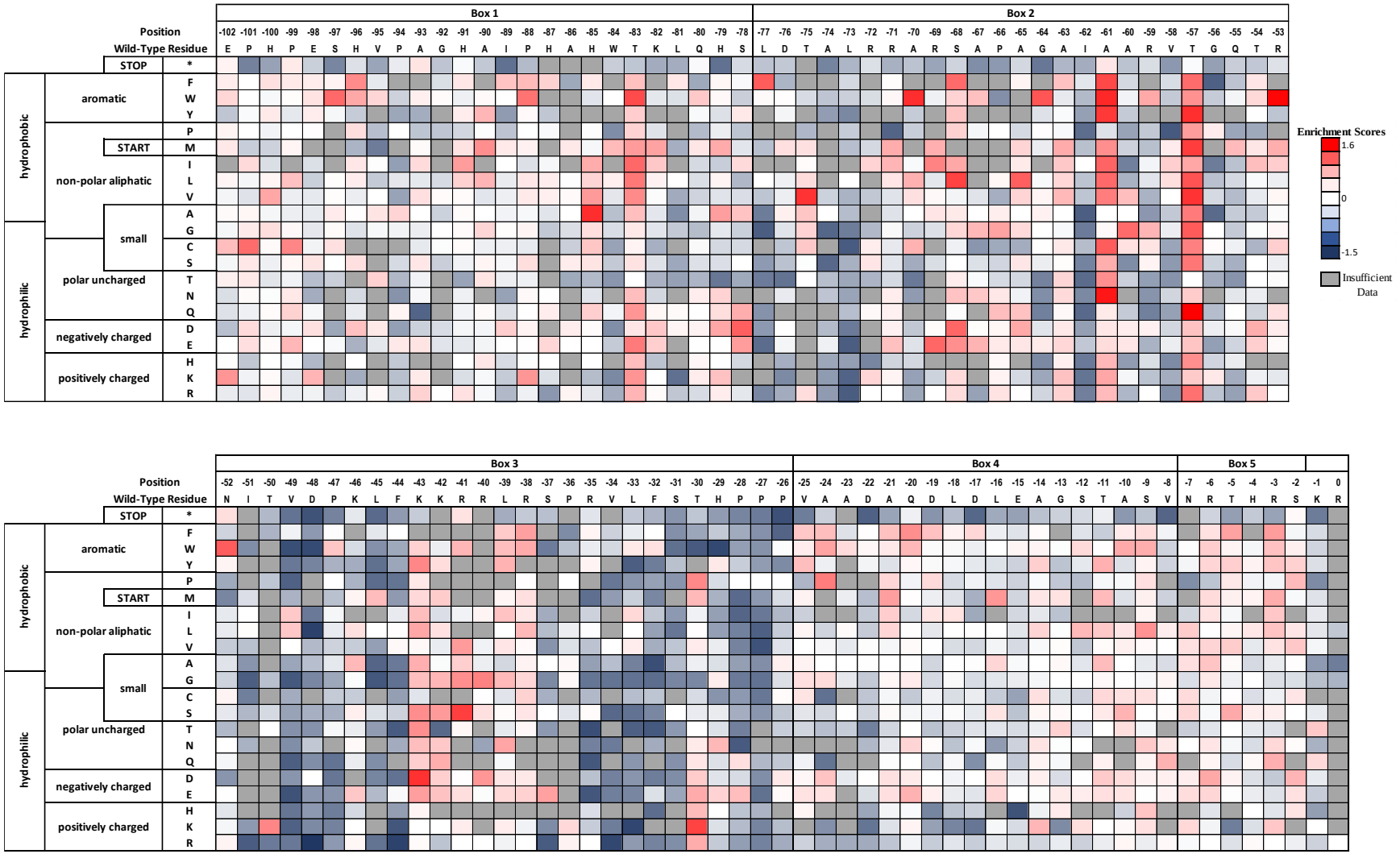


Figure D 7. Per-position heatmap of enrichment scores for pro-cNGF mutants after 1 sort with tanezumab.

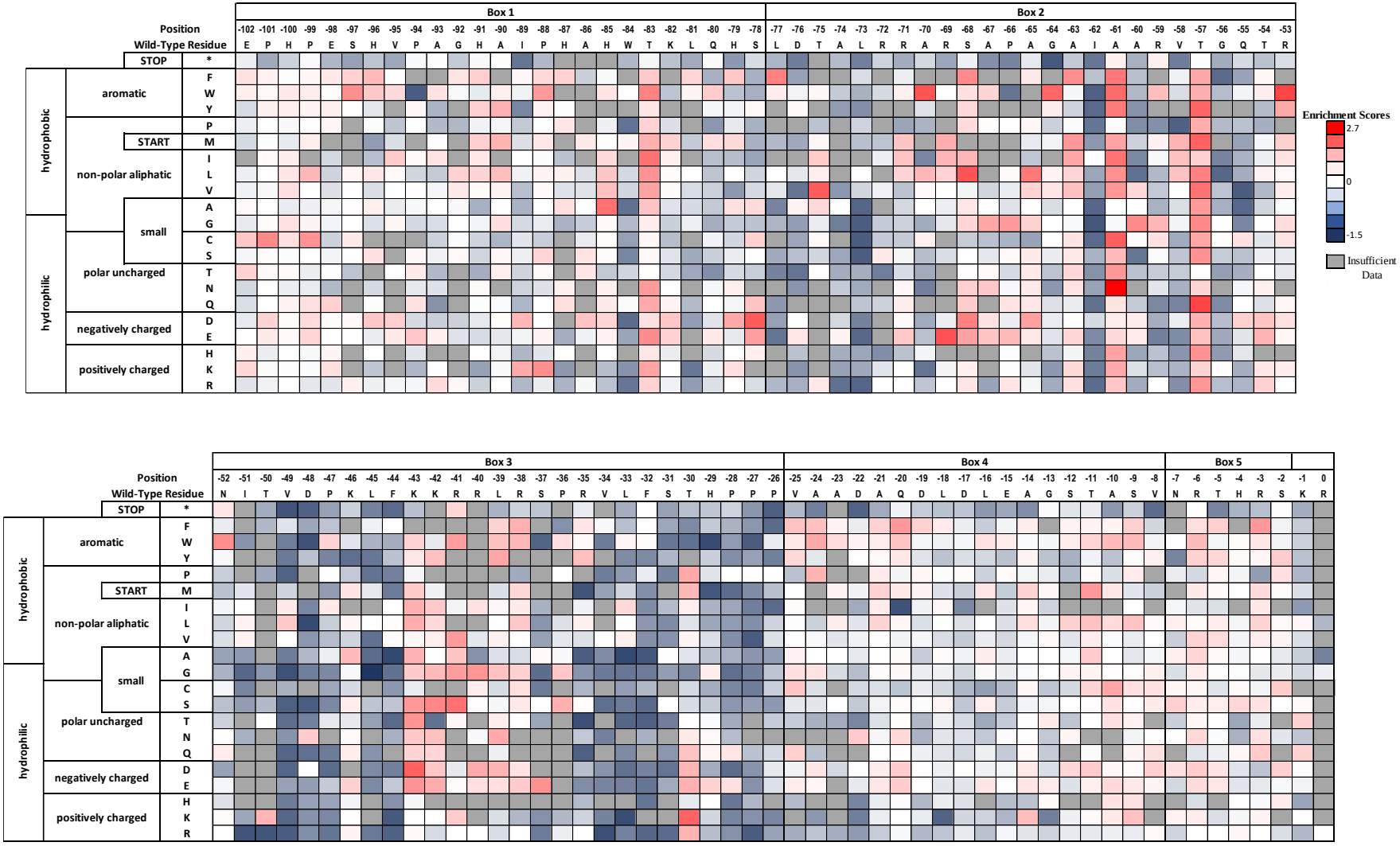


Figure D 8. Per-position heatmap of enrichment scores for pro-cNGF mutants after 1 sort with mAb #1.

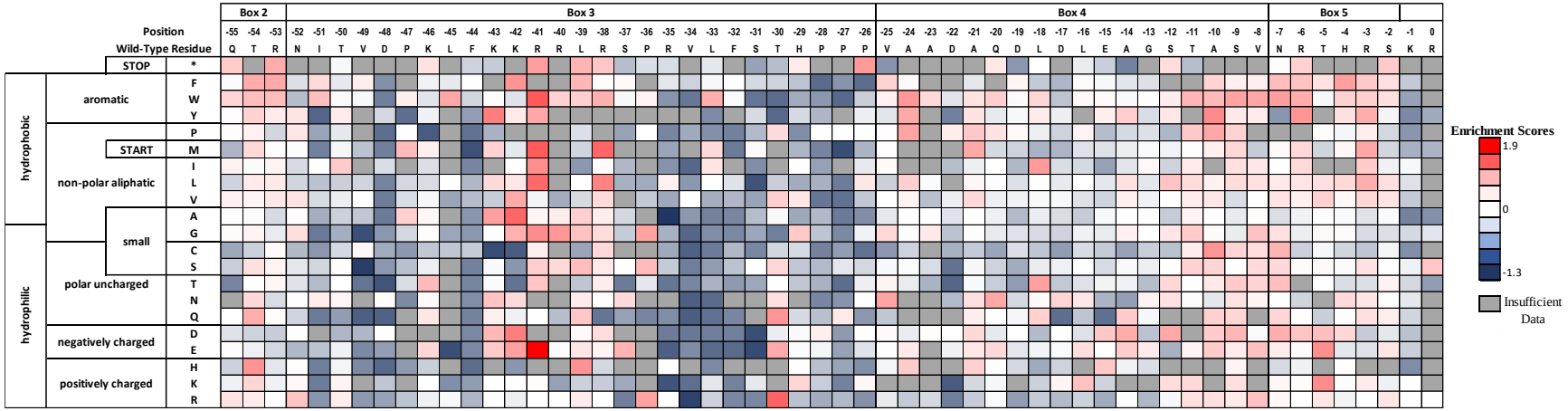


Figure D 9. Per-position heatmap of enrichment scores for proΔ1,2-cNGF mutants after 1 sort with tanezumab.

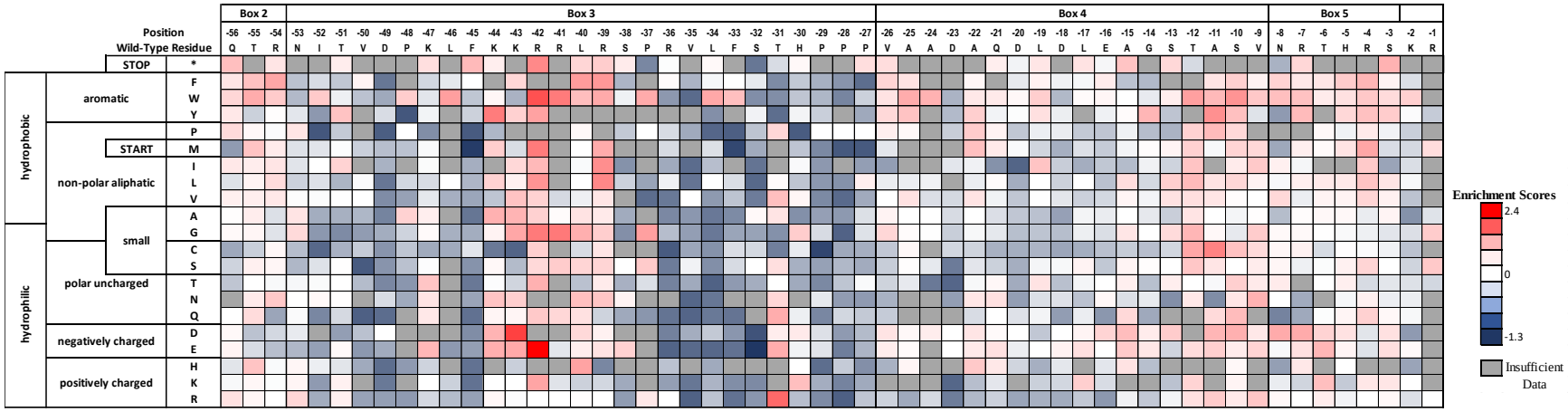


Figure D 10. Per-position heatmap of enrichment scores for proΔ1,2-cNGF mutants after 1 sort with mAb #1.

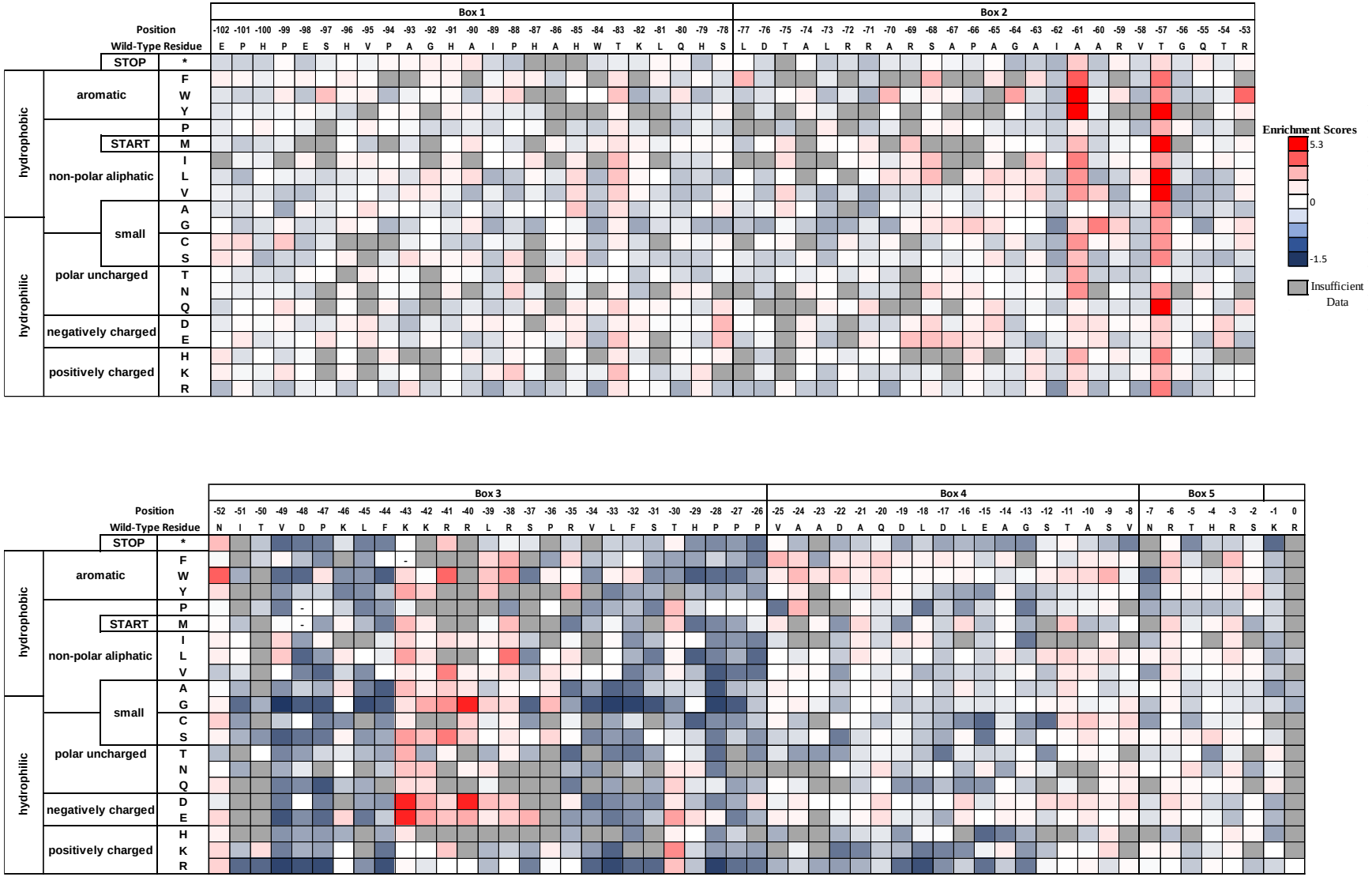


Figure D 11. Per-position heatmap of enrichment scores for pro-cNGF mutants after 2 sorts with mAb #1.

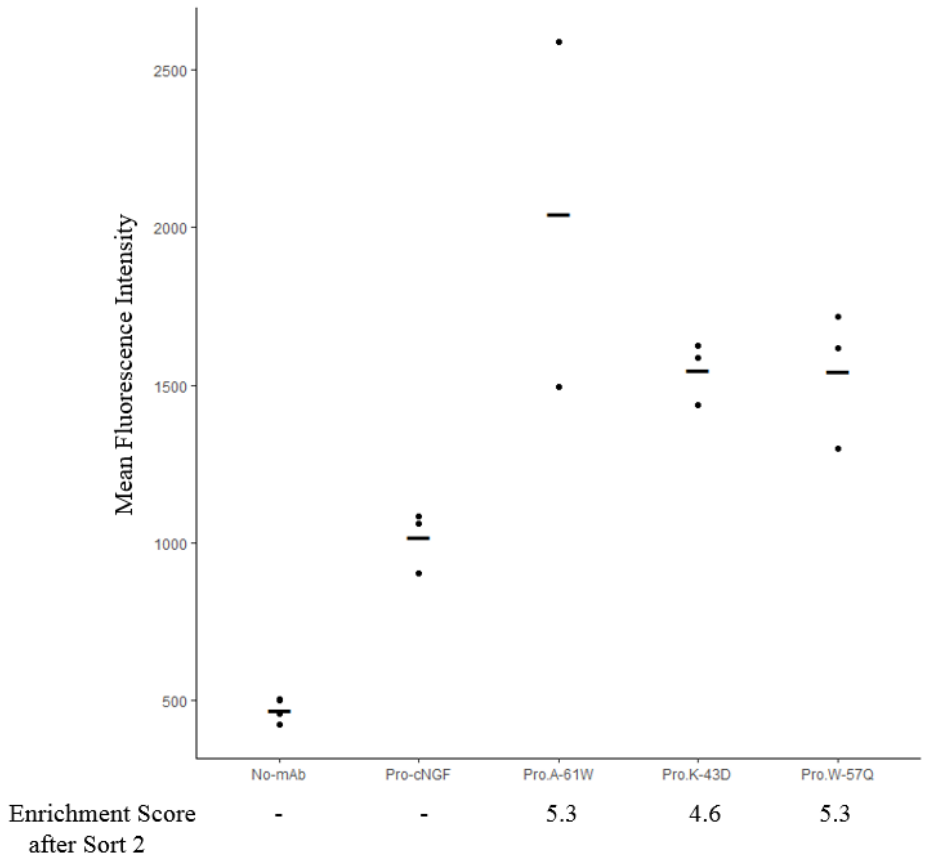


Figure D 12. Mean fluorescence intensities for individual point mutants compared with wild-type pro-cNGF.

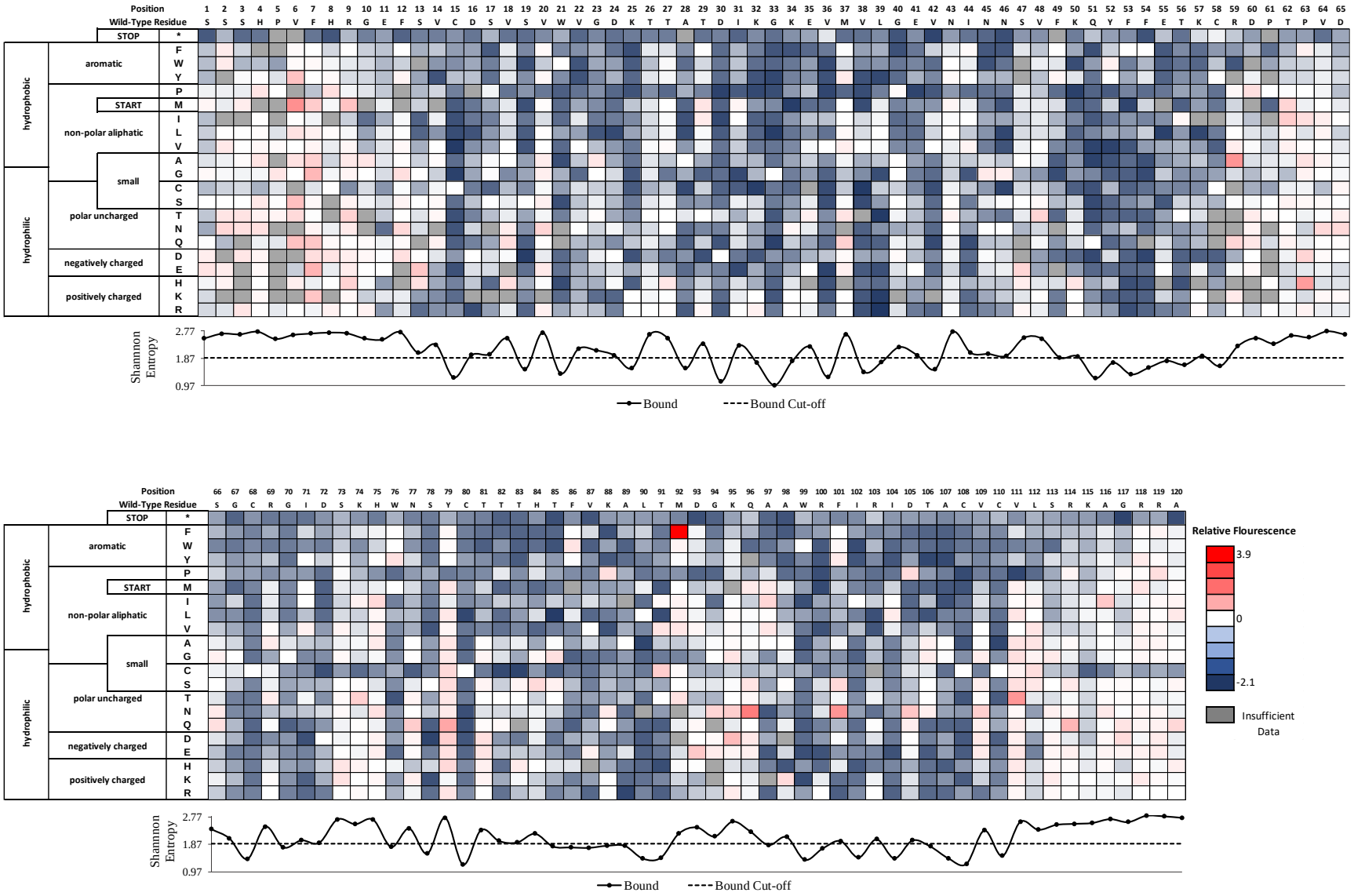


Figure D 13. Determination of conformational epitope for cNGF_tanezumab. Fitness metric heatmap of the top 7% bound population vs the unselected population. Shannon entropy is plotted below with its respectively cut-off (dashed line).

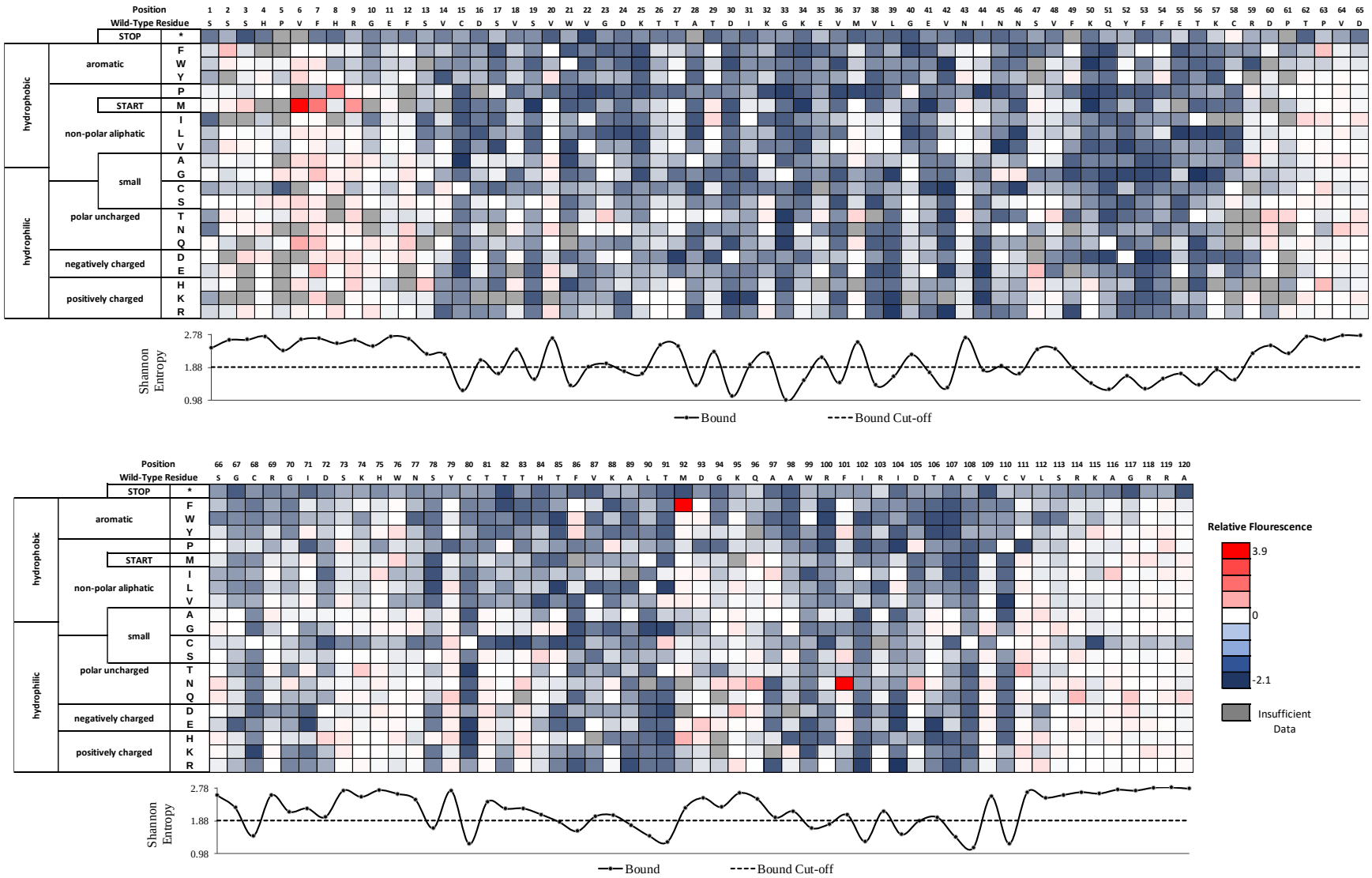


Figure D 14. Determination of conformational epitope for cNGF_mAb #1. Fitness metric heatmap of the top 7% bound population vs the unselected population. Shannon entropy is plotted below with its respectively cut-off (dashed line).

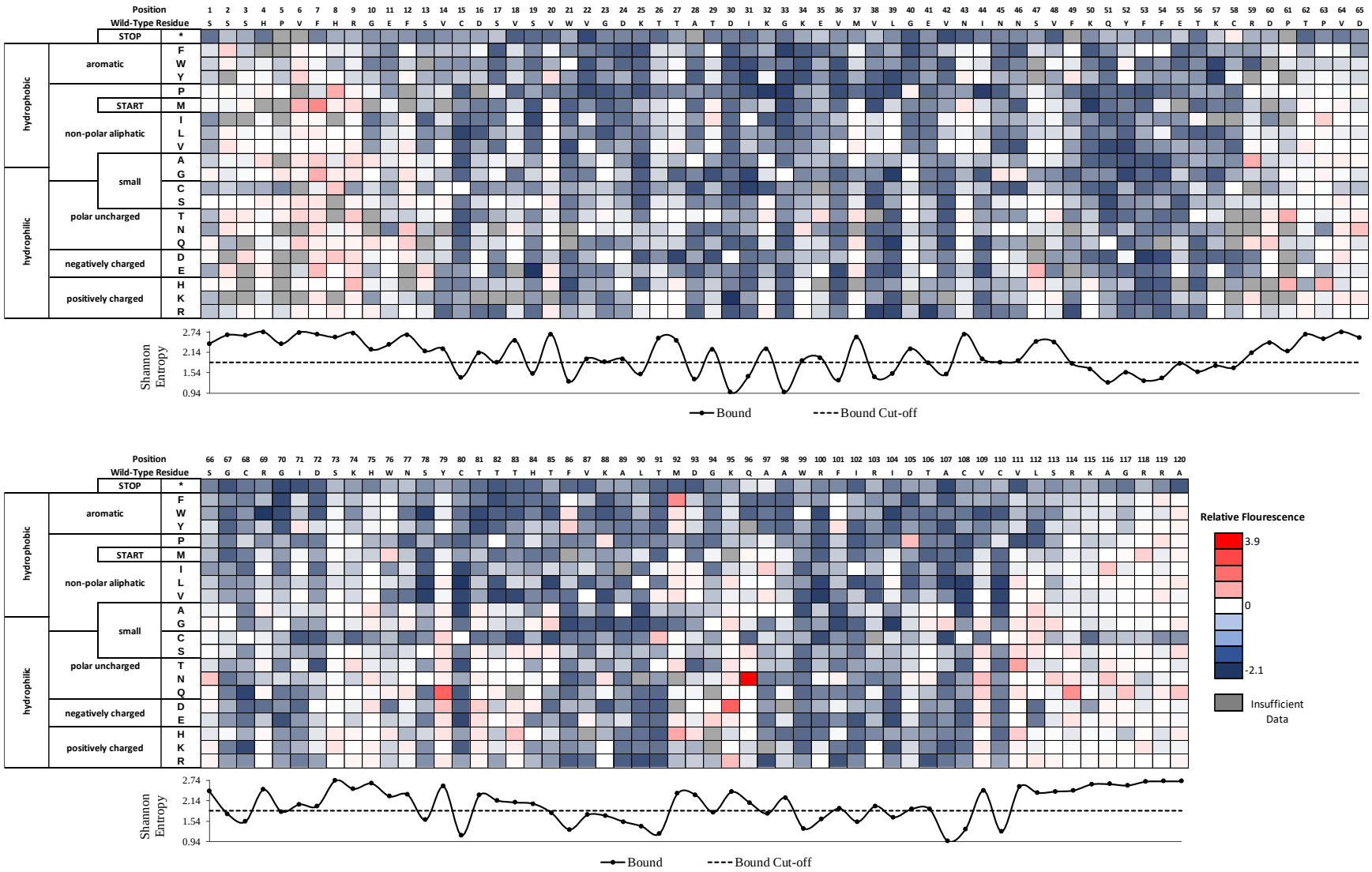


Figure D 15. Determination of conformational epitope for cNGF_mAb #2. Fitness metric heatmap of the top 7% bound population vs the unselected population. Shannon entropy is plotted below with its respectively cut-off (dashed line).

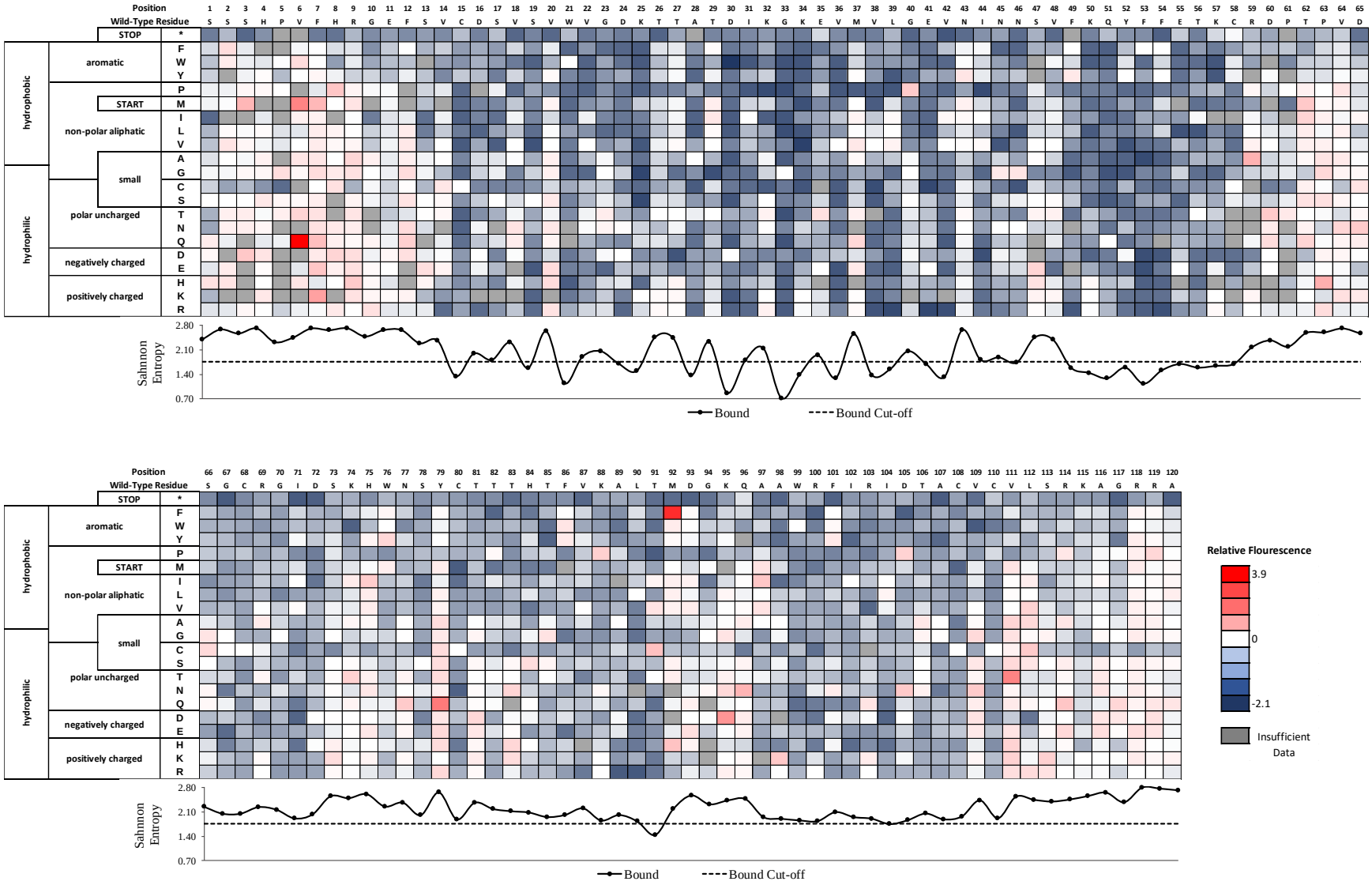


Figure D 16. Determination of conformational epitope for cNGF_mAb #3. Fitness metric heatmap of the top 7% bound population vs the unselected population. Shannon entropy is plotted below with its respectively cut-off (dashed line).

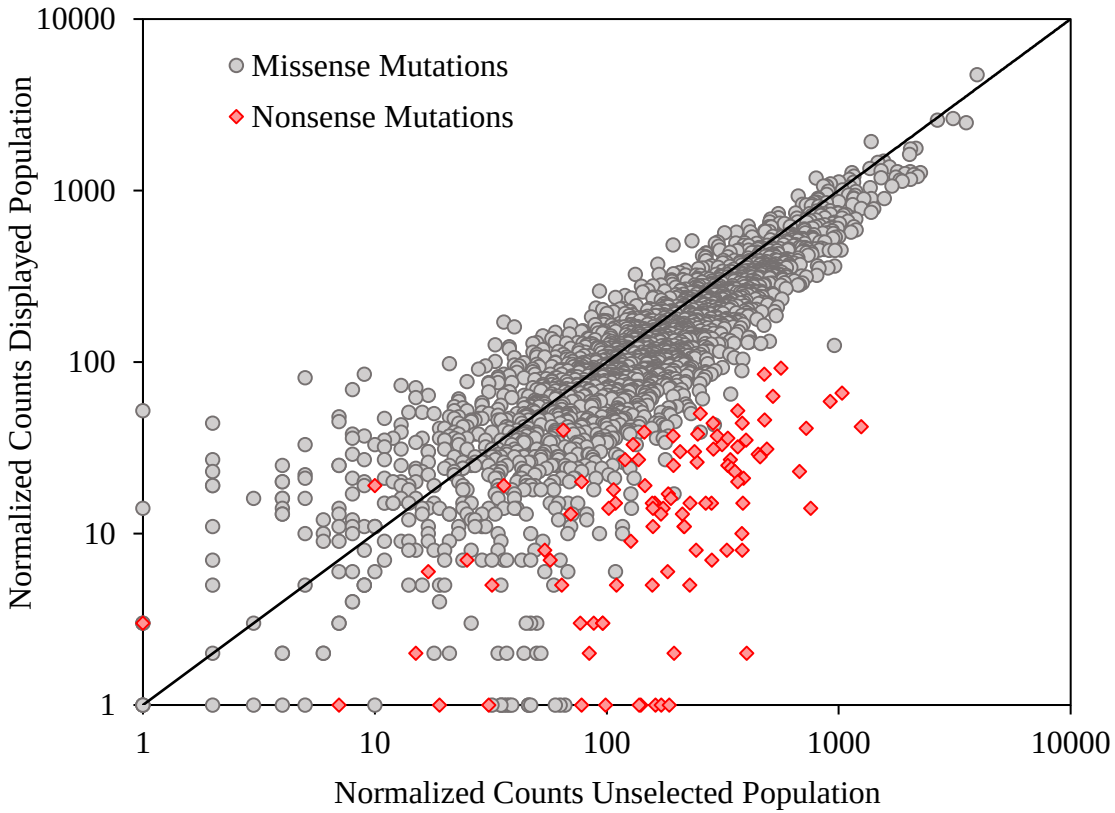
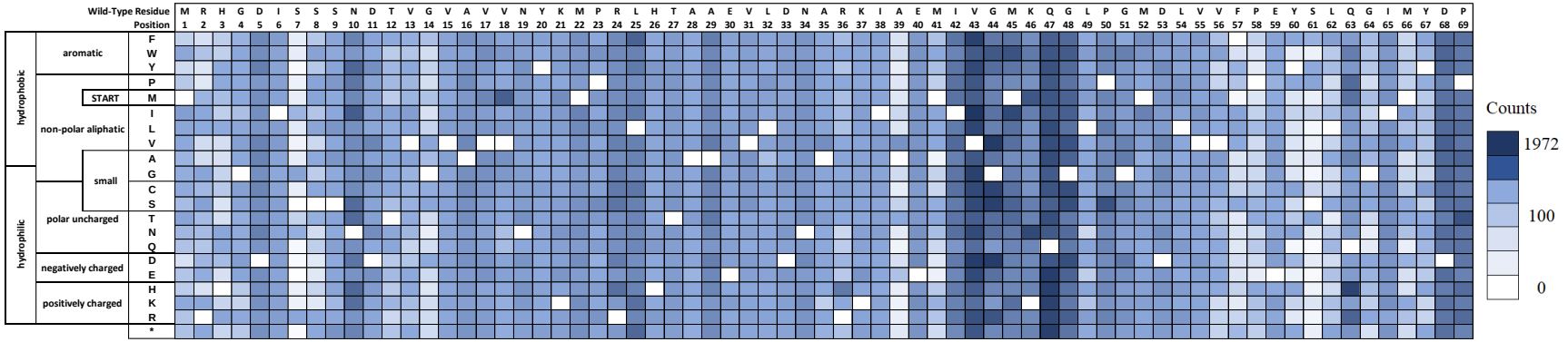


Figure D 17. Correlation between counts in the displayed population relative to the counts in the unselected population for cNGF.

a.



b.

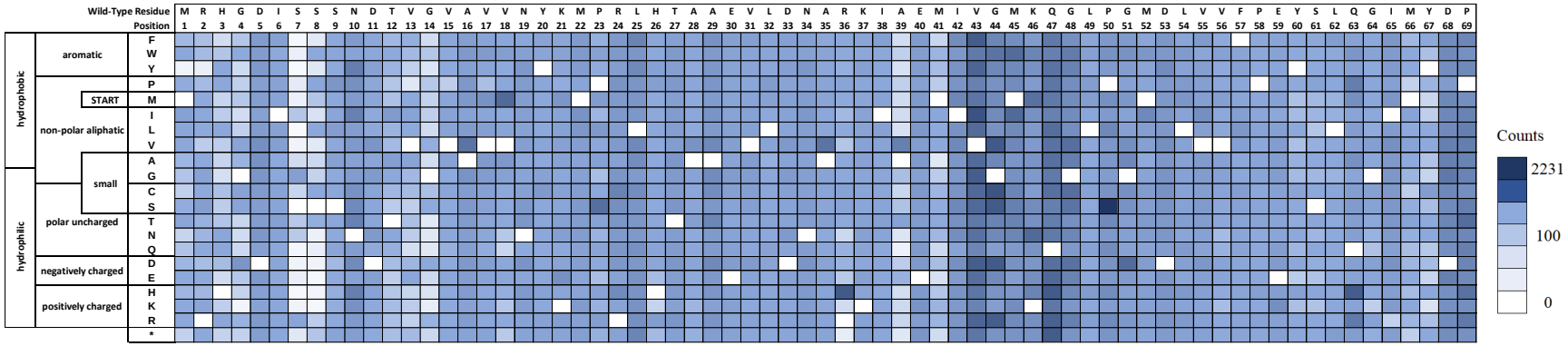


Figure D 18. Per-position heatmap of sequencing counts for AmiE mutants in (a.) replicate 1 and (b.) replicate 2 using the oligo pool.

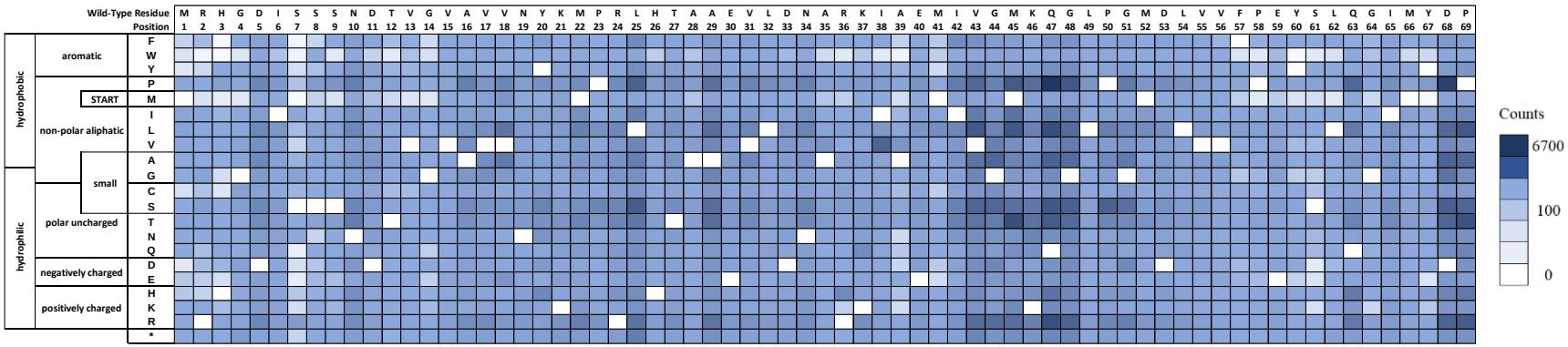
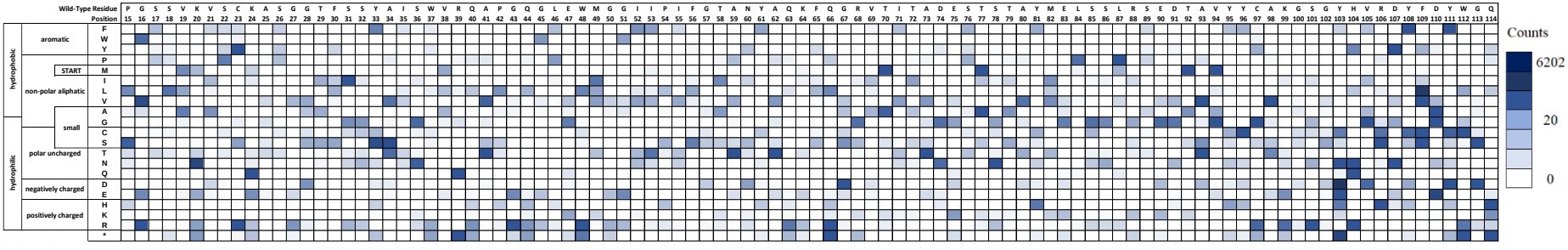


Figure D 19. Per-position heatmap of sequencing counts for AmiE mutants using degenerate "NNN" oligos.

a.



b.

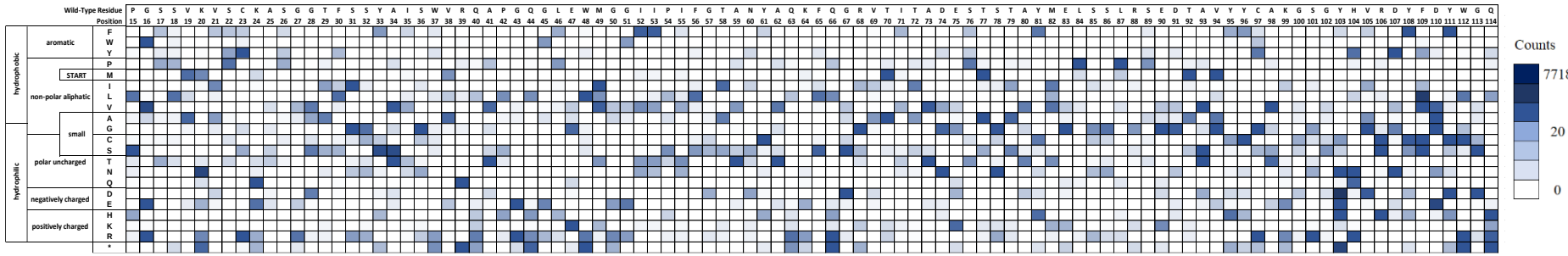
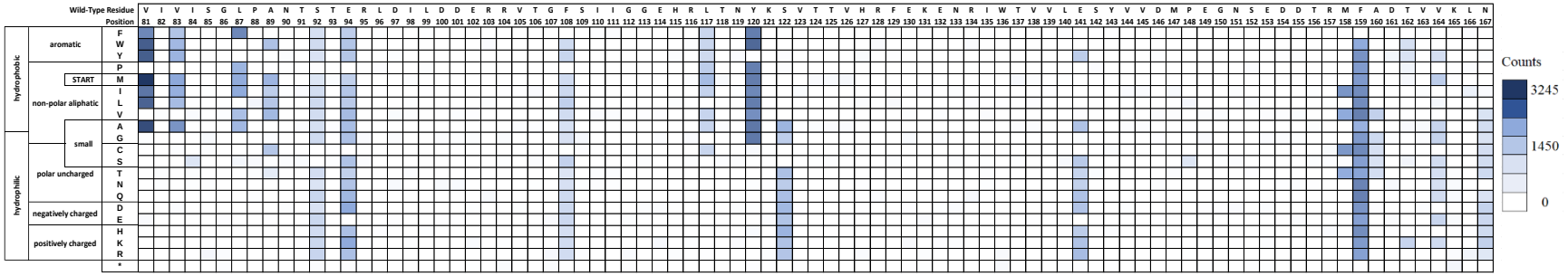


Figure D 20. Per-position heatmap of sequencing counts of UCA9 mutants in (a.) replicate 1 and (b.) replicate 2 using the oligo pool.

a.



b.

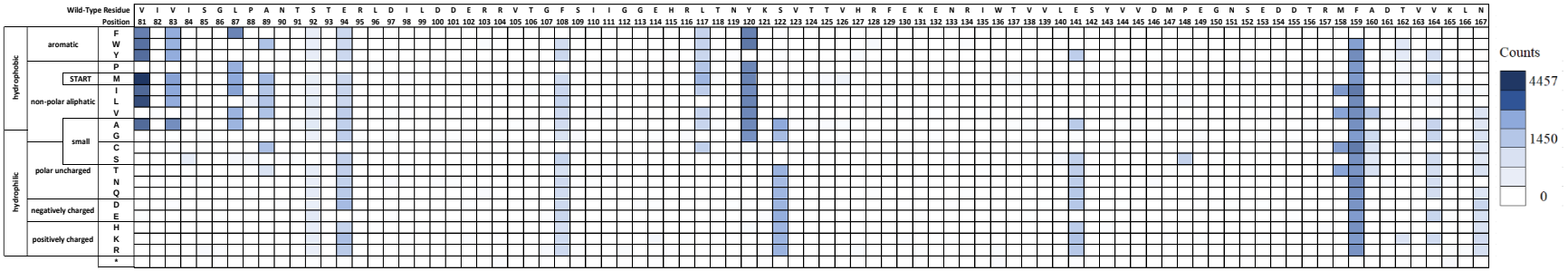


Figure D 21. Per-position heatmap of sequencing counts for PYR1 mutants (a.) replicate 1 and (b.) replicate 2 using the oligo pool.

APPENDIX E:

Supplementary Tables

Table E 1. Sorting conditions and FACS collection statistics for fIL-31 libraries

Antigen	*Amount of Collected Cells*	*Labeling Concentration [nM]*	*Percent Sorted Library 1*	*Percent Sorted Library 2*
fOSMR-ECD	300,000	5.1	6.5%	6.7%
fIL31RA-1FNIII	300,000	52.8	8.5%	6.0%
mAb #1	300,000	0.13	7.6%	7.8%

Table E 2. Primers for deep sequencing. L1: Library 1, L2: Library 2, Blue: Illumina Universal Sequence, **NNNNNN:** Indexing Barcode, Purple: Illumina Adapter, UN: Unselected Population, and CMYC: Displayed Population

Name	*Sequence*
Inner Primer:	
fIL-31_L1_FWD	5' - GTTCAGAGTTCTACAGTCCGACGATCAGGGTCGGCTAGCCATATG - 3'
fIL-31_L1_REV	5' - CCTTGGCACCCGAGAATTCCATCTGACAATGGTCTAATAGCTCT - 3'
fIL-31_L2_FWD	5' - GTTCAGAGTTCTACAGTCCGACGATCTCAGCAATCTTGCCATACTTT - 3'
fIL-31_L2_REV	5' - CCTTGGCACCCGAGAATTCCAATAAGCTTTTGTTCGGATCCG - 3'
Outer Primers:	
Illumina_FWD	5' - AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATC - 3'
Top 5% Displayed Experiment:	
fIL-31_L1_UN_REV	5' - CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L1_CMYC_REV	5' - CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L2_UN_REV	5' - CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L2_CMYC_REV	5' - CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Labeled with mAb#1 or fOSMR-ECD:	
fIL-31_L1_UN_REV	5' - CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L1_mAb#1_REV	5' - CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L1_OSMR_REV	5' - CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL31_L2_UN_REV	5' - CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'

Table E2 (cont'd)

fIL-31_L2_mAb#1_REV	5' - CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L2_OSMR_REV	5' - CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Labeled with fIL31RA-1FNIII:	
fIL31_L1_UN_REV	5' - CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L1_IL31RA-1FNIII_REV	5' - CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL31_L2_UN_REV	5' - CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
fIL-31_L2_IL31RA-1FNIII_REV	5' - CAAGCAGAAGACGGCATACGAGATGCTCATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'

Table E 3. fIL-31 library statistics results

	fOSMR-ECD and mAb#1		***fIL31RA-1FNIII***		***Top 5% Displayed***	
	Library 1	Library 2	Library 1	Library 2	Library 1	Library 2
Percent of possible codon substitutions observed in the unselected population:						
1-base substitution	100%	100.3%	94.4%	95.8%	97.4%	97.4%
2-base substitutions	52.0%	56.4%	39.9%	42.0%	38.5%	42.8%
3-base substitutions	38.3%	41.5%	37.1%	38.0%	37.3%	38.8%
Percent of unselected reads with:						
No nonsynonymous mutations:	41.8%	45.6%	39.1%	45.3%	44.5%	45.9%
One nonsynonymous mutation:	52.6%	50.8%	58.6%	52.3%	48.3%	49.3%
Multiple nonsynonymous mutations:	5.5%	3.5%	2.3%	2.3%	7.2%	4.8%
Coverage of possible single nonsynonymous amino acid mutations:	**77.8%**	**83.2%**	**75.7%**	**80.0%**	**77.9%**	**81.0%**

Table E 4. FACS collection statistics for Pro-cNGF and ProΔ1,2-cNGF library screening experiments

	Library Size (aa's)	Sort Round 1		Sort Round 2	
		tanezumab	mAb #1	tanezumab	mAb #1
Pro-cNGF_tanezumab library 1	51	2.9%	2.4%	3.1%	2.7%
Pro-cNGF_tanezumab library 2	52	3.3%	3.1%	2.8%	2.8%
ProΔ1,2-cNGF_tanezumab	56	3.3%	3.6%	-	-

Table E 5. Summary of Average Dissociation Constant, K_D values using Surface Plasmon Resonance (SPR) for human pro-NGF and canine NGF and Yeast Surface Display (YSD) for pro.v4-cNGF, and sorting conditions for library screening using pro.v4-cNGF. The K_D values from SPR were obtained using the 1:1 biding model. Error bars represent 1 standard deviation of the regression. One tail t-test assuming unequal variances was used to calculate p-values for Hill coefficients. ($n \geq 3$)

	Surface Plasmon Resonance Data						Yeast Surface Display Data				
mAb	Human Pro-NGF		Human NGF		Canine NGF		Average K_D values with Hill coefficient, H=1 [pM]	Labeling Concentrations for Screening Libraries [pM]	Average K_D values varying Hill coefficient, H [pM]	Hill coefficient values	p- values for Hill coefficient
	Average K_D values [pM]	Chi2	Average K_D values [pM]	Chi2	Average K_D values [pM]	Chi2					
tanezumab	1610	0.01	15.9	1.26	19	0.77	801 ± 164	400.6	1319 ± 200	0.66 ± 0.05	0.0033
mAb #1	286	0.22	0.243	2.32	0.118	1.92	209 ± 65	104.6	189 ± 36	0.70 ± 0.18	0.0103
mAb #2	48	0.05	0.308	1.47	0.074	1.29	307 ± 173	153.6	461 ± 300	0.62 ± 0.09	0.0003
mAb #3	2500	0.07	1.24	5.36	0.179	4.92	143 ± 44	71.6	325 ± 328	0.58 ± 0.19	0.0106

Table E 6. FACS collection statistics for cNGF libraries

mAb	Amount of collected cells	cNGF	
		Library 1 (90 aa's)	Library 2 (90 aa's)
tanezumab	250,000	8.01%	6.10%
mAb #1	250,000	7.36%	6.33%
mAb #2	250,000	7.28%	6.60%
mAb #3	250,000	6.59%	8.48%

Table E 7. Primers set for deep sequencing. L1: library 1, L2: library 2, Blue: Illumina Universal Sequence, **NNNNNN**: Indexing Barcode, and Green: Illumina Adapter

Name	Sequence
Inner Primers	
Pro-cNGF_L1_FWD	5'- GTTCAGAGTTCTACAGTCCGACGATCGATGACGACAAGCATATG -3'
Pro-cNGF_L1_REV	5'- CCTTGGCACCCGAGAATTCCAAACTTTGGATCAACTGTGAT -3'
Pro-cNGF_L2_FWD	5'- GTTCAGAGTTCTACAGTCCGACGATC TTACAGGTCAAACTAGAAAC-3'
Pro-cNGF_L2_REV	5'- CCTTGGCACCCGAGAATTCCAAACTGGATGAGATGAAGA -3'
ProΔ1,2-cNGF_FWD	5'- GTTCAGAGTTCTACAGTCCGACGATCGATGACGACAAGCATATG-3'
ProΔ1,2-cNGF_REV	5'- CCTTGGCACCCGAGAATTCCAAACTGGATGAGATGAAGA -3'
cNGF_L1_FWD	5'- GTTCAGAGTTCTACAGTCCGACGATCAACTGGATGAGATGAAGA -3'
cNGF_L1_REV	5'- CCTTGGCACCCGAGAATTCCAATCAACTGGAGTTGG -3'
cNGF_L2_FWD	5'- GTTCAGAGTTCTACAGTCCGACGATCCTTTTTTGAAACAAAATGTAGAGAT -3'
cNGF_L2_REV	5'- CCTTGGCACCCGAGAATTCCAGCCTCCTCCACC -3'
Foward Outer Primer	
Illumina_FWD	5'- AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATC - 3'
Reverse Outer Primers	
Pro-cNGF_L1_Unsel	5'- CAAGCAGAAGACGGCATACGAGAT**TGACAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L1_Display	5'- CAAGCAGAAGACGGCATACGAGAT**GGACGG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L1_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**CTCTAC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L1_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**GCGGAC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'

Table E7 (cont'd)

Pro-cNGF_L2_Unsel	5'- CAAGCAGAAGACGGCATACGAGAT**TTTCAC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L2_Display	5'- CAAGCAGAAGACGGCATACGAGAT**GGCCAC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L2_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**CGAAAC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L2_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**CGTACG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
ProΔ1,2-cNGF_Unsel	5'- CAAGCAGAAGACGGCATACGAGAT**CCACTC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
ProΔ1,20-NGF_Display	5'- CAAGCAGAAGACGGCATACGAGAT**GCTACC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
ProΔ1,2-cNGF_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**ATCAGT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
ProΔ1,2-cNGF_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**GCTCAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L1_S2_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**AGGAAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L1_S2_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**CTTTTG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L2_S2_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**TAGTTG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
Pro-cNGF_L2_S2_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**CCGGTG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L1_Unsel	5'- CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L1_Display	5'- CAAGCAGAAGACGGCATACGAGAT**ACATCG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L1_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**GCCTAA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L1_mAb #2	5'- CAAGCAGAAGACGGCATACGAGAT**TGGTCA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L1_mAb #3	5'- CAAGCAGAAGACGGCATACGAGAT**CACTGT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L1_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**ATTGGC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L2_Unsel	5'- CAAGCAGAAGACGGCATACGAGAT**CTGATC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L2_Display	5'- CAAGCAGAAGACGGCATACGAGAT**AAGCTA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cGF_L2_mAb #1	5'- CAAGCAGAAGACGGCATACGAGAT**GTAGCC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L2_mAb #2	5'- CAAGCAGAAGACGGCATACGAGAT**TACAAG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L2_mAb #3	5'- CAAGCAGAAGACGGCATACGAGAT**TTGACT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'
cNGF_L2_tanezumab	5'- CAAGCAGAAGACGGCATACGAGAT**GGAACT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3'

Table E 8. Libraries Statistics Results for cNGF constructs

	Pro-cNGF		***ProΔ1,2-cNGF***	***NGF***	
	Tile 1	***Tile 2***		***Tile 1***	***Tile 2***
Percent of possible codon subsititions observed in the unselected population:					
1-base substitution	96.50%	95.10%	94.20%	98.70%	99.80%
2-base substitutions	59.50%	52.80%	53.20%	55.70%	60.00%
3-base substitutions	47.40%	43.20%	44.80%	44.50%	47.90%
Percent of unselected reads with:					
No nonsynonymous mutations:	35.20%	30.80%	28.60%	39.50%	34.50%
One nonsynonymous mutation:	56.00%	61.80%	56.60%	55.40%	57.10%
Multiple nonsynonymous mutations:	8.80%	7.40%	14.80%	5.10%	8.50%
Coverage of possible single nonsynonymous amino acid mutations:	**85.50%**	**83.20%**	**84.20%**	**92.90%**	**98.20%**

Table E 9. Primers set for deep sequencing. DSM: double site mutant library, Rep 1: Replicate 1, Rep 2: replicate 2, Red: Illumina Universal Sequence, **NNNNNN**: Indexing Barcode, and Blue: Illumina Adapter

Name	Sequence
Inner Primers	
AmiE_FWD	5' - GTTCAGAGTTCTACAGTCCGACGATCTTAACTTTAAGAAGTTTTTATACAT - 3'
AmiE_REV	5' - CCTTGGCACCCGAGAATTCCAAAGCACGGCTAAAGAT - 3'
PYR1_FWD	5' - GTTCAGAGTTCTACAGTCCGACGATCATGCACGCGCGAC - 3'
PYR1_REV	5' - CCTTGGCACCCGAGAATTCCACGCGAGTTTCTGCAA - 3'
UCA9_FWD	5' - GTTCAGAGTTCTACAGTCCGACGATCGGGCTGAGGTGAAGAAG - 3'
UCA9_REV	5' - CCTTGGCACCCGAGAATTCCAGGTGACCAGGGTTCC - 3'
Forward Outer Primer	
Illumina FWD	5’ - AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATC - 3’
Reverse Outer Primers	
AmiE_Rep 1	5’ - CAAGCAGAAGACGGCATACGAGAT**GATCTG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
AmiE_Rep 2	5’ - CAAGCAGAAGACGGCATACGAGAT**TCAAGT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
PYR1_Rep 1	5’ - CAAGCAGAAGACGGCATACGAGAT**CACTGT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
PYR1_Rep 2	5’ - CAAGCAGAAGACGGCATACGAGAT**ATTGGC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
PYR1_DSM	5’ - CAAGCAGAAGACGGCATACGAGAT**GCCTAA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
UCA9_Rep 1	5’ - CAAGCAGAAGACGGCATACGAGAT**TTGACT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
UCA9_Rep 2	5’ - CAAGCAGAAGACGGCATACGAGAT**AGGAAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’
AmiE_"NNN"	5’ - CAAGCAGAAGACGGCATACGAGAT**CTCTAC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA - 3’

Table E 10. Summary of statistics for single site saturation mutagenesis (SSM) and double site saturation mutagenesis (DSM) libraries

	Oligo Pool Primers							**Degenerate "NNN" Oligos**
	SSM						**DSM**	**SSM**
	AmiE Replicate 1	AmiE Replicate 2	PYR1 Replicate 1	PYR1 Replicate 2	UCA9 Replicate 1	UCA9 Replicate 2	PYR1	AmiE
Percentage of reads with:								
No nonsynonymous mutations:	50.8%	54.6%	63.0%	55.8%	71.1%	71.5%	60.0%	39.0%
One nonsynonymous mutation:	38.0%	34.9%	24.1%	29.0%	14.3%	13.9%	32.0%*	48.0%
Multiple nonsynonymous mutations:	11.2%	10.5%	11.8%	15.2%	15.5%	14.6%	9.0%‡	13.0%
Coverage of all possible nonsynonymous amino acid mutations:	100%	100%	100%	100%	96.7%	98%	79.2%*	100%

*1 to 2 nonsynonymous mutations

‡More than 2 nonsynonymous mutations

REFERENCES

REFERENCES

1. Schreiber, G. & Fleishman, S. J. Computational design of protein–protein interactions. *Curr. Opin. Struct. Biol.* **23,** 903–910 (2013).

2. Ecker, D. M., Jones, S. D. & Levine, H. L. The therapeutic monoclonal antibody market. *MAbs* **7,** 9–14 (2015).

3. Grilo, A. L. & Mantalaris, A. The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.* **37,** 9–16 (2019).

4. Scott, A. M., Allison, J. P. & Wolchok, J. D. Monoclonal antibodies in cancer therapy. *Cancer Immun.* **12,** 14 (2012).

5. Mata-Fink, J. *et al.* Rapid Conformational Epitope Mapping of Anti-gp120 Antibodies with a Designed Mutant Panel Displayed on Yeast. *J. Mol. Biol.* **425,** 444–456 (2013).

6. Sukupolvi-Petty, S. *et al.* Structure and function analysis of therapeutic monoclonal antibodies against dengue virus type 2. *J. Virol.* **84,** 9227–39 (2010).

7. Correia, B. E. *et al.* Proof of principle for epitope-focused vaccine design. *Nature* **507,** 201–206 (2014).

8. Whitehead, T. A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30,** 543–548 (2012).

9. Wang, L.-F. & Yu, M. Epitope Identification and Discovery Using Phage Display Libraries: Applications in Vaccine Development and Diagnostics. *Curr. Drug Targets* **5,** 1–15 (2004).

10. Van Blarcom, T. *et al.* Precise and Efficient Antibody Epitope Determination through Library Design, Yeast Display and Next-Generation Sequencing. *J. Mol. Biol.* **427,** 1513–1534 (2015).

11. Doolan, K. M. & Colby, D. W. Conformation-Dependent Epitopes Recognized by Prion Protein Antibodies Probed Using Mutational Scanning and Deep Sequencing. *J. Mol. Biol.* **427,** 328–340 (2015).

12. Kowalsky, C. A. *et al.* Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing. *J. Biol. Chem.* **290,** 26457–70 (2015).

13. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11,** 801–807 (2014).

14. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8950–4 (2000).

15. Chao, G., Cochran, J. R. & Dane Wittrup, K. Fine Epitope Mapping of anti-Epidermal

Growth Factor Receptor Antibodies Through Random Mutagenesis and Yeast Surface Display. *J. Mol. Biol.* **342,** 539–550 (2004).

16. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Publ. Gr.* **7,** 741–746 (2010).

17. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1,** 755–768 (2006).

18. Deventer, J. A. Van & Wittrup, K. D. Yeast Surface Display for Antibody Isolation: Library Construction, Library Screening, and Affi nity Maturation. *Methods Mol. Biol.* **1131,** 151–181 (2014).

19. Adams, R. M., Mora, T., Walczak, A. M. & Kinney, J. B. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *Elife* **5,** 5980–5985 (2016).

20. Le Saux, S. *et al.* Molecular dissection of human interleukin-31-mediated signal transduction through site-directed mutagenesis. *J. Biol. Chem.* **285,** 3470–7 (2010).

21. Burns, M. L. *et al.* Directed evolution of brain-derived neurotrophic factor for improved folding and expression in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **80,** 5732–42 (2014).

22. Huang, E. J. & Reichardt, L. F. Neurotrophins: Roles in Neuronal Development and Function. *Annu. Rev. Neurosci.* **24,** 677–736 (2001).

23. Wrenbeck, E. E. *et al.* Plasmid-based one-pot saturation mutagenesis. *Nat. Methods* **13,** 928–930 (2016).

24. Medina-Cucurella, A. V. & Whitehead, T. A. Characterizing Protein-Protein Interactions Using Deep Sequencing Coupled to Yeast Surface Display. *Methods Mol. Biol.* **1764**, 101-121 (2018).

25. Kowalsky, C. A. *et al.* High-Resolution Sequence-Function Mapping of Full-Length Proteins. *PLoS One* **10,** e0118193 (2015).

26. Nelson, A. L., Dhimolea, E. & Reichert, J. M. Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discov.* **9,** 767–774 (2010).

27. Smith, S. A. & Crowe, Jr., J. E. Use of Human Hybridoma Technology To Isolate Human Monoclonal Antibodies. *Microbiol. Spectr.* **3,** 141–156 (2015).

28. Adler, A. S. *et al.* Rare, high-affinity mouse anti-PD-1 antibodies that function in checkpoint blockade, discovered using microfluidics and molecular genomics. *MAbs* **9,** 1270–1281 (2017).

29. Adler, A. S. *et al.* Rare, high-affinity anti-pathogen antibodies from human repertoires, discovered using microfluidics and molecular genomics. *MAbs* **9,** 1282–1296 (2017).

30. Adler, A. S. *et al.* A natively paired antibody library yields drug leads with higher sensitivity and specificity than a randomly paired antibody library. *MAbs* **10,** 431–443 (2018).

31. Medina-Cucurella, A. *et al.* Preferential Identification of Agonistic OX40 Antibodies by Using Cell Lysate to Pan Natively Paired, Humanized Mouse-Derived Yeast Surface Display Libraries. *Antibodies* **8,** 17 (2019).

32. Dillon, S. R. *et al.* Interleukin 31, a cytokine produced by activated T cells, induces dermatitis in mice. *Nat. Immunol.* **5,** 752–760 (2004).

33. Neis, M. M. *et al.* Enhanced expression levels of IL-31 correlate with IL-4 and IL-13 in atopic and allergic contact dermatitis. *J. Allergy Clin. Immunol.* **118,** 930–937 (2006).

34. Rabenhorst, A. & Hartmann, K. Interleukin-31: A Novel Diagnostic Marker of Allergic Diseases. *Curr. Allergy Asthma Rep.* **14,** 423 (2014).

35. Cornelissen, C., Lüscher-Firzlaff, J., Baron, J. M. & Lüscher, B. Signaling by IL-31 and functional consequences. *Eur. J. Cell Biol.* **91,** 552–566 (2012).

36. Takaoka, A. *et al.* Involvement of IL-31 on scratching behavior in NC/Nga mice with atopic-like dermatitis. *Exp. Dermatol.* **15,** 161–167 (2006).

37. Sonkoly, E. *et al.* IL-31: A new link between T cells and pruritus in atopic skin inflammation. *J. Allergy Clin. Immunol.* **117,** 411–417 (2006).

38. Lewis, K. E. *et al.* Interleukin (IL) 31 induces in cynomolgus monkeys a rapid and intense itch response that can be inhibited by an IL-31 neutralizing antibody. *J. Eur. Acad. Dermatology Venereol.* **31,** 142–150 (2017).

39. Raap, U. *et al.* Correlation of IL-31 serum levels with severity of atopic dermatitis. *J. Allergy Clin. Immunol.* **122,** 421–3 (2008).

40. Ezzat, M., Hasan, Z. & Shaheen, K. Serum measurement of interleukin-31 (IL-31) in paediatric atopic dermatitis: elevated levels correlate with severity scoring. *J. Eur. Acad. Dermatology Venereol.* **25,** 334–339 (2011).

41. Pantazi, E., Valenza, G., Hess, M. & Hamad, B. The atopic dermatitis market. *Nat. Rev. Drug Discov.* **17,** 237 (2017).

42. Nemoto, O. *et al.* The first trial of CIM331, a humanized antihuman interleukin-31 receptor A antibody, in healthy volunteers and patients with atopic dermatitis to evaluate safety, tolerability and pharmacokinetics of a single dose in a randomized, double-blind, placebo-co. *Br. J. Dermatol.* **174,** 296–304 (2016).

43. Michels, G. M. *et al.* A blinded, randomized, placebo-controlled, dose determination trial of lokivetmab (ZTS-00103289), a caninized, anti-canine IL-31 monoclonal antibody in client owned dogs with atopic dermatitis. *Vet. Dermatol.* **27,** 478-e129 (2016).

44. Diveu, C. *et al.* Predominant expression of the long isoform of GP130-like (GPL) receptor is required for interleukin-31 signaling. *Eur. Cytokine Netw.* **15,** 291–302 (2004).

45. Zhang, Q., Putheti, P., Zhou, Q., Liu, Q. & Gao, W. Structures and biological functions of IL-31 and IL-31 receptors. *Cytokine Growth Factor Rev.* **19,** 347–356 (2008).

46. Diveu, C. *et al.* GPL, a novel cytokine receptor related to GP130 and leukemia inhibitory factor receptor. *J. Biol. Chem.* **278,** 49850–9 (2003).

47. Dambacher, J. *et al.* Interleukin 31 mediates MAP kinase and STAT1/3 activation in intestinal epithelial cells and its expression is upregulated in inflammatory bowel disease. *Gut* **56,** 1257–65 (2007).

48. Dreuw, A. *et al.* Characterization of the signaling capacities of the novel gp130-like cytokine receptor. *J. Biol. Chem.* **279,** 36112–20 (2004).

49. Boulanger, M. J., Chow, D., Brevnova, E. E. & Garcia, K. C. Hexameric Structure and Assembly of the Interleukin-6/IL-6 α-Receptor/gp130 Complex. *Science* **300,** 2101–2104 (2003).

50. Araya, C. L. & Fowler, D. M. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* **29,** 435–42 (2011).

51. Medina-Cucurella, A. V., Zhu, Y., Bowen, S. J., Bergeron, L. M. & Whitehead, T. A. Pro region engineering of nerve growth factor by deep mutational scanning enables a yeast platform for conformational epitope mapping of anti-NGF monoclonal antibodies. *Biotechnol. Bioeng.* **115,** 1925–1937 (2018).

52. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738 (2010).

53. Zhang, Y. *et al.* I-TASSER server for protein 3D structure prediction. *BMC Bioinforma. 2008 91* **59,** 305–309 (2008).

54. Chen, V. B. *et al.* Biological Crystallography MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* **66**, 12-21 (2010).

55. Lessmann, V., Gottmann, K. & Malcangio, M. Neurotrophin secretion: current facts and future prospects. *Prog. Neurobiol.* **69,** 341–374 (2003).

56. Park, H. & Poo, M. Neurotrophin regulation of neural circuit development and function. *Nat. Rev. Neurosci.* **14,** 7–23 (2013).

57. Rattenholl, A. *et al.* The pro-sequence facilitates folding of human nerve growth factor from Escherichia coli inclusion bodies. *Eur. J. Biochem.* **268,** 3296–3303 (2001).

58. Watson, J. J., Allen, S. J. & Dawbarn, D. Targeting Nerve Growth Factor in Pain. *BioDrugs* **22,** 349–359 (2008).

59. Slosky, L. M., Largent-Milnes, T. M. & Vanderah, T. W. Use of Animal Models in Understanding Cancer-induced Bone Pain. *Cancer Growth Metastasis* **8,** 47–62 (2015).

60. La Porte, S. L. *et al.* Generation of a high-fidelity antibody against nerve growth factor using library scanning mutagenesis and validation with structures of the initial and optimized Fab-antigen complexes. *MAbs* **6,** 1059–1068 (2014).

61. Chang, D. S., Hsu, E., Hottinger, D. G. & Cohen, S. P. Anti-nerve growth factor in pain management: current evidence. *J. Pain Res.* **9,** 373–83 (2016).

62. Kumar, V. & Mahal, B. A. NGF - the TrkA to successful pain treatment. *J. Pain Res.* **5,** 279–87 (2012).

63. Burns, M. L. *et al.* Pro-region engineering for improved yeast display and secretion of brain derived neurotrophic factor. *Biotechnol. J.* **11,** 425–436 (2016).

64. Kliemannel, M., Golbik, R., Rudolph, R., Schwarz, E. & Lilie, H. The pro-peptide of proNGF: Structure formation and intramolecular association with NGF. *Protein Sci.* **16,** 411–419 (2007).

65. Nomoto, H., Takaiwa, M., Mouri, A. & Furukawa, S. Pro-region of neurotrophins determines the processing efficiency. *Biochem. Biophys. Res. Commun.* **356,** 919–924 (2007).

66. Hauburger, A., Kliemannel, M., Madsen, P., Rudolph, R. & Schwarz, E. Oxidative folding of nerve growth factor can be mediated by the pro-peptide of neurotrophin-3. *FEBS Lett.* **581,** 4159–4164 (2007).

67. Rattenholl, A. *et al.* Pro-sequence assisted folding and disulfide bond formation of human nerve growth factor. *J. Mol. Biol.* **305,** 523–533 (2001).

68. Feng, D. *et al.* Molecular and Structural Insight into proNGF Engagement of p75NTR and Sortilin. *J. Mol. Biol.* **396,** 967–984 (2010).

69. Kliemannel, M. *et al.* The mature part of proNGF induces the structure of its pro-peptide. *FEBS Lett.* **566,** 207–212 (2004).

70. Suter, U., Heymach, J. V, Shooter, E. M. & Shooter, E. M. Two conserved domains in the NGF propeptide are necessary and sufficient for the biosynthesis of correctly processed and biologically active NGF. *EMBO J.* **10,** 2395–400 (1991).

71. Pagadala, P. C., Dvorak, L. A. & Neet, K. E. Construction of a mutated pro-nerve growth factor resistant to degradation and suitable for biophysical and cellular utilization. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 17939–43 (2006).

72. Wrenbeck, E. E., Faber, M. S. & Whitehead, T. A. Deep sequencing methods for protein engineering and design. *Curr. Opin. Struct. Biol.* **45,** 36–44 (2017).

73. Kitamura, T. *et al.* Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin. *J. Cell. Physiol.* **140,** 323–334 (1989).

74. Kowalsky, C. A. & Whitehead, T. A. Determination of binding affinity upon mutation for type I dockerin-cohesin complexes from C lostridium thermocellum and C lostridium cellulolyticum using deep sequencing. *Proteins Struct. Funct. Bioinforma.* **84,** 1914–1928 (2016).

75. Klesmith, J. R., Bacik, J.-P., Wrenbeck, E. E., Michalczyk, R. & Whitehead, T. A. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 2265–2270 (2017).

76. Abeliovich, H. An Empirical Extremum Principle for the Hill Coefficient in Ligand-Protein Interactions Showing Negative Cooperativity. *Biophys. J.* **89,** 76–79 (2005).

77. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol.*

Cell Biol. **4,** 181–191 (2003).

78. Park, S. *et al.* Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Eng. Des. Sel.* **19,** 211–217 (2006).

79. Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **17,** 467–473 (2007).

80. Firnberg, E. & Ostermeier, M. PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS One* **7,** e52031 (2012).

81. Cozens, C. & Pinheiro, V. B. Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis. *Nucleic Acids Res.* **46,** e51–e51 (2018).

82. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods.* **11,** 499–507 (2014).

83. Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D. & Kosuri, S. Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science* **359,** 343–347 (2018).

84. Wrenbeck, E. E., Azouz, L. R. & Whitehead, T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8,** 15695 (2017).

85. Pappas, L. *et al.* Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature* **516,** 418–422 (2014).

86. Park, S.-Y. *et al.* Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* **520,** 545–548 (2015).

87. Klesmith, J. R. & Hackel, B. J. Improved mutant function prediction via PACT: Protein Analysis and Classifier Toolkit. *Bioinformatics.* (2018).

88. Wrenbeck, E., Klesmith, J., Stapleton, J. & Whitehead, T. Nicking Mutagenesis: comprehensive single-site saturation mutagenesis. *Protoc. Exch.* (2016).

89. Das, R. & Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **77,** 363–382 (2008).

90. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487,** 545–574 (2011).

91. Chaudhury, S. *et al.* Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. *PLoS One* **6,** e22477 (2011).

92. van der SchaarJan, H. M., Wilschut, J. C. & Smit, J. M. Role of antibodies in controlling dengue virus infection. *Immunobiology* **214,** 613–629 (2009).

93. Schwartz, L. M., Halloran, M. E., Durbin, A. P. & Longini, I. M. The dengue vaccine pipeline: Implications for the future of dengue control. *Vaccine* **33,** 3293–3298 (2015).

94. Katzelnick, L. C. *et al.* Antibody-dependent enhancement of severe dengue disease in humans. *Science* **358,** 929–932 (2017).

95. Murrell, S., Wu, S.-C. & Butler, M. Review of dengue virus and the development of a vaccine. *Biotechnol. Adv.* **29,** 239–247 (2011).

96. Dejnirattisai, W. *et al.* Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. *Nat. Immunol.* **17,** 1102–1108 (2016).

97. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2,** 31–34 (2007).

98. Sambrook, J. & Russell, D. W. Transformation of E. coli by Electroporation. *CSH Protoc.* **2006,** pdb.prot3933 (2006).

99. Fowler, D. M., Araya, C. L., Gerard, W. & Fields, S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* **27,** 3430–3431 (2011).

100. KÖHLER, G. & MILSTEIN, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256,** 495–497 (1975).

101. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348,** 552–554 (1990).

102. Kohl, T. O. & Ascoli, C. A. Direct and Indirect Cell-Based Enzyme-Linked Immunosorbent Assay. *Cold Spring Harb. Protoc.* **2017,** pdb.prot093732 (2017).

103. Even-Desrumeaux, K. & Chames, P. Phage display and selections on cells. *Methods Mol Biol.* **907**, 225–235 (2012).

104. Spencer, S., Bethea, D., Raju, T. S., Giles-Komar, J. & Feng, Y. Solubility evaluation of murine hybridoma antibodies. *MAbs* **4,** 319–325 (2012).

105. Jain, T. *et al.* Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 944–949 (2017).

106. Rajan, S. *et al.* Recombinant human B cell repertoires enable screening for rare, specific, and natively paired antibodies. *Commun. Biol.* **1,** 5 (2018).

107. Wang, B. *et al.* Functional interrogation and mining of natively paired human VH:VL antibody repertoires. *Nat. Biotechnol.* **36,** 152–155 (2018).

108. Linch, S. N., McNamara, M. J. & Redmond, W. L. OX40 Agonists and Combination Immunotherapy: Putting the Pedal to the Metal. *Front. Oncol.* **5,** 34 (2015).

109. Willoughby, J., Griffiths, J., Tews, I. & Cragg, M. S. OX40: Structure and function – What questions remain? *Mol. Immunol.* **83,** 13–22 (2017).

110. Compaan, D. M. & Hymowitz, S. G. The Crystal Structure of the Costimulatory OX40-OX40L Complex. *Structure* **14,** 1321–1330 (2006).

111. Tillotson, B. J., De Larrinoa, I. F., Skinner, C. A., Klavas, D. M. & Shusta, E. V. Antibody affinity maturation using yeast display with detergent-solubilized membrane proteins as antigen sources. *Protein Eng. Des. Sel.* **26,** 101–112 (2013).

112. Cho, Y. K. & Shusta, E. V. Antibody library screens using detergent-solubilized

mammalian cell lysates as antigen sources. *Protein Eng. Des. Sel.* **23,** 567–577 (2010).

113. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31,** 3476–3482 (2015).

114. Lefranc, M.-P. *et al.* IMGT(R), the international ImMunoGeneTics information system(R). *Nucleic Acids Res.* **37,** D1006–D1012 (2009).

115. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

116. Csardi, G. & Nepusz, T. The igraph software package for complex network research | BibSonomy. *InterJournal, Complex Systems* **1695**, 1–9 (2006).

117. Dreyer, A. M., Beauchamp, J., Matile, H. & Pluschke, G. An efficient system to generate monoclonal antibodies against membrane-associated proteins by immunisation with antigen-expressing mammalian cells. *BMC Biotechnol.* **10,** 87 (2010).

118. Rezaei, M. & Ghaderi, A. Production of a Mouse Monoclonal Antibody Against Mortalin by Whole Cell Immunization. *Monoclon. Antib. Immunodiagn. Immunother.* **36,** 169–175 (2017).

119. Tamura, T. & Chiba, J. Production of antibodies against multipass membrane proteins expressed in human tumor cells using dendritic cell immunization. *J. Biomed. Biotechnol.* **2009,** 673098 (2009).

120. Reddy, S. T. *et al.* Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat. Biotechnol.* **28,** 965–969 (2010).

121. Saggy, I. *et al.* Antibody isolation from immunized animals: comparison of phage display and antibody discovery via V gene repertoire mining. *Protein Eng. Des. Sel.* **25,** 539–549 (2012).

122. Wilson, J. R. *et al.* Diversity of the murine antibody response targeting influenza A(H1N1pdm09) hemagglutinin. *Virology* **458–459,** 114–124 (2014).