Reduction of Nonspecificity Motifs in Synthetic Antibody Libraries

As Published: 10.1016/J.JMB.2017.11.008

Publisher: Elsevier BV

Persistent URL: https://hdl.handle.net/1721.1/134710

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2019 January 05.

Published in final edited form as:

J Mol Biol. 2018 January 05; 430(1): 119–130. doi:10.1016/j.jmb.2017.11.008.

Reduction of nonspecificity motifs in synthetic antibody libraries

Ryan L. Kelly[1], **Doris Le**[2], **Jessie Zhao**[2], and **K. Dane Wittrup**[1,2,*]

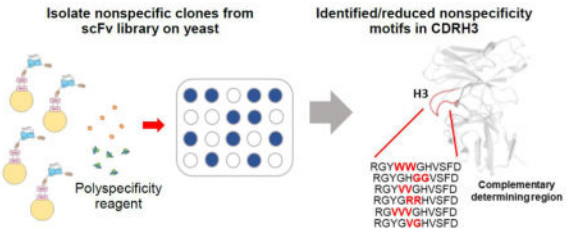
[1]Department of Biological, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA U.S.A

[2]Department of Chemical Engineering, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA U.S.A

Abstract

Successful antibody development requires both functional binding and desirable biophysical characteristics. In the current study we analyze the causes of one hurdle to clinical development, off-target reactivity or nonspecificity. We used a high-throughput nonspecificity assay to isolate panels of nonspecific antibodies from two synthetic scFv libraries expressed on the surface of yeast, identifying both individual amino acids and motifs within the complementarity determining regions (CDRs) which contribute to the phenotype. We find enrichment of glycine, valine, and arginine as both individual amino acids and as a part of motifs, and additionally enrichment of motifs containing tryptophan. Insertion of any of these motifs into the CDR H3 of a “clean” antibody increased its nonspecificity, with greatest increases in antibodies containing Trp or Val motifs. We next applied these rules to the creation of a synthetic diversity library based on natural frameworks with significantly decreased incorporation of such motifs and demonstrated its ability to isolate binders to a wide panel of antigens. This work both provides a greater understanding of the drivers of nonspecificity and provides design rules to increase efficiency in the isolation of antibodies with drug-like properties.

Graphical Abstract



*Correspondence to: K. Dane Wittrup; wittrup@mit.edu.

Keywords

Nonspecificity; monoclonal antibody; synthetic scFv library; complementarity determining region; cross-interaction; polyreactivity

Introduction

Improved techniques have allowed for the rapid isolation of candidate monoclonal antibodies (mAbs) against a wide range of antigens, although significant challenges remain in ensuring candidate clones are suitable for clinical development. In addition to on-target binding, mAbs must display desirable biophysical characteristics, including resistance to aggregation and nonspecificity. Deficiencies in either can lead to undesirable effects including off target effects and accelerated clearance rates.[1–6] Developability is naturally selected for during the affinity maturation process in vivo,[7] and attempts to improve the in vitro selection process have focused on improved library designs[8–12] and better early stage characterization of biophysical characteristics.[3,13–17]

Many library designs have focused on mimicking the natural antibody repertoire, selecting for optimal VH/VL pairing[11] and frequencies of amino acids in the various complementarity determining regions (CDRs),[10] while other synthetic designs have favored use of a minimal set of amino acids for antigen recognition.[18–21] Some of these minimal designs have allowed for systematic evaluation of the contribution of certain amino acids to nonspecificity, observing increased cross-interactions in libraries solely consisting of Arg/Ser and to a lesser extent Trp/Ser and Tyr/Gly, as well as a general increase in nonspecificity with increased Arg levels within CDR H3. [19,22] The contribution of positive charge to nonspecificity has been similarly implicated in other antibody case studies,[23,24] while Trp is usually implicated as a significant driver of aggregation due to its hydrophobicity.[5,25,26]

Early assessment of nonspecificity is valuable to clinical development, allowing early elimination of potentially poor candidates. Multiple assays exist to detect this off-target binding, with many traditional assays utilizing binding to multiple noncognate antigens or baculovirus particles in an ELISA assay as an early stage readout.[19,27,28] We have previously reported the use of a polyspecificity reagent (PSR) in a binding assay on the surface of yeast with similar predictive power, correlating to pharmacokinetics in mice.[3,14]

In this study we utilize two synthetic yeast surface displayed single chain variable fragment (scFv) libraries to identify common features and motifs which can drive nonspecificity. Libraries were subjected to a modified version of the PSR assay to isolate nonspecific clones, revealing a predominance of motifs containing Trp, Val, and Arg. We next applied this information to the design of a new synthetic antibody library, from which we readily isolated multiple clones against a panel of antigens.

Results

Sorting a synthetic library reveals prevalence of Trp in CDR H3 of nonspecific clones

As an initial screen to isolate a panel of nonspecific antibodies, we sorted a synthetic antibody library from the Sidhu laboratory that is based off of natural frameworks, library 'G'.[29] The library was screened iteratively against four nonspecificity reagents: a preparation of soluble cytosolic or membrane proteins (SCP or SMP) from human embryonic kidney (HEK293) or Spodoptera frugiperda (Sf9) cells. These reagents were chosen to maximize orthogonality between preparations and avoid isolation of specific binders to components in one particular reagent. Populations were positively or negatively selected for binding against the four reagents sequentially, trying to maximize round to round differences in selection reagent. After four rounds of selection, the resulting populations were validated by assessing binding to all four reagents (Supplementary Figure 1A). As a secondary validation, we measured binding to polyclonal human IgG, an antigen not present in any of the PSR reagents (Supplementary Figure 1B). Sequencing of a sampling of 96 clones revealed six unique clones, whose CDR sequences are enumerated in Table 1.

Immediately evident was the increase of Trp residues in the H3 sequences of four of the clones over the expected 5% abundance in the library (Table 2). To assess the relative contribution of these Trp containing loops, we individually amplified each of the six CDRs from these nonspecific clones as well as six PSR negative clones and randomly recombined them into a new library of chimeric antibodies. This small library was again screened for nonspecificity. High throughput sequencing of the output library revealed an enrichment of the H3 sequences containing multiple Trp residues (Supplementary Figure 1), indicating their likely role in driving the nonspecificity in these clones. In clones without Trp in their H3, there was limited enrichment of any single CDR, indicating a combined role of multiple CDRs in determining the phenotype.

Constructing a tryptophan rich library based on the natural diversity

While our initial screening pointed to a prominent role of Trp in the H3 loop, we wanted to refine this understanding from a more diverse panel of nonspecific antibodies, and decided to construct a library rich in Trp residues in the H3 region. We constructed the library with a minimal number of antibody scaffolds, with most of the diversity focused in the H3 region to allow for a focused analysis of this region with minimal interference from other CDRs. Using previous knowledge from the natural antibody repertoire and similar libraries constructed in phage,[10,11,30] we selected five VH segments (VH1-69, VH3-15, VH3-23, VH4-39, VH5-51) and three VL segments (VK1-39, VK3-20, Vλ1-40). All sequences are found in Supplementary Data, and junctional diversity in the VL segments was limited to the most common residue for the given segment. All VH segments were equally represented in the final design, while the VL segments were designed to be at a 2:2:1 ratio, with the two kappa segments twice as likely to appear as the lambda segment. Diversity in the H3 region was created via trinucleotide synthesis using defined ratios of amino acids based on natural repertoire frequencies,[31] with the exception of the increase in Trp frequency from ~3% to 15% (Table 2). To simplify analysis for this library design, the length of the H3 was

restricted to eleven amino acids, the most common length found in the natural repertoire.[31] The segments were assembled in a VL-VH scFv format and transformed into a 10^8 member library in yeast display format and sequence verified prior to experiments.

Nonspecificity sorting of a Trp rich library

The Trp rich library was screened against four polyspecificity reagents as before to isolate a positive, nonspecific binding population (Figure 1). Sequencing of this population revealed 860 unique clones out of 5748 clones sequenced, which were analyzed against 3808 unique clones out of the 3814 sequences from the naïve, unsorted library.

As intended, the three VL segments had close to a 2:2:1 distribution in the naïve library, while the nonspecific population saw a modest enrichment of the VK1-39 segment (Supplementary Figure 2A). Although VH segments were intended have an even distribution in the library, VH families were highly skewed toward the VH1-69 segment, likely due to PCR biases during construction. A modest enrichment of VH1-69 segment was additionally observed in the sorted population (Supplementary Figure 2B).

Next, overall amino acid frequencies in the H3 were compared for the sorted and naïve populations against expected frequencies. Residues including Arg, Gly, and Val were enriched over the naïve library (Figure 2A, p-value < 0.001). Amino acid enrichment was further analyzed by clustering based on enrichment as a function of prevalence of a given amino acid in the H3 sequence. Using a k-means clustering approach, amino acids were best clustered into four groups, broadly categorized as those more enriched in nonspecific sequences with increased frequency, those which have little effect, those which show a modest de-enrichment at higher levels, and Trp. The positively charged amino acids Arg and to a lesser extent Lys, as well as Gly and Val saw greater enrichment as occurrence in the H3 increased. Trp is displayed with these residues, though it was classified in its own unique cluster and was de-enriched for 1 or 2 occurrences and only enriched when there were 4 or more copies present (Figure 2B). Most small aliphatic amino acids showed no enrichment over the naïve population (Figure 2C) whereas many of the charged (Glu, His) or hydrophilic (Ser, Thr) amino acids were modestly de-enriched. Of some interest, the bulkier amino acids Phe and Tyr were also included in this third group, with Tyr as the most highly de-enriched amino acid overall.

Tryptophan, Glycine and Valine are enriched at the center four positions

Analysis of the nonspecific population by sequence logos revealed a high frequency of Trp, Gly, and Val in the center four positions of the H3 (Figure 3A). Based on these results, we further analyzed the center four positions by comparing the amino acid frequencies of the naïve and nonspecific populations in those positions. This comparison showed a significant enrichment of Trp, Gly, and Val in the nonspecific population (Figure 3B, p value < 0.001). To see if these enrichments could be linked to the center four positions in the H3 specifically, we compared the enrichment of amino acids in those positions to overall enrichment in the H3. As expected, most residues were represented at similar or lower frequencies in the center compared to overall frequency in the H3. One notable de-enrichment was Arg, which was only enriched along the edges of the H3. In contrast,

enrichment of Trp, Gly, and Val was greater in the center four positions compared to enrichment in the H3 overall (Figure 3C). Looking at enrichment of these residues as a function of prevalence in the center four positions, Trp enrichment followed a similar trend as that seen in the H3 overall, showing de-enrichment for 1 or 2 occurrences and enrichment for 3 or more copies. Gly and Val again saw greater enrichment as occurrence in the center four positions increased (Figure 3D).

Tryptophan, Glycine, and Valine motifs are enriched in nonspecific antibodies

Analysis of individual amino acid enrichments suggested a strong role for motifs of multiple amino acids, and as such we next applied two motif discovery tools, DREME[32] and a hidden Markov model. In both cases, relative frequency of motifs in the sorted population was compared against the background of the naïve population. The hidden Markov model identified eight significantly enriched motifs (adjusted p-value < 0.01): GG, GGG, RR, VG, VV, VVV, WW, and WWW, along with other variants including these motifs (e.g. GWW, LVV). The DREME algorithm revealed similar enrichments, with significant (E-value < 0.05) presence of the motifs (in order of significance) WWW, VV, GGG, RR, QRWW, and VG. Expanding the search to include motifs with one residue of separation, additionally the WXW motif (p-value <0.01) was significantly enriched above background, while motifs containing other amino acids were not.

We next wanted to confirm whether these motifs could dominantly drive nonspecificity by splicing these motifs across the H3 sequence of an otherwise PSR negative antibody. Eight of the described motifs were inserted at each position along the H3 and resulting clones were subjected to binding to both the HEK SCP and HEK SMP. In addition to the eight identified motifs, we constructed a library containing the motif YY, as Tyr was modestly de-enriched at higher levels in analysis of the sorted output. Insertion of all the tested motifs at all positions increased binding to the SCP reagent (Figure 4A). Binding to the SMP reagent (Figure 4B), was less affected, with the most significant increase in binding coming from clones with VV, VVV, WW, or WWW insertions. Of some interest, the clone with highest binding in all cases contained the motif at or near the center of the H3 sequence, in accordance with the enrichments found at the center four amino acids in the original sort. Of note, the YY motifincreased nonspecific binding to a similar degree as the Gly and Arg containing motifs, which may point to some underlying instability in the scFv caused by mutation.

Isolating binders from a library lacking tryptophan

With the prevalence of Trp containing motifs in nonspecific clones, we next decided to explore the consequences of constructing a library free of Trp. Using the same design as the Trp rich library, we constructed a similar library instead lacking any Trp in the H3 sequence (Table 2). We subsequently subjected the library to antigen selections via yeast surface display in order to probe whether knockout of tryptophan would compromise antigen binding capacity. The Trp negative library was sorted against a panel of seven antigens: FcIL2, PDGFR-β, TA99, Lysozyme, sso7d, KRas, and TNFα. We were able to isolate scFvs against each antigen, and affinity titrations were conducted on selected individual clones for each antigen (Figure 5). For every antigen, binders with nanomolar binding affinity were

successfully isolated from the Trp negative library, suggesting that antigen binding is not severely compromised upon knockout of Trp in the H3.

Construction of a new library which minimizes cross-interaction prone motifs

Based on the knowledge of nonspecificity driving motifs, we next constructed a new general purpose library which minimizes these motifs. The overall structure of the library matches the Trp rich library, using five VH and three VL frameworks with the majority of diversity focused on the H3 segment. Key differences include eliminating Trp and significantly dropping the frequency of Arg and Val (Table 2). We additionally allowed for loop length diversity which mimics the natural repertoire, and ensured a library size of at least one billion members. Sequence validation of the library reveals an amino acid and length distribution which matches the intended design (Fig 6A & B). This library was next screened for scFvs against a wide panel of antigens to demonstrate library robustness. We were able to isolate binding clones to all antigens tested, and measured the affinity of three clones for each campaign. The library was frequently able to isolate multiple clones of nanomolar affinity (Fig 6C), suggesting high functional utility of the library. Additionally, we tested PSR binding of all isolated clones, finding a low incidence of PSR positive clones (Fig 6D). Of some interest, there was antigen dependence on polyspecificity, with higher nonspecific binding mostly focused to binders of the small, charged cytokine IFNα. Comparing these PSR scores against a large data set of clinical candidate and approved antibodies,[16] we find that our clones are significantly less (p-value <0.001) nonspecific than clones from phage derived sources and comparable to those from non-phage sources (Supplementary Figure 3).

Discussion

The initial goal of this study was to elucidate some causes of nonspecificity and apply the findings towards an improved library design. Towards this goal, here we have identified the enrichment of four amino acids in the CDR H3 of nonspecific antibodies: Gly, Val, Trp, and Arg. While individual amino acid enrichments were apparent in three of these instances, effects of Trp were most evident in motifs. Applying this knowledge towards generation of a new library additionally yielded robust binding to a wide variety of antigens with low levels of nonspecificity.

Comparing our findings to previous work in the field, we find similar results with regards to enriched amino acids. Glycine, a small residue, is likely contributing to increased flexibility of the H3 loop, which may allow for greater plasticity and broader antigen recognition. Insertion of glycine motifs into a clean scaffold only modestly increased nonspecific binding, which suggests that the context in which the motif exists may have an impact on whether or not it can drive the phenotype. Glycine is the most common amino acid in CDR H3 loops of natural antibodies, and is likely tolerated by deletion of potentially harmful contexts. While not commonly linked with nonspecificity, increased binding associated with Glycine was previously seen in binary amino acid libraries consisting solely of Gly/Tyr residues.[22]

The enrichment of valine was more novel, with no previous studies to our knowledge identifying it as a potential liability in antibody CDRs. This hydrophobic, branched chain amino acid was significantly enriched individually and had a significant impact as a motif of two or three consecutive residues. Valine is frequently found in the core of proteins, and negatively contributes to the free energy of folding if found on the surface.[33–35] While buried in a protein-protein interaction, valine would instead positively contribute, which may explain its proclivity to drive nonspecific interactions with other proteins. Interestingly, this phenotype was unique to valine, and was not seen with the similar aliphatic amino acids leucine and isoleucine.

Enrichment of positive charge, seen in this study with the enrichment of arginine as an individual amino acid and as a part of motifs, has been previously reported to drive accelerated clearance of clinical antibodies, indicative of nonspecificity.[2,23,24,36,37] Additionally, in limited diversity libraries, increased charge has been previously shown to drive nonspecific binding.[19]

In our initial screens tryptophan was found to be a major contributor to nonspecificity, and we expected that constructing a library with five-fold more Trp residues might lead to even higher enrichment of this amino acid in sorted pools. Somewhat surprisingly, as an individual amino acid there was no enrichment, however when looking at motifs containing Trp there was significant enrichment. This points to the importance of hydrophobic patches on the antibody surface, consistent with previously described aggregation prone regions.[5,26,38] The inclusion of Trp in our data set also suggests overlap between the phenotypes of nonspecificity and aggregation, which may be linked by the preference of hydrophobic regions to be buried rather than surface exposed. Other studies analyzing preclinical properties of panels of antibodies have shown similar overlaps, with a significant number of nonspecific clones also displaying signs of self-association.[3,16,39]

Tryptophan is commonly found in antibody paratopes,[40] and in eliminating tryptophan from our library design, we were initially concerned there may be some cost to the ability to isolate antigen specific antibodies. In both the test library and final design this was not the case, as we were able to isolate panels of binders against a wide variety of antigens. While it is beyond the scope of the current work, it remains to be seen if there are more subtle sacrifices in this design decision, such as a decrease in epitopic coverage.

Taken in whole, we have presented nonspecificity motifs which can be applied to the antibody development process. It is important to note that all the current work was completed in the scFv format, and while we expect these findings translate to the full IgG format, this could be worth further evaluation in future studies. Additionally, here we have focused on CDR H3, which, while it can often be a major driver of nonspecificity, is not the main driver in all cases. In sum, the motifs and enrichments described here can account for no more than 50% of the nonspecific clones isolated. To close this gap and expand knowledge of nonspecificity, structural modeling or more sophisticated machine learning algorithms will be necessary to identify the more subtle causes.

Materials and Methods

Antigen Preparation

Soluble membrane preparations (SMP) were prepared as previously described.[14] Briefly, one billion HEK or Sf9 cells were pelleted, washed and homogenized. The homogenate was separated into membrane and cytosolic fractions via centrifugation at 40,000 x *g* for 1h at 4° C. The supernatant was collected as the soluble cytosolic preparation (SCP) and the pellet was further washed and solubilized to create the enriched membrane fraction (EMF). Both the SCP and EMF were diluted down to a concentration of 1mg/mL and biotinylated using NHS-LC-Biotin (Pierce, Thermo Fisher Cat# 21336). Biotin solution was prepared according to manufacturer's protocol and 20μL 10mM biotin reagent was added per 1mg sample. The reaction mixture was incubated for 3h at 4° C with gentle agitation, after which the biotinylated EMF (b-EMF) was pelleted at 40,000 x *g* for 1h at 4° C and the biotinylated SCP (b-SCP) was buffer exchanged into phosphate buffered saline, pH 7.4 (PBS; Corning) using a Zeba Spin Desalting column (Thermo, 7K MWCO). The b-EMF was subsequently solubilized overnight using a buffer containing 1% n-dodecyl-b-D-maltopyranoside (DDM) and then centrifuged at 40,000 x *g* for 1h at 4° C. The resulting supernatant was collected as the final, biotinylated SMP (b-SMP) product.

Other antigens were obtained as follows: hen egg lysozyme (Sigma) and human TNF-α (Gold Bio) were purchased, bFc-IL2, TA99, and EGFR were produced in HEK cells as previously described,[29] and sso7d, TF, and K-Ras were expressed in E. coli as 6x His-Sumo fusions and purified using TALON resin (Clontech).

Library Construction

Libraries of scFvs were constructed in a VL-VH format via two-piece homologous recombination in yeast. The insert for the H3 diversity libraries was constructed in four segments as follows: the H3 piece was synthesized as a primer using trinucleotide synthesis using a defined ratio of amino acids for H3 positions 95–100H (Ella Biotech); the VH and VL sequences, as obtained from IMGT, as well as the FR4 region were synthesized as GeneBlock fragments (IDT). Full sequences for synthesized segments are found in Supplementary data and amino acid compositions for each constructed library are detailed in Table 2. For both the Trp Rich and Trp Free libraries H3 length was restricted 11 amino acids, while in the Final H3 library length varied from 6–17 as defined by Kabat numbering, with a distribution exhibited in Figure 6B.

Sequences coding for the overlap region with the pCTCON2 vector as well as the scFv linker were appended as appropriate to segment designs, and each segment contained 20bp overlap with its neighboring piece to allow for splicing via PCR (See Supplementary Data for overlap sequences). First, H3 library primer segments were appended to the FR4 construct with pCTCON2Rev primer. The VH segment was subsequently added, followed by addition of the VL segment and subsequent amplification using the pCTCON2Fwd and Rev primers. The pCTCON2 vector was prepared for homologous recombination by digestion overnight with SalI-HF followed by digestion overnight with NheI-HF and BamI-HF (New England Biolabs). The scFv library inserts and the digested vector were

transformed into yeast using previously described electroporation methods.[41,42] The Trp Free and Trp Rich libraries were constructed to a size of 10^8 while the Final H3 library contained 10^9 members. Library quality was assessed by SMRT sequencing of a sample of each library.

Yeast display experiments

All yeast experiments were conducted in the RJY100 strain.[29] Nonspecificity sorting experiments were conducted using the synthetic library 'G'[29] or the Trp Rich scFv library, and sorts for specific binders were conducted using the Trp Free and Final H3 libraries using previously described yeast display techniques.[41,42] Initial rounds of sorting were conducted using Dynabeads biotin binder beads (Life Technologies), ensuring at least ten fold coverage of the library or population size in each round. After two rounds, subsequent sorting was completed on a FACS Aria 3 (BD). As there was concern over selection of scFvs specific to a single protein for nonspecificity sorting, the order of antigens was chosen to maximize round to round differences. In order, the population was positively screened against: HEK bSMP, Sf9 bSCP, Sf9 bSMP, HEK bSCP. To ensure no enrichment of binders to biotin, 1mM biotin (Sigma) was added to sort rounds three and four. In all instances, nonspecificity was evaluated using a 1:10 dilution of each biotinylated nonspecificity reagent as primary and streptavidin AlexaFluor 488 conjugate (1:100, Life Technologies) as secondary. Successful scFv display was assessed using chicken anti-cMyc primary (1:250, Gallus) followed by goat anti-chicken AlexaFluor647 secondary (1:100, Life Technologies).

All sorts against specific antigens were conducted using two rounds of bead sorting followed by two to three rounds of sorting via FACS. Successful antibody display was assessed as above, and all antigens were biotinylated to allow for detection via streptavidin AlexaFluor 488 secondary (1:100, Life Technologies). For single clones, affinity was assessed on the yeast surface using previously described methods,[29] utilizing the iQue Screener PLUS (intellicyt) for all flow cytometry analysis.

Loop Swapping library construction

Twelve total clones from positively and negatively sorted populations were selected, and are listed in Table 1. As the framework was identical for all clones, a single set of primers was designed to amplify each of the individual CDR regions. Primers were designed to overlap at a region of approximately 20bp to allow for homologous recombination between segments (Supplementary Data). Due to their proximity in sequence space, there was not a suitable primer location in between CDR H1 and H2, and as such these were amplified as a pair. Once individual segments were amplified, the pool of five individual insert pieces, along with the digested pCTCON2 backbone was transformed into RJY100 yeast using previously reported electroporation protocols.[41] Insert pieces were proportionately added based on their size, with the total amount of insert DNA equal to 4ug for each transformation. The library was constructed to a size of 1.2×10^7, which is four-fold coverage of the maximum theoretical library size (3.0×10^6). Sequence verification via Sanger sequencing of 96 clones revealed a diverse library, with a significant number of crossover events between H1 and H2 segments, likely due to homologous recombination during the transformation process.

Motif Library Construction

All H3 motif mutants were constructed via homologous recombination of three DNA fragments in yeast. Designed overlap sequences were located at the two ends of the scFv construct as well as directly in the H3 region. For the fragment prior to and containing the H3, each motif was inserted once at each position down the length of the H3. For each motif, all positional variants of the H3-containing fragment were amplified with the pCTCON2 RecombFwd and appropriate PreH3 primers, and the fragment after and also containing the H3 was amplified using the pCTCON2 RecombRev and proper PostH3 primers (Primer sequences in Supplementary Data). The pCTCON2 vector was prepared for homologous recombination by digestion with SalI-HF followed by digestion with NheI-HF and BamI-HF (New England Biolabs). The two scFv fragments along with the digested vector were co-transformed into chemically competent yeast using the Frozen-EZ Yeast Transformation II Kit (Zymo Research) and resultant clones were sequence verified prior to experiments.

Data Analysis

Sequence analysis was conducted using Matlab. Motifs were identified using the DREME program[32] or a custom Matlab code. Statistical significance of motif enrichments was conducted by comparing differential enrichment of motifs against enrichment in a background of randomized sequences with identical composition. Motifs were deemed to be significant if the margin was greater than all of 100 simulations of randomized motif identification, implying a p-value < 0.01.

Sequencing

Libraries were sequenced using SMRT sequencing (Pacific Biosciences). Individual clones for motif and sort validation were sequenced using Sanger sequencing (Macrogen). Library inserts were bar code amplified by PCR using primers detailed in Supplementary Data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Wu H, Pfarr DS, Tang Y, An L-L, Patel NK, Watkins JD, Huse WD, Kiener Pa, Young JF. Ultra-potent antibodies against respiratory syncytial virus: effects of binding kinetics and binding valence on viral neutralization. J Mol Biol [Internet]. 2005; 350:126–44. [cited 2013 Feb 13]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15907931.
2. Datta-Mannan A, Lu J, Witcher DR, Leung D, Tang Y, Wroblewski VJ. The interplay of non-specific binding, target-mediated clearance and FcRn interactions on the pharmacokinetics of humanized antibodies. MAbs [Internet]. 2015; 7:1084–93. Available from: http://www.tandfonline.com/doi/full/10.1080/19420862.2015.1075109.
3. Kelly RL, Sun T, Jain T, Caffry I, Yu Y, Cao Y, Lynaugh H, Brown M, Vásquez M, Wittrup KD, et al. High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. MAbs [Internet]. 2015; 7:00–00. Available from: http://www.tandfonline.com/doi/full/10.1080/19420862.2015.1043503.
4. Kelly RL, Yu Y, Sun T, Caffry I, Lynaugh H, Brown M, Jain T, Xu Y, Wittrup KD. Target-independent variable region mediated effects on antibody clearance can be FcRn independent.

MAbs [Internet]. 2016; 8:1269–75. Available from: http://dx.doi.org/10.1080/19420862.2016.1208330.

5. Agrawal NJ, Kumar S, Wang X, Helk B, Singh SK, Trout BL. Aggregation in Protein-Based Biotherapeutics: Computational Studies and Tools to Identify Aggregation-Prone Regions. J Pharm Sci [Internet]. 2011; 100:5081–95. [cited 2013 Jan 24]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/jps.22705/full.
6. Sigounas G, Harindranath N, Donadel G, Notkins AL. Half-life of polyreactive antibodies. J Clin Immunol [Internet]. 1994; 14:134–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8195315.
7. Wang, F., Sen, S., Zhang, Y., Ahmad, I., Zhu, X., Wilson, Ia, Smider, VV., Magliery, TJ., Schultz, PG. Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. Proc Natl Acad Sci U S A [Internet]. 2013. [cited 2013 Feb 26] Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1301810110
8. Knappik, a, Ge, L., Honegger, a, Pack, P., Fischer, M., Wellnhofer, G., Hoess, a, Wölle, J., Plückthun, a, Virnekäs, B. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J Mol Biol. 2000; 296:57–86. [PubMed: 10656818]
9. Rothe C, Urlinger S, Löhning C, Prassler J, Stark Y, Jäger U, Hubner B, Bardroff M, Pradel I, Boss M, et al. The Human Combinatorial Antibody Library HuCAL GOLD Combines Diversification of All Six CDRs According to the Natural Immune System with a Novel Display Method for Efficient Selection of High-Affinity Antibodies. J Mol Biol [Internet]. 2008; 376:1182–200. Available from: http://dx.doi.org/10.1016/j.jmb.2007.12.018.
10. Prassler J, Thiel S, Pracht C, Polzer A, Peters S, Bauer M, Nörenberg S, Stark Y, Kölln J, Popp A, et al. HuCAL PLATINUM, a synthetic Fab library optimized for sequence diversity and superior performance in mammalian expression systems. J Mol Biol [Internet]. 2011; 413:261–78. [cited 2014 Sep 29]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21856311.
11. Tiller T, Schuster I, Deppe D, Siegers K, Strohner R, Herrmann T, Berenguer M, Poujol D, Stehle J, Stark Y, et al. A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. MAbs [Internet]. 2013; 5:1–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23571156.
12. Sidhu SS, Fellouse FA. Synthetic therapeutic antibodies. Nat Chem Biol [Internet]. 2006; 2:682–8. [cited 2013 Jan 10]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17108986.
13. Sule SV, Dickinson CD, Lu J, Chow C-K, Tessier PM. Rapid analysis of antibody self-association in complex mixtures using immunogold conjugates. Mol Pharm [Internet]. 2013; 10:1322–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23383873.
14. Xu Y, Roach W, Sun T, Jain T, Prinz B, Yu T-Y, Torrey J, Thomas J, Bobrowicz P, Vásquez M, et al. Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. Protein Eng Des Sel [Internet]. 2013; 26:663–70. [cited 2014 Jan 21]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24046438.
15. Jayaraman J, Wu J, Brunelle MC, Cruz AMM, Goldberg DS, Lobo B, Shah A, Tessier PM. Plasmonic measurements of monoclonal antibody self-association using self-interaction nanoparticle spectroscopy. Biotechnol Bioeng [Internet]. 2014; 111:1513–20. [cited 2014 Aug 12]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25367343.
16. Jain, T., Sun, T., Durand, S., Hall, A., Houston, NR., Nett, JH., Sharkey, B., Bobrowicz, B., Caffry, I., Yu, Y., et al. Biophysical properties of the clinical-stage antibody landscape. Proc Natl Acad Sci [Internet]. 2017. 201616408. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1616408114
17. He, F., Razinkov, VI., Middaugh, CR., Becker, GW. High-Throughput Biophysical Approaches to Therapeutic Protein Development [Internet]. In: Narhi, LO., editor. Biophysics for Therapeutic Protein Development. New York, NY: Springer New York; 2013. [cited 2013 Mar 20]. page 7–31.Available from: http://link.springer.com/10.1007/978-1-4614-4316-2
18. Fellouse, Fa, Wiesmann, C., Sidhu, SS. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. Proc Natl Acad Sci U S A [Internet]. 2004; 101:12467–72. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=515084&tool=pmcentrez&rendertype=abstract.

19. Birtalan S, Zhang Y, Fellouse Fa, Shao L, Schaefer G, Sidhu SS. The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. J Mol Biol [Internet]. 2008; 377:1518–28. [cited 2013 Feb 12]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18336836.
20. Fellouse FA, Esaki K, Birtalan S, Raptis D, Cancasci VJ, Koide A, Jhurani P, Vasser M, Wiesmann C, Kossiakoff Aa, et al. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. J Mol Biol [Internet]. 2007; 373:924–40. [cited 2012 Nov 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17825836.
21. Persson H, Ye W, Wernimont A, Adams JJ, Koide A, Koide S, Lam R, Sidhu SS. CDR-H3 Diversity Is Not Required for Antigen Recognition by Synthetic Antibodies. J Mol Biol [Internet]. 2013; 425:803–11. [cited 2013 Feb 11]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23219464.
22. Birtalan S, Fisher RD, Sidhu SS. The functional capacity of the natural amino acids for molecular recognition. Mol Biosyst [Internet]. 2010; 6:1186–94. [cited 2013 Jan 15]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20383388.
23. Datta-Mannan A, Thangaraju A, Leung D, Tang Y, Witcher DR, Lu J, Wroblewski VJ, Derrick R, Lu J, Wroblewski VJ. Balancing charge in the complementarity-determining regions of humanized mAbs without affecting pI reduces non-specific binding and improves the pharmacokinetics. MAbs [Internet]. 2015; 7:483–93. Available from: http://www.tandfonline.com/doi/full/10.1080/19420862.2015.1016696.
24. Igawa T, Tsunoda H, Tachibana T, Maeda a, Mimoto F, Moriyama C, Nanami M, Sekimori Y, Nabuchi Y, Aso Y, et al. Reduced elimination of IgG antibodies by engineering the variable region. Protein Eng Des Sel [Internet]. 2010; 23:385–92. [cited 2014 Aug 29]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20159773.
25. Voynov V, Chennamsetty N, Kayser V, Helk B, Trout BL. Predictive tools for stabilization of therapeutic proteins. MAbs [Internet]. 2009; 1:580–2. [cited 2013 Jan 29]. Available from: http://www.landesbioscience.com/journals/22/article/9773/.
26. Lauer TM, Agrawal NJ, Chennamsetty N, Egodage K, Helk B, Trout BL. Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. J Pharm Sci [Internet]. 2012; 101:102–15. [cited 2013 Jan 29]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/jps.22758/full.
27. Frese, K., Eisenmann, M., Ostendorp, R. An automated immunoassay for early specificity profiling of antibodies… [Internet]; 2013. p. 1-9.[cited 2013 Feb 25]Available from: http://www.landesbioscience.com/journals/mabs/article/23539/
28. Hotzel I, Theil F-P, Bernstein LJ, Prabhu S, Deng R, Quintana L, Lutman J, Sibia R, Chan P, Bumbaca D, et al. A strategy for risk mitigation of antibodies with fast clearance. MAbs [Internet]. 2012; 4:753–60. [cited 2013 Jan 24]. Available from: http://migrate.landesbioscience.com/journals/mabs/article/22189/?show_full_text=true.
29. Van Deventer, JA., Kelly, RL., Rajan, S., Wittrup, KD., Sidhu, SS. A switchable yeast display/secretion system; Protein Eng Des Sel [Internet]. 2015. p. 1-9.Available from: http://www.ncbi.nlm.nih.gov/pubmed/26333274
30. Jayaram N, Bhowmick P, Martin ACR. Germline VH/VL pairing in antibodies. Protein Eng Des Sel [Internet]. 2012; 25:523–9. [cited 2014 Nov 17]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22802295.
31. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L, Sundar PD, Day GMR, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci U S A [Internet]. 2009; 106:20216–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2787155&tool=pmcentrez&rendertype=abstract.
32. Bailey TL. DREME: Motif discovery in transcription factor ChIP-seq data. Bioinformatics. 2011; 27:1653–9. [PubMed: 21543442]
33. Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. Protein Sci. 2003; 12:1406–17. [PubMed: 12824487]
34. Kauzmann W. Some Factors in the Interpretation of Protein Denaturation. Adv Protein Chem. 1959; 14:1–63. [PubMed: 14404936]

35. Schulz, GE., Schirmer, RH. Principles of Protein Structure [Internet]. New York, NY: Springer New York; 1979. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Springer+Advanced+Texts+in+Chemistry#0%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Springer+advanced+texts+in+chemistry%230

36. Boswell CA, Tesar DB, Mukhyala K, Theil FP, Fielder PJ, Khawli La. Effects of charge on antibody tissue distribution and pharmacokinetics. Bioconjug Chem. 2010; 21:2153–63. [PubMed: 21053952]

37. Hong G, Chappey O, Niel E, Scherrmann J-M. Enhanced Cellular Uptake and Transport of Polyclonal Immunoglobulin G and Fab After Their Cationization. J Drug Target [Internet]. 2000; 8:67–77. Available from: http://informahealthcare.com/doi/abs/10.3109/10611860008996853.

38. Courtois, F., Schneider, CP., Agrawal, NJ., Trout, BL. Rational Design of Biobetters with Enhanced Stability. J Pharm Sci [Internet]. 2015. :n/a–n/a. Available from: http://doi.wiley.com/10.1002/jps.24520

39. Dobson CL, Devine PWA, Phillips JJ, Higazi DR, Lloyd C, Popovic B, Arnold J, Buchanan A, Lewis A, Goodman J, et al. Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. Sci Rep [Internet]. 2016; 6:38644. Available from: http://www.nature.com/articles/srep38644.

40. Robin G, Sato Y, Desplancq D, Rochel N, Weiss E, Martineau P. Restricted diversity of antigen binding residues of antibodies revealed by computational alanine scanning of 227 antibody-antigen complexes. J Mol Biol [Internet]. 2014; 426:3729–43. Available from: http://www.sciencedirect.com/science/article/pii/S0022283614004562.

41. VanDeventer, JA., Wittrup, KD. Yeast Surface Display for Antibody Isolation: Library Construction, Library Screening, and Affinity Maturation [Internet]. In: Ossipow, V., Fischer, N., editors. Monoclonal Antibodies. Totowa, NJ: Humana Press; 2014. p. 151-81.[cited 2014 Jul 10]Available from: http://link.springer.com/10.1007/978-1-62703-992-5

42. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nat Protoc [Internet]. 2006; 1:755–68. [cited 2012 Nov 1]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17406305.

Highlights

- Screened two libraries for nonspecific antibodies using a high throughput yeast display assay
- Identified motifs containing Trp, Val, Gly, and Arg as drivers of nonspecificity
- Created a new, Trp-free synthetic library based off natural frameworks which minimized these motifs
- Isolated high affinity binders to a wide panel of antigens

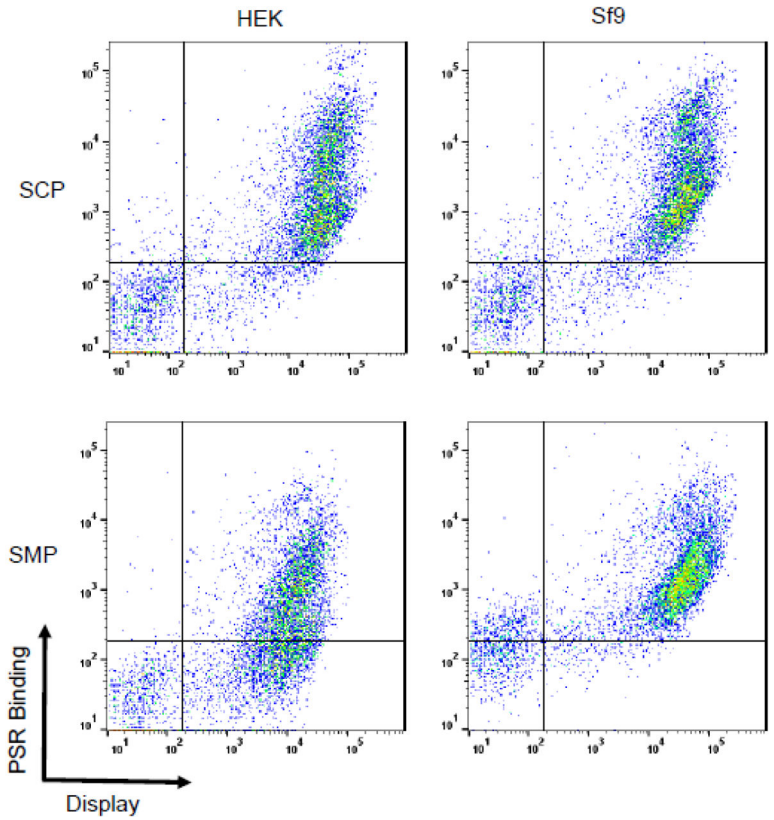


Figure 1. Sorting for nonspecificity

A nonspecific population was isolated from the Trp rich library and validated to bind to four polyspecificity reagents. Clones were evaluated for successful display via staining of the cMyc tag and for binding to PSR reagents derived from both soluble membrane preparations(SMP) and soluble cytoplasmic preparations (SCP). To ensure nonspecificity, reagents from both mammalian (HEK) and insect (Sf9) sources were used for evaluation.

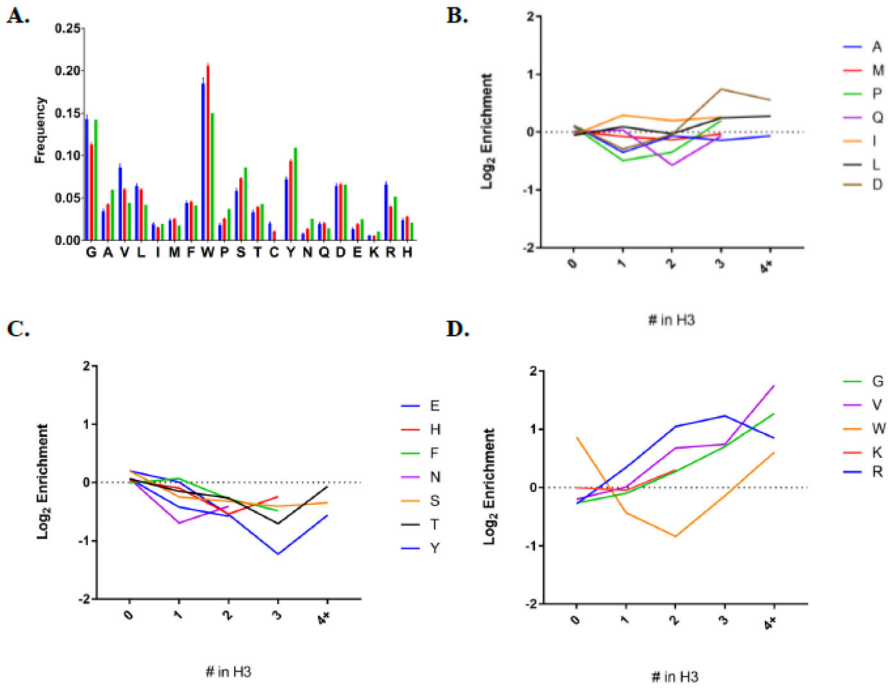


Figure 2. Enrichment of Arg, Val, Gly in nonspecific clones

Frequencies of individual amino acids in CDR H3 loops were evaluated for the theoretical (green), naïve (red), and sorted populations (blue) (A). For figures B–D, the log2 enrichments for the sorted population over the naïve population were calculated for sequences with 0, 1, 2, 3, and 4 or more amino acids in the H3 loop. Amino acids broadly fell into three categories, those that were not enriched (B), those that were modestly de-enriched (C), and those that were enriched (D) as the amino acid was more frequently found in the loop. Error bars represent standard error of the mean over 860 (sorted) or 3808 (naïve) clones.

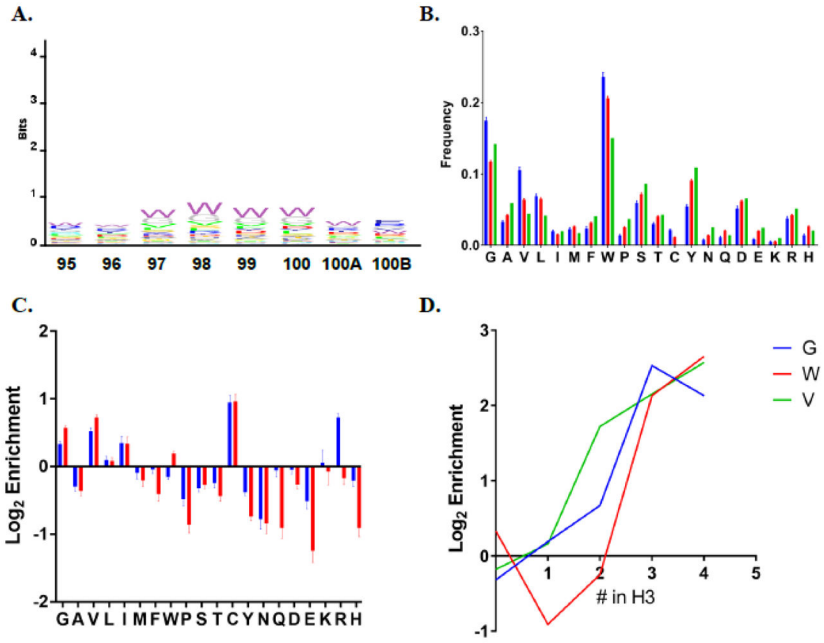


Figure 3. Enrichments of Val, Gly, and Trp are enhanced in the center four positions of the H3

Conserved amino acids of the H3 loop were analyzed by sequence logos for the sorted population (A). Residues are numbered using Kabat standard numbering. Individual amino acid frequencies in the center four positions of the H3 were calculated for the theoretical (green), naïve (red), and sorted populations (blue) (B). For the sorted population, the individual amino acid enrichments in the center four positions (red) were compared to the overall enrichments in the H3 (blue) (C). Log2 enrichment of Gly, Trp, and Val were calculated for sequences with 0, 1, 2, 3, and 4 amino acids in the center four positions for the sorted population over the naïve population (D).

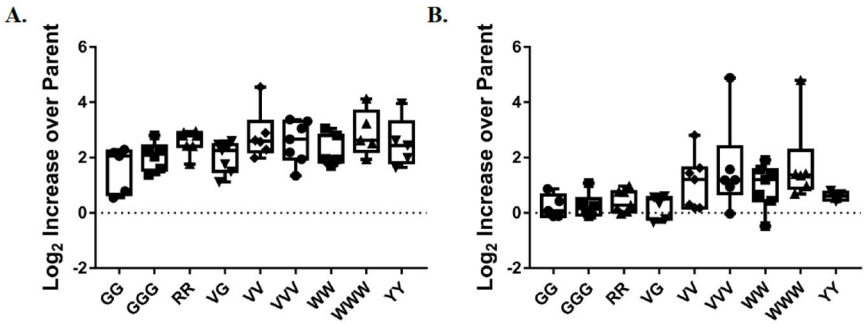


Figure 4. Motifs which drive nonspecificity

Putative motifs were inserted across the H3 sequence of one clean scFv sequence and PSR binding was measured against both a cytosolic (A) and membrane (B) preparation. For each motif depicted, the log2 enrichment over the parent clone for individual clones are displayed as dots; the median, lower quartile, upper quartile, and extrema for each collection of clones is represented in the box and whiskers plot.

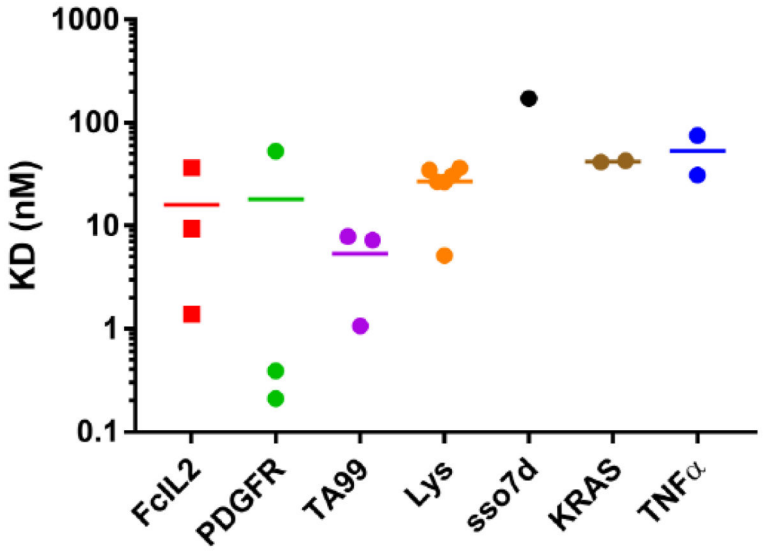


Figure 5. High affinity scFvs from a library lacking tryptophan

Affinities of scFvs isolated against a panel of antigens were evaluated on the surface of yeast. Points represent individual clones and bars represent the mean affinity for all antibodies isolated against a given antigen.

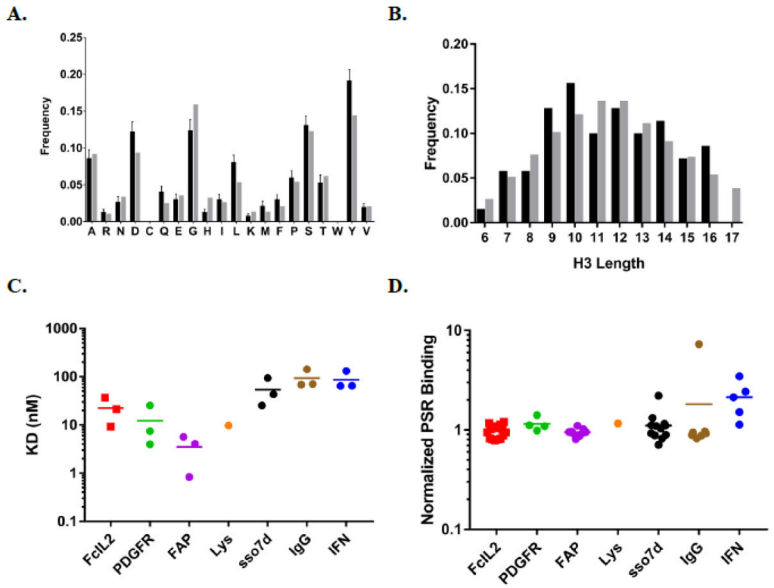


Figure 6. Construction of a new library with reduced nonspecificity

A new library was constructed which minimizes motifs which drive nonspecificity. The composition of this library (black bars) was analyzed by SMRT sequencing and compared against the theoretical library design (grey bars). Both the abundance of individual amino acids (A) and the length distribution of the H3 segment (B) closely match the intended design. Selections were conducted against a panel of antigens and affinities of resulting clones were assessed on the surface of yeast (C). Additionally, nonspecific binding of each clone was assessed and normalized against PSR binding of adalimumab, a clinical antibody with no measurable PSR reactivity.[16] Points represent individual clones and bars represent the mean affinity or PSR score of all clones.

Table 1

Sequences of Positively (P) and negatively (N) isolated clones from Library G.

Clone Name	CDR Sequences					
	VL			VH		
	L1	L2	L3	H1	H2	H3
N1	QSYSGSYV	YASSLY	QSSYSLIT	NISSGSGIH	ASIYPYYGSTGY	GAFYGYPYYSYAMD
N2	QSYYYV	GASYLY	QPSGLIT	NLSSGYSYMH	AGIYSYSGYTGY	TVRGSKKPYFSGWAMD
N3	QGSYSYV	SASSLY	QHGYPIT	NISYGGGMH	AGIYPSYGSTGY	FWYGSALD
N4	QYSSGV	SASYLY	QSVSYHGSLFT	NLGYSSMH	ASIGPYGGSTSY	YYGVSGWPGFD
N5	QSSYSYSV	GASSLY	QSSSVSAPIT	NIGYYGGMH	AGIYPGSSSTSY	SYYGGGWYFSAMD
N6	QYYGSSV	YASSLY	QSWYGLIT	NISYYYSSMH	ASIYPYGGSTSY	GYPAYVYYHAMD
P1	QGYGSSYV	GASSLY	QAYVPFT	NIYGSSGVH	AYIYPSYGYTSY	RSYYYGAYAMD
P2	QYYSYV	GASSLY	QSYSGLIT	NLGYGYGGMH	AYISSYYGYTYY	RSYHAMD
P3	QSYSGSV	SASSLY	QSSVLFT	NLYYGYMH	AGIGPYYGGTSY	RGYHYSWWFWSSSAMD
P4	QSYGGYV	YASYLY	QSSSVGSLFT	NISSSSYMH	ASISPYYSSTYY	RSASHSSYHWWYGMD
P5	QYYGSV	YASGLY	QYSGLIT	NLGGYSMH	AYIGPYSGYTSY	RSGSWWWSWYWYFAMD
P6	QGSGSSGV	SASSLY	QSSYSLIT	NIYGSSMH	ASISPYSGSTYY	RDYSWPWYWSYWHFHSYSGFYGLD

J Mol Biol. Author manuscript; available in PMC 2019 January 05.

Table 2

Amino Acid Frequencies for CDR H3

		Library composition (% frequency)				
		Library 'G'	**Natural repertoire**	**Trp Rich**	**Trp Free**	**Final**
A	Ala	10.0	7.8	6.0	7.0	9.1
C	Cys	0.0	0.8	0.0	0.0	0.0
D	Asp	0.0	7.9	6.6	7.7	9.3
E	Glu	0.0	3.0	2.5	2.9	3.5
F	Phe	5.0	1.7	4.1	4.8	2.0
G	Gly	20.0	13.5	14.2	16.6	15.9
H	His	5.0	2.7	2.1	2.4	3.2
I	Ile	0.0	2.2	2.0	2.3	2.6
K	Lys	0.0	1.1	1.0	1.2	1.3
L	Leu	0.0	4.5	4.2	4.9	5.3
M	Met	0.0	1.1	1.8	2.1	1.3
N	Asn	0.0	2.8	2.5	3.0	3.3
P	Pro	5.0	4.6	3.7	4.3	5.3
Q	Gln	0.0	2.1	1.4	1.7	2.4
R	Arg	0.0	6.3	5.2	6.0	1.0
S	Ser	20.0	10.4	8.6	10.1	12.2
T	Thr	0.0	5.2	4.3	5.0	6.1
V	Val	5.0	5.9	4.4	5.2	2.0
W	Trp	5.0	3.1	15.0	0.0	0.0
Y	Tyr	25.0	12.3	10.9	12.8	14.3