

Combining random mutagenesis, structure-guided design and next-generation sequencing to mitigate polyreactivity of an anti-IL-21R antibody

Sharon M. Campbell^a, Joseph DeBartolo^b, James R. Apgar^b, Lydia Mosyak^b, Virginie McManus^a, Sonia Beyer^a, Eric M. Bennett^b, Matthew Lambert^a, and Orla Cunningham^a 

^aBiomedicine Design, Pfizer, Dublin, Ireland; ^bBiomedicine Design, Pfizer, Cambridge, MA, USA

ABSTRACT

Despite substantial technological advances in antibody library and display platform development, the number of approved biotherapeutics from displayed libraries remains limited. *In vivo*, 20–50% of peripheral B cells undergo a process of receptor editing, which modifies the variable and junctional regions of light chains to delete auto-reactive clones. However, *in vitro* antibody evolution relies primarily on interaction with antigen, with no in-built checkpoints to ensure that the selected antibodies have not acquired additional specificities or biophysical liabilities during the optimization process. We had previously observed an enrichment of positive charge in the complementarity-determining regions of an anti-IL-21 R antibody during affinity optimization, which correlated with more potent IL-21 neutralization, but poor *in vivo* pharmacokinetics (PK). There is an emerging body of data that has correlated antibody nonspecificity with poor PK *in vivo*, and established a series of screening assays that are predictive of this behavior. In this study we revisit the challenge of developing an anti-IL-21 R antibody that can effectively compete with IL-21 for its highly negatively charged paratope while maintaining favorable biophysical properties. *In vitro* deselection methods that included an excess of negatively charged membrane preparations, or deoxyribonucleic acid, during phage selection of optimization libraries were unsuccessful in avoiding enrichment of highly charged, nonspecific antibody variants. However, a combination of structure-guided rational library design, next-generation sequencing of library outputs and application of linear regression models resulted in the identification of an antibody that maintained high affinity for IL-21 R and exhibited a desirable stability and biophysical profile.

ARTICLE HISTORY

Received 2 December 2020
Revised 18 January 2021
Accepted 26 January 2021

KEYWORDS

Antibody nonspecificity; developability; IL-21R; deselection; optimization



Introduction


Phage display remains a workhorse for the generation and optimization of biotherapeutics. It provides a versatile platform to represent in excess of 10^{10} unique, fully human, synthetic or immune repertoires for panning against conserved proteins, specific epitopes, particular protein conformations or protein complexes.^{1–4} Additionally, carefully designed selection strategies can drive for cross-species reactivity, homolog specificity and thermal stability.^{5–7} Conversely, in the absence of the natural *in vivo* processes of B-cell receptor editing and negative selection, the iterative enrichment that underpins phage display can result in the emergence of poor biophysical properties, such as reduced stability, increased aggregation propensity and nonspecific binding.^{8–10} Increased net complementarity-determining regions (CDR) loop charge, and the presence of positively charged patches, have been associated with nonspecificity, poor *in vivo* pharmacokinetics (PK) and ultimately unfavorable developability.^{8,11–13}

Considerable efforts have focused on engineering strategies to reduce CDR-based charge while still maintaining high affinity.^{8,12–15} In addition, several groups have developed *in vitro* screening assays to identify nonspecific or polyreactive antibodies, thus preventing the costly progression of such monoclonal antibodies (mAbs) through the development

pipeline.^{16–18} We have developed one such suite of appropriately high-throughput *in vitro* assays for which standardized scores have been correlated with *in vivo* human PK.¹⁸ We identified deoxyribonucleic acid (DNA) and insulin-binding enzyme-linked immunosorbent assays (ELISAs) as the most sensitive, robust and amenable to high-throughput automation.^{16,18} As a measure of self-association, a property correlated with poor solubility and viscosity, we adopted the affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS) assay.^{17–19} Our previous studies concluded that mAbs scoring > 11 on all three *in vitro* assays (DNA-binding, insulin-binding and AC-SINS assays) are predicted to have poor PK, enabling the removal of poorly behaved mAbs from the pipeline.¹⁸ However, with increasingly complex mechanisms-of-action and more elusive targets, obtaining both highly functional and biophysically well-behaved antibodies remains challenging.

With a goal of delivering as much functional diversity as possible to the downstream screening process, we sought to explore the effects of applying deselection pressures to our phage libraries to remove nonspecific clones. This has been previously applied with some success using discovery libraries in yeast display where a soluble membrane protein (SMP)

CONTACT Orla Cunningham  cunningorla@yahoo.co.uk  Biomedicine Design, Pfizer, Grange Castle Business Park, Clondalkin, Dublin 22, Ireland.

 Supplemental data for this article can be accessed on the [publisher's website](#).

© 2021 Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

preparation was used to identify nonspecific clones.²⁰ These clones could be subsequently sorted and separated from antigen-specific binding populations using fluorescence-activated cell sorting (FACS). Based on our previous experience, dsDNA and SMP derived from chicken cells (DT40) were prioritized as our deselection agents.¹⁸

As an antibody optimization test case we chose MJ4-2, a neutralizing anti-IL-21 R antibody derived from a rat immunization campaign, which competes with IL-21 for binding to its receptor. IL-21-IL-21 R interaction has been shown to be mediated primarily via charge, with 80% of the IL-21 paratope representing positive charge.²¹ A previous anti-IL-21 R affinity optimization effort perhaps unsurprisingly resulted in positive charge enrichment during phage display, and this was shown to directly correlate with poor PK.^{22,23} MJ4-2, derived through *in vivo* evolution, has a nanomolar (nM) affinity for IL-21 R. However, it has sub-optimal biophysical properties, precluding any further development. The goal of this study was to use mutagenesis strategies and phage display in combination with deselection strategies to maintain MJ4-2 affinity for IL-21 R, but mitigate its undesirable specificity and developability issues. A number of deselection strategies were used, but positive charge enrichment proved a substantial challenge that was not possible to avoid through deselection alone. A co-crystal structure of MJ4-2/IL-21 R, together with next-generation sequencing (NGS) datasets derived from optimization library outputs after antigen selection, and deselection, informed linear regression models that ultimately identified a lead clone with greatly improved biophysical properties.

Results

Deselection approaches do not significantly affect selection outcomes

As a first step toward optimization, we humanized rat anti-IL21R MJ4-2 using alternative framework grafting, selecting five variable heavy (VH) and four variable light (VL) germlines as acceptor frameworks for the parental rat CDRs (Figure S1). Clone 2 (C2; V_{H3-30}/V_{K1-39}) demonstrated good periplasmic single-chain variable fragment (scFv) expression and antigen binding while retaining the characteristic sub-optimal biophysical liability scores (Table S1, Figure S2). Soft mutagenesis across all 6 CDRs of C2 generated 15 mutagenic libraries (9 VH and 6 VL sub-libraries) each containing 10^7 – 10^8 variants. Libraries were rescued independently prior to pooling, generating a single VH and VL library pool for selection. Given legacy challenges of extreme positive charge enrichment in CDRs during anti-IL-21 R optimization, two different selection strategies were tested: In-solution deselection and nonspecificity pre-clearing (Figure 1(a)). The in-solution deselection strategy involved standard antigen selection in the presence or absence of competing DNA or SMP (Figure 1(a), Figure S3A). The nonspecificity pre-clearing strategy went a step further, combining the in-solution deselection strategy with a pre-clearing step at each round, i.e., the phage pool was first incubated with DNA/SMP-coated immunotubes for 2 hours prior to subsequent antigen selection in the presence of the

deselective agent (Figure 1(a), Figure S3B). Four rounds of panning were performed with antigen concentration decreasing 10-fold/round.

In order to ensure that selection outputs derived from R1-R4 maintained the same epitope as the parental mAb, a competitive homogeneous time-resolved fluorescence (HTRF) assay was used in screening. In this assay format, the ability of periplasmic scFv outputs to compete with parental C2 binding to IL-21 R is measured, with a decrease in fluorescence representing direct competition, and magnitude of decrease allowing an indirect assessment of comparative affinity. In-solution deselection was successful in retrieving competitive clones from all rounds of selection (Figure 1(b)). In contrast, nonspecificity pre-clearing failed to identify any competitive clones after round 2 of selection. We hypothesized that a R1 starting concentration of 0.3 nM hIL-21 R may be too stringent when combined with aggressive library pre-clearing. Thus, the soft libraries were re-selected via nonspecificity pre-clearing with a R1 starting concentration of 30 nM IL-21 R. Providing a higher antigen concentration from R1 allowed the identification of functional clones from this branch (Figure 1(b), Figure S3C). Competitive clones identified via HTRF were analyzed via Sanger sequencing. The sequencing results revealed limited mutational variation in the C2 variants, with most variants possessing 1–2 substitutions from parental with a clear preference for mutation in VL-CDR3. Within the functional VL population, nine clones with increased positive charge dominated across all branches of selection, regardless of the deselection agent or selection method applied (Figure 1(c,d); Figure S4A). This positive charge enrichment was recapitulated in the VH library outputs, with VH-CDR1 mutation being most dominant (VH results are summarized in Figures S4B–D). The 9 dominant clones, which represented 80% of the Sanger sequences, were reformatted to IgG for assessment in the specificity ELISAs (DNA and insulin) and AC-SINS assay (Figure 1(e)). As expected, given the charge enrichment observed, dominant clones displayed higher scores than the parental antibody in all assays. Initial data suggest that neither dsDNA nor SMP were effective in avoiding charge enrichment during antibody optimization.

Next-generation sequencing provides a comprehensive dataset for computational design approaches

Deselective pressures applied to the C2 mutational libraries failed to yield clones with reduced charge or improved biophysical properties. However, given the limited sampling possible via traditional screening (typically 88 of 10^6 clones/round of selection are assessed in functional screening assays), R3 library outputs were assessed via NGS using the Illumina 300 bp pair-end MiSeq platform (R3 outputs from all deselection conditions displayed maximal IL-21 R binding in a polyclonal phage ELISA and were therefore chosen for subsequent NGS analysis; data not shown). The high-quality reads (average Phred score 37.7) yielded 5.6 – 20×10^6 raw reads/sample (Table S2). Reads were processed for V/J-gene assignment and canonical numbering using in-house bioinformatic tools. To limit the

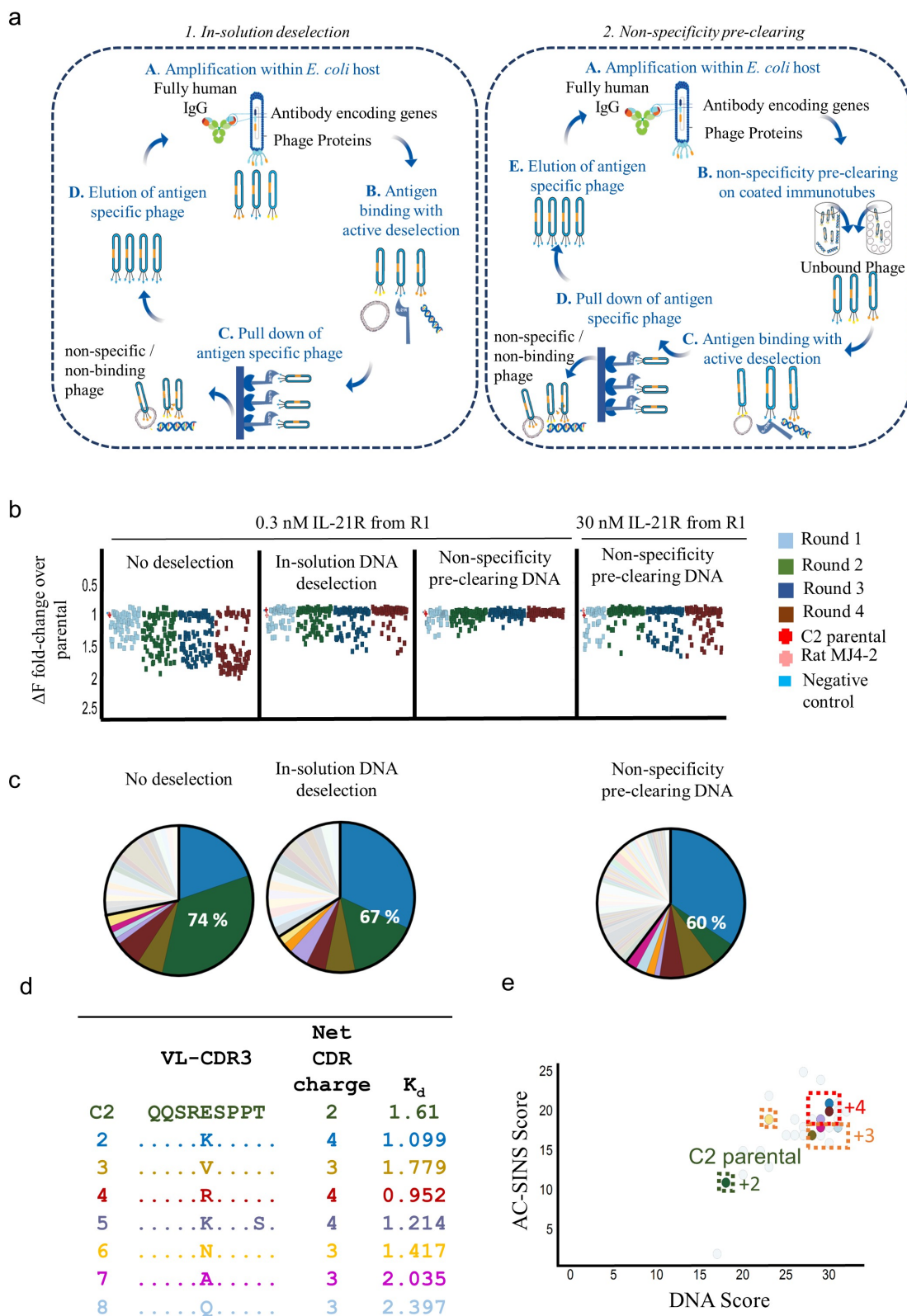


Figure 1. Deselection approaches do not prevent positive charge enrichment. (a) Schematic illustrating two differing selection strategies used to select mutagenic libraries. (b) Selection outputs (Round 1–4) from separate strategies were induced to express scFv and inhibitory activity was assessed via periprep HTRF. (c) Sanger sequencing of clones with equivalent or improved ΔF over parental revealed the percentage of the screening output represented by 9 dominant clones. (d) Sequences of the nine dominant clones, net VL CDR charge and BIAcore affinity measurements (e) AC-SINS and DNA binding scores for the dominant clones.

influence of polymerase chain reaction (PCR) or sequencing generated errors, sequences observed only once and any clone

with a frequency $< 10^{-6}$ was omitted from further analysis. Species accumulation curves highlighted that the diversity of

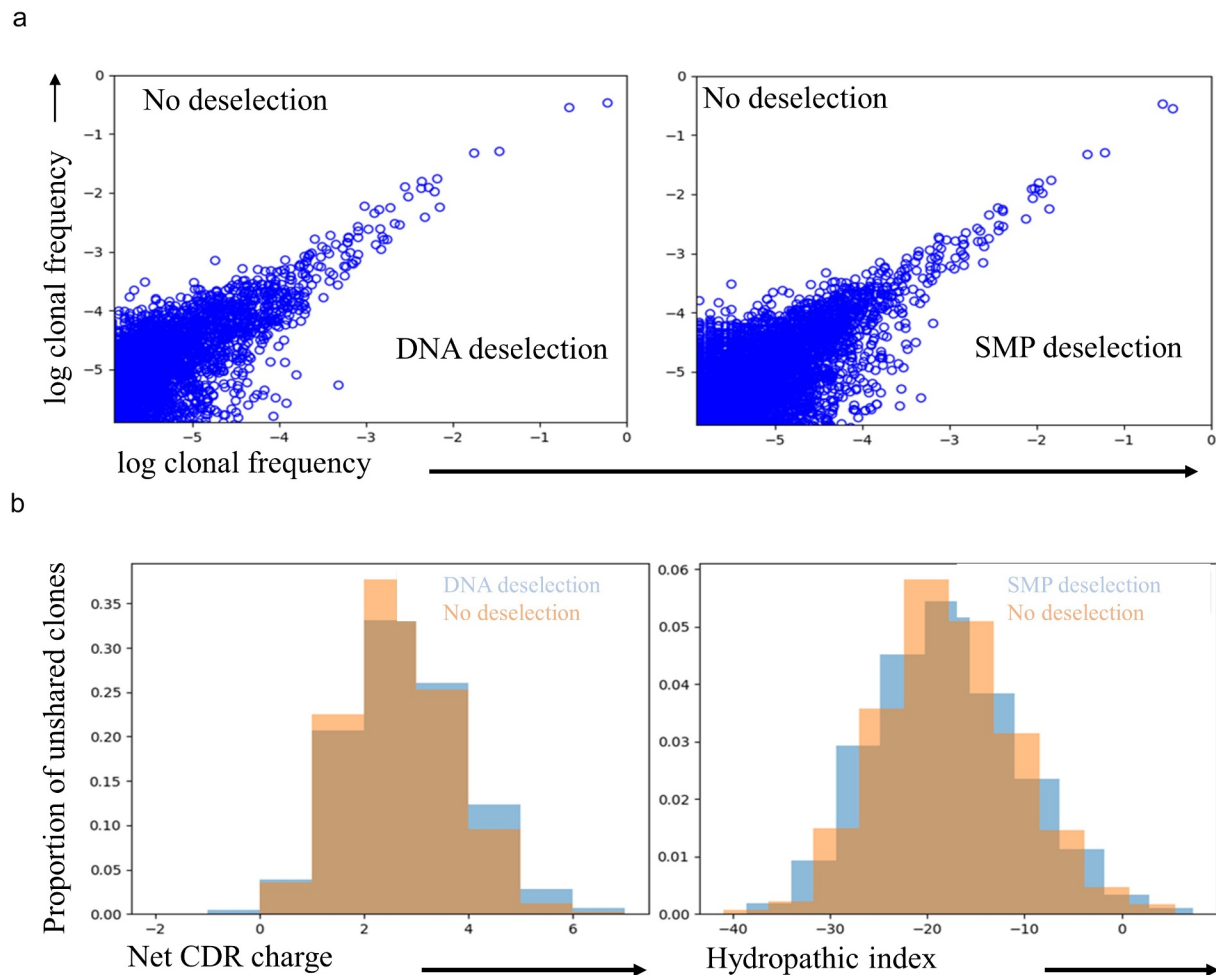


Figure 2. NGS of R3 selection outputs. Next-generation sequencing was performed on R3 outputs of the following selection branches: No deselection, in-solution DNA deselection and in-solution DT40 membrane deselection. (a) A slope of unity was generated with the frequency of shared clones between the non-deselected and deselected samples. (b) Overlap histograms highlighting the net CDR charge of unshared clones between non-deselected and deselected samples.

each output pool had been adequately sampled, with the no deselection sample saturating significantly faster as a result of being allocated a greater sequencing space (Figure S5, Table S2).

To assess the convergence of samples generated in the absence or presence of deselection agents, the degree of shared clones and their relative frequencies within a population were assessed. Across all branches, shared clones illustrated highly correlated frequencies and represented 95–99% of the sample

abundance (Figure 2(a)). This further highlighted that deselection mechanisms did not significantly alter the output clones or their abundance. We next asked if there were notable differences between the remaining un-shared clones, which accounted for less than 0.1% of each sample. Neither DNA nor SMP induced significant change in CDR charge or hydrophobic index (a measure of hydrophobicity) of the remaining unshared clones (Figure 2(b)). Surprisingly, NGS revealed no

Table 1. Frequency of dominant clones within NGS dataset.

	VL-CDR1	VL-CDR2	VL-CDR3	Frequency within each de-selective branch		
				None	DNA	SMP
C2 Parental	GASQSVSISRFNLMH	RASNLAS	CQQSRESPTF	0.268	0.596	0.268
1K.....	0.223	0.216	0.350
2V.....	0.038	0.017	0.037
3R.....	0.040	0.033	0.058
4K...S.	0.009	0.006	0.011
5KH.....	0.004	0.007	0.013
6N.....	0.009	0.004	0.010
7A.....	0.012	0.004	0.010
8Q.....	0.010	0.003	0.009
Sum of frequencies of dominant clones				0.61	0.89	0.77
Percentage of sequence space consumed by dominant clones				61.46	88.79	76.59

NGS data was generated for R3 outputs of the VL soft mutagenic anti-IL-21 R library. Datafiles were searched and frequencies noted for the C2 parental clone and 8 additional clones which dominated the traditional selection output.

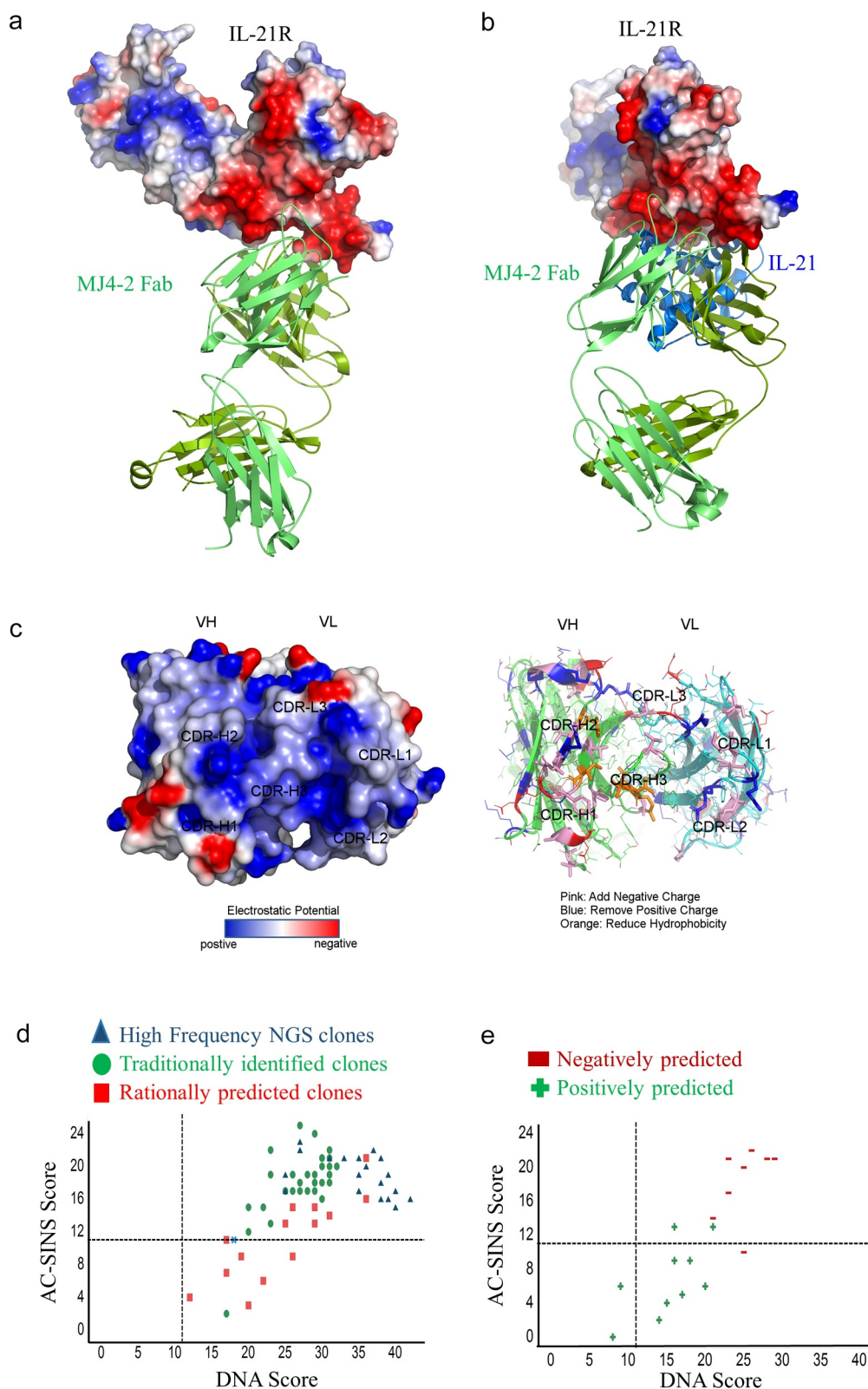


Figure 3. Structurally guided interrogation of NGS datasets. (a) The structure of hIL-21 R interacting with the parental lead molecule MJ4-2 (b) Overlay of the structure of hIL-21 R in complex with both MJ4-2 Fab and IL-21. (c) Top down view of the MJ4-2 binding pocket highlighting the electrostatic potential and mutations within the binding pocket predicted to improve biophysical properties. (d) AC-SINs and DNA score for clones identified as high-frequency NGS clones, clones identified through traditional screening methods and clones rationally designed from the NGS data set or (e) Clones predicted via linear regression to have improved or poor biophysical properties.

difference between deselection agents and indeed no discernible difference to the no deselection control branch.

Nonetheless, NGS did provide valuable insights into the traditional screening processes: dominant clones identified

via traditional screening represented a similar proportion of the NGS generated pool (Table 1) with less dominant clones present at both high and low frequencies (Table S3). Furthermore, NGS provided a vast dataset of previously unsampled clones for interrogation.

In order to further explore the NGS dataset for potential IL-21 R binders with reduced biophysical liabilities, we solved the co-crystal structure of MJ4-2 interacting with IL-21 R and combined both datasets to inform a computational design approach. The crystal structure of MJ4-2 Fab in complex with the extracellular domain of IL-21 R revealed that MJ4-2 binds almost exclusively to the negatively charged D1 domain of IL-21 R, making only one contact with the D2 domain. The binding of MJ4-2 to D1 is mediated by the strong positive electrostatic surface potential of the VL CDR loops (Figure 3(a)), with each donating one arginine to form hydrogen-bonds with the negatively charged residues on D1. Comparison with the structure of IL-21 bound to IL-21 R, confirmed that MJ4-2 antagonizes IL-21 R by competing with IL-21 for overlapping binding determinants (Figure 3(b)). A top-down view of the binding pocket highlights the positively charged patches across both the VH and VL CDRs (Figure 3(c)). Computational binding affinity and stability predictions were used to identify non-contact points that would tolerate either increased negative charge, reduced positive charge or reduced hydrophobicity, while retaining antigen binding (Figure 3(c)). Of 40 rationally designed VL variants, 14 of these already existed in the NGS dataset.

These rationally predicted clones, alongside 21 high-frequency NGS clones that were not identified via traditional screening, were produced in IgG format for assessment in our suite of biophysical liability assays (Figure 3(d)). While the rationally predicted IgGs trended toward improved biophysical behaviors when compared with high-frequency NGS and traditionally identified IgGs, none possessed biophysical liability scores within the acceptable thresholds (<11 in all assays).¹⁸ Nevertheless, these additional populations contributed to an IgG dataset linking performance to sequence. This dataset (consisting of 180 IgGs with associated DNA- and insulin-binding ELISA scores, AC-SINs scores and fold change in half-maximal inhibitory concentration (IC₅₀) as determined by HTRF) was used to generate linear regression models for each assay parameter. These models were generated by converting the sequences of the variable domains into vectors of length 20xN where N is the number of sites, and each site is encoded as a 20 × 1 vector that is all zeros except for a 1 corresponding to the particular amino acid. One vector for each of the 180 IgGs was then combined into a single matrix with corresponding assay data for each of the three assays. We then used the linear regression module of matlab (regress) to fit the weight matrix for each assay.²⁴ New sequences were scored by matrix multiplication of the sequence matrix and the weight matrix. The models for the DNA, insulin and AC-SINS scores were applied to the NGS datafiles, generating predicted scores for all sequences. Sequences predicted to be improved over the parental for each parameter were filtered. Of the 119 sequences identified via linear regression (88 VH and 32 VL), only 10 had predictive scores <11 on all three biophysical assays. Moreover, all 10 of these sequences came from the VH dataset, with none

of the 32 VL identified sequences having a predicted cumulative score < 33 (sum of predicted DNA, insulin and AC-SINS). To evaluate the use of such a model in extracting improved clones from an NGS dataset, 18 sequences were produced in IgG format, 10 with predictive scores <11 for all 3 assays and 8 with poor predicted scores. When produced in IgG format, there was a notable difference in the DNA binding and AC-SINS scores, with clones predicted to be well-behaved having improved scores over the poorly predicted clones (Figure 3(e)). While two of these clones had favorable AC-SINs and DNA binding scores (Figure 3(e)), both molecules had significantly reduced affinity (7–15-fold; data not shown).

Structure-guided rational library design enables identification of clones that meet developability criteria while maintaining IL-21 R affinity

In addition to informing predictive modeling, the co-crystal structure of IL-21 R/MJ4-2 was used to design charge-reduced libraries for the VL and VH, allowing combinatorial sampling of mutations previously identified as being amenable to negative charge introduction, positive charge removal or reduction of hydrophobicity (Figure 4(a)). Each library was built and selected against hIL-21 R either in the absence of deselection pressure or with nonspecificity preclearing with DNA or SMP (Figure S6). After two rounds of selection, the VH and VL outputs were combined to generate a VH-VL library pool, which was then selected through round 3 and round 4. Periprep HTRF screening was carried out on output rounds 3 and round 4, these outputs were also prepped for NGS analysis. Initial screening identified 29 clones that were expressed as IgG and characterized. Of these rational library clones, 5 did indeed meet the developability criteria (Figure 4(b)). Inherent to library design, these rational library clones were mutationally more distant from the parental when compared with IgGs identified from the soft mutagenic library; all containing 2 or more mutations from parental across all CDRs, not just in the previously observed hotspots of VL-CDR3 and VH-CDR1 (Figure 4(b)). However, all displayed poor binding to IL-21 R when assessed using BIAcore (data not shown). Despite not achieving the desired combination of reduced biophysical properties and increased affinity, the rational library outputs did provide an additional dataset in which combinatorial mutations across the VH-VL sequence could be linked with function. Incorporation of the rational library screening and NGS outputs into the previously applied linear regression models, and application of these models to the rational library NGS dataset, resulted in the identification of 33 clones predicted to have nonspecificity scores < 11 in all three assays. When produced as IgG, 5 clones met all developability criteria (Figure 4(c)). Furthermore, BIAcore assessment revealed that mAb5-Rational retained IL-21 R binding with an affinity comparable to the parental control (Figure 4(c,d)).

Discussion

The development and implementation of high-throughput *in vitro* assays at early stages of drug discovery has helped to identify mAbs with poor developability profiles. However,

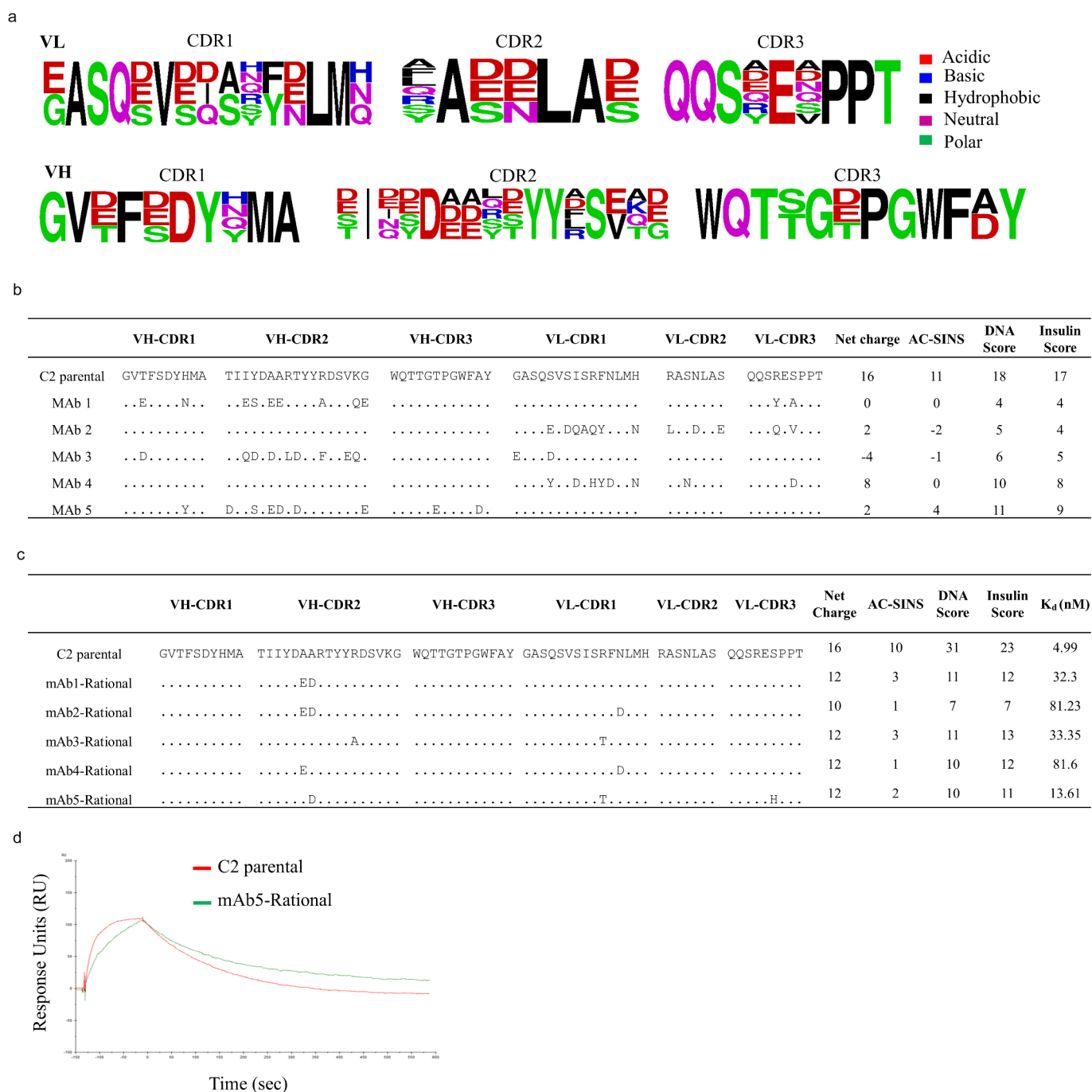


Figure 4. Design, selection and screening of a computationally designed rational library. (a) Amino acid diversity introduced across all 6 CDRs within the rational library design. Sequence and biophysical scores of clones identified from selection and screening of a (b) rationally designed library or (c) application of linear regression models to the rational library NGS outputs. (d) Overlay of BIAcore trace of C2 parental and mAb5-Rational.

during the antibody library selection process, mAbs with poor biophysical properties or excessively charged CDR loops can dominate selection outputs. In this study, we aimed to maintain affinity for IL-21 R while improving the poor biophysical properties of the anti-IL-21 R mAb MJ4-2. Using parallel deselection methods, in combination with traditional screening and NGS analysis of outputs, we demonstrate that neither dsDNA nor SMP can prevent positive charge enrichment. Previous studies using SMP in combination with yeast display have shown success in decreasing the number of hits with poor biophysical properties.^{17,20,25} Of note, the yeast display

platform uses stringent FACS gating, preventing the carry-over and iterative enrichment of these poorly behaved clones through rounds of selection. In the particular case of the IL-21 R mAb, the combination of the low target isoelectric point (pI; 4.57) together with a reasonably high starting affinity for target (5 nM) may have added to the challenges of using low affinity, generic deselection agents.

The ability of NGS to provide > 10,000 antibody sequences/binding population has the potential to revolutionize antibody discovery, providing access to an unprecedented number of potential lead candidates while simultaneously providing

critical insights into library build and selection practices. As such, NGS enabled comprehensive coverage of the binding populations, overcoming a common concern of insufficient output sampling through traditional screening methods. This allowed an empirical demonstration of the failure of DNA and SMP deselection to significantly alter the output population. With NGS we observed 96% clonal abundance overlap between selection branches, building on previous studies in which up to 75% clonal overlap was observed between replicate pannings and analysis via Sanger sequencing.²⁶ We found that dominant clones identified via traditional screening are present at a similarly high frequency within the NGS dataset, highlighting the likelihood of high-frequency NGS clones representing functionally relevant hits. Conversely, random colony picking, which underpins traditional screening, can isolate functional clones present within the NGS dataset at frequency as low as 10^{-6} while simultaneously omitting many additional high-frequency binders.

The lack of associated function for NGS discovered sequences, while a major hurdle for the incorporation of NGS into naïve discovery, is somewhat overcome in the affinity optimization setting, where epitope and mode of action are frequently known. In this setting, NGS can enable multi-parametric optimization through the provision of vast datasets for *in silico* filtering on CDR charge, hydrophathy index, post-translational modification associated sequence liabilities and predicted immunogenicity.^{27–30} Indeed, the application of such filtering tools by Mason *et al* to a population of *in silico* predicted binders enabled the reduction of potential leads from 7.2×10^7 to 40,000.³¹ As an alternative method of navigating the vast sequence space provided by NGS, we generated a significant dataset (~ 200 IgGs) with associated functional and biophysical characterization data and implemented linear regression models to predict improvements in biophysical behavior and affinity. The poor biophysical behavior of predicted clones, conferred via increased negative charge in VH-CDR1, directly correlated with a significant decrease in antigen binding. This intricate relationship between positive charge, antigen binding and PK was previously reported with a reduction in positive charge across CDRs reducing nonspecificity, and improving *in vivo* PK, but ultimately at the cost of antigen binding.⁸

The fundamental role of positive charge in IL-21 R binding was further highlighted in analysis of the structurally designed library outputs. Rational library-derived clones with favorable DNA, insulin and AC-SINS scores had reduced hIL-21 R binding. However, the functional data generated with these rational library clones provided additional diversity to the existing dataset linking genotype-phenotype, allowing refinement of the existing linear regression models that could be applied to the associated NGS data set. Ultimately, this refined analysis identified a lead clone that retained affinity for IL-21 R while also meeting all criteria in the biophysical characterization assays.

The ability of the parental mAb MJ4-2 to out-compete the existing highly electrostatic interaction between hIL-21 and its receptor is dependent on the presence of a large positively charged patch, which also contributes to nonspecific

interactions resulting in poor *in vivo* PK. In this study, where antigen specificity and charge are intricately linked, we have highlighted the difficulty in maintaining functionality while attempting to reduce charge to improve overall biophysical behavior. Surrogate *in vitro* deselection pressures attempting to replace the process of *in vivo* receptor editing were ineffective in avoiding charge enrichment. Indeed, the low diversity, highly dominant clones that resulted from the iterative rounds of selection of the single-CDR soft mutagenesis libraries demonstrated significant charge enrichment and worse behavior than the parental molecule. Where complex, multi-parameter optimization is required, it seems that combinatorial and potentially compensatory mutations are essential to achieve the fine balance between affinity and developability. With this in mind, we feel that neither single-CDR, soft mutagenesis libraries or deselection pressures coupled with phage display were helpful in achieving our ultimate goal. Instead, using structure to design a mutationally restricted library that sampled across all six CDRs in parallel provided the necessary information around combinatorial mutations that could not have been predicted *a priori*. Coupling this structure-guided design with NGS analysis and linear regression models, we demonstrated that it is possible to circumvent the complex charge-affinity relationship to achieve sufficient binding in combination with acceptable developability to support further therapeutic development.

Materials and methods

Humanization

Rat MJ4-2 CDR sequences were grafted onto the most similar human germline V-gene sequences (IGHV3-7, IGHV3-30, IGHV2-23, IGHV1-69, IGHV5-51, IGKV1-39, IGKV2-28, IGKV3-20, IGKV4-1). Germline sequences were derived from the international ImMunoGeneTics information system® database (<http://www.imgt.org/>). Humanized VH/VL genes were cloned into expression vectors and synthesized at Blue Heron.

IgG expression

IgG expression plasmids were transfected with expifectamine (Life Technologies) into human embryonic kidney-293 (HEK-293) expi cells and conditioned media was harvested and filtered 5 days later. IgG purification was achieved using ProPlus resin tips on the MEA system (Phynexus) followed by buffer exchange into phosphate-buffered saline (PBS).

Biophysical analysis

Purified IgGs were analyzed in ELISA format in a high-throughput automated workflow for binding to DNA and insulin against in-house controls, as previously described.¹⁸ A similarly high-throughput method was used for IgG assessment in the AC-SINS assay using purified IgGs at a concentration of 1 mg/mL as previously described.¹⁸

ScFv assessment

Clones 2, 4, 6 and 8 that were prioritized from the IgG analysis were synthesized as scFv fragments in both V_H - V_L and V_L - V_H orientation by Blue Heron, ligated into a proprietary in-house phagemid vector and transformed into electrocompetent *Escherichia coli* (*E.coli*) TG1 cells (Lucigen). For each clone, a 10 ml culture was grown to mid-log phase ($OD_{600} = 0.5$) and scFv expression was induced with addition of isopropyl β -D-1-thiogalactopyranoside to a final concentration of 0.25 nM, followed by overnight growth at 30°C with shaking at 350 rpm. The periplasmic fraction was prepared as previously described.³² The presence of soluble scFv within these fractions was assessed by Western blot (anti-c-myc horse radish peroxidase, Roche) and sodium dodecyl sulfate–polyacrylamide gel electrophoresis. A periprep binding ELISA was carried out as described previously starting at a concentration of 50% (v/v) in a 2-fold dilution series prepared in casein buffer with plates coated at 1 μ g/mL IL-21 R.³³

ScFv library design and construction

For soft-mutagenic library design, introduction of diversity along the 6 CDRs of Clone 2 was achieved through the design of mutagenic primers, which allow approximately 50:50 incorporation of wild-type amino acid:19 amino acids at any given position. Primers were designed to target 5–7 consecutive amino acids along the CDRs, generating 15 diversity cassettes: VHCDR1a, VHCDR2b, VHCDR2a, VHCDR2b, VHCDR2c, VHCDR2d, VHCDR3a, VHCDR3b, VHCDR3c, VLCDR1a, VLCDR1b, VLCDR1c, VLCDR2a, VLCDR3a, VLCDR3b (Table S4). Each diversity cassette was then incorporated into a library insert using the splicing by overhang extension polymerase chain reaction (SOE-PCR) assembly method.^{34,35} For the charge-reduced rational library design, VH/VL diversity cassettes were generated by TWIST Bioscience with each amplicon subsequently paired with the corresponding parental chain (VL/VH). Combination of the charge-reduced rational library outputs after 2 rounds of independent selection was achieved through isolating the mutated VH/VL chain from the respective pool via PCR using vector-specific primers, and combining the VH and VL mutants to generate a single paired amplicon using the SOE-PCR assembly method. Library inserts were ligated into in-house phagemid vectors, transformed into electrocompetent *E.coli* TG1 cells (Lucigen) and rescued independently with helper phage M13K07 (NEB), as previously described.^{32,36} Library quality control was carried out with sanger sequencing of 48 clones per sub-library. For the soft-mutagenic libraries, rescued phage from the VH and VL sub-libraries were pooled in equimolar concentrations to generate a master VH and VL library for selection.

DT40 soluble membrane protein preparation

The chicken lymphoblast cell line DT40 was obtained from ATCC (CRL-2111) and cultured at 37°C in Dulbecco's modified Eagle's medium with 4 mM L-glutamine modified to

contain 4.5 g/L glucose, 1.5 g/L sodium bicarbonate and 0.05 mM 2-mercaptoethanol, supplemented with 10% tryptose phosphate broth, 10% fetal bovine serum, and 5% chicken serum. Preparation of soluble membranes was accomplished using previously described methods.²⁵ Briefly, 1×10^9 cells (at 10^6 – 10^7 /ml at confluency) were pelleted at 550 x g for 3 min. Cells were kept at 4°C during centrifugation and for all subsequent steps. Cells were washed with 100 ml of PBS-F (1 x PBS, 1 mg/ml bovine serum albumin) followed by a wash with 30 ml Buffer B (50 mM HEPES, 0.15 M NaCl, 2 mM $CaCl_2$, 5 mM KCl, 5 mM $MgCl_2$, 10% Glycerol, pH 7.2). The cell pellet was resuspended in Buffer B (3x pellet volume) with protease inhibitors (2.5x pellet volume; Complete protease inhibitor (+P), Roche). The cells were homogenized and the membrane fraction was collected by centrifugation at 40,000 x g for 1 hr at 4°C. The membrane pellet was gently rinsed with 1 ml of cold Buffer B (+P), and supernatant discarded. The pellet was resuspended in 3 ml of Buffer B (+P), transferred to a Dounce homogenizer and disrupted by 30–35 strokes with a tight-fitting pestle. The protein concentration of the resulting enriched membrane fraction (EMF) was determined using the bicinchoninic acid assay (BCA assay; Pierce).

Library selection and screening

Phage were blocked for 1 hr in 3% M-PBS followed by deselection on streptavidin-coated magnetic beads (Dynabeads, Invitrogen, Cat # 112–06) for 30 min. Once blocked and deselected on beads, nonspecificity pre-clearing was achieved through 2 hr incubation on immunotubes, coated with DNA (5 mg/ml, Sigma) or SMP (100 μ g/ml). In-solution deselection was achieved through co-incubation of blocked, bead-deselected phage with DNA (25 μ g/ml)/SMP (100 μ g/ml) and biotinylated hIL-21 R (in house-source) for 1 hr. The concentration of hIL-21 R was decreased 10-fold/round from a starting concentration of 30 nM/300 pM for nonspecificity pre-clearing and in-solution methods, respectively. Streptavidin-coated magnetic beads were used to pull-down the hIL-21 R-biotin-phage complexes (15 min), washing of bead-antigen-phage complexes was increased from 5 to 15 washes from round 1–4. Phage were dissociated from magnetic beads with 400 μ l triethylamine (100 mM, 10 min, Sigma) before neutralization with 200 μ l Tris-HCl pH 7.5 (1 M, Sigma) and re-infection into mid-log electrocompetent *E.coli* TG1 cells ($OD_{600} = 0.5$). The output population was diluted on carbenicillin-supplemented 2YT agar plates (100 μ g/ml), allowing for calculation of the output titer and selection of 88 individual colonies/branch/round for high-throughput screening. ScFv expression, periplasmic preparation and HTRF screening was carried out as previously described.³² The final HTRF reaction mixture contained 2.5 nM biotinylated hIL-21 R, 1:500 dilution of SA-XL665 (CisBio), 1:500 dilution of europium cryptate-labeled MJ4-2 (prepared using a cryptate labeling kit as per manufacturer's instructions, CisBio) and 50% (v/v) of periplasmic extract from clones of interest.

NGS of phage derived outputs

Plasmid DNA was isolated directly from the glycerol stock of the selection round of interest (approx. 20 μ l at OD₆₀₀ = 20–50) using the QIAprep spin miniprep kit (QIAGEN). Isolated dsDNA was quantified on the Qubit 3.0 fluorometer using the Qubit[®] dsDNA HS kit (Thermo). The generation of a VH, VL, VH-VL amplicon for sequencing was generated through a nested PCR. For the VH-VL amplicon, the initial PCR utilized a forward primer specific to the vector leader sequence prior to the VH-VL insert and a reverse primer downstream of the VL-FW4. For the generation of individual VH/VL amplicons the site-specific reverse/forward primers were complimentary to the scFv glycine-serine-linker. The initial PCR (40 μ l reaction) was performed using Phusion HF 2X master mix (Thermo) and the following cycle conditions: 98°C for 30 sec, 15 cycles of 98°C for 10 sec, 70°C for 30 sec, 72°C for 30 sec and a final extension at 72°C for 8 min, storage at 4°C. The initial PCR primers also incorporated an 8-nucleotide index, 6-nucleotide unique molecular identifier and partial P5/P7 tag for the Illumina platform. PCR 1 product was run of a 2% agarose gel, purified using the Wizard[®] SV gel and PCR clean-up kit (Promega), quantified using the Qubit 3.0 and a second PCR performed to generate complete P5/P7 adaptors for annealing to the Illumina MiSeq chip. PCR 2 (100 μ l) conditions were as follows: Phusion GC 2X master mix (Thermo) with cycle conditions: 98°C for 30 sec, 15 cycles of 98°C for 8 sec, 69°C for 12 sec, 72°C for 10 sec and a final extension at 72°C for 5 min with storage at 4°C. PCR 2 products were run on a 2% agarose gel, gel purified, quantified and amplicon purity was assessed on the 2100 Bioanalyzer (Agilent) using the High Sensitivity DNA kit (Agilent).

NGS data analysis

Individual samples were pooled, spiked with 2% PhiX and sequenced on the Illumina MiSeq platform using the v3 kit (2 X 300 cycles, paired-end). Paired-end reads were merged using FLASH.³⁷ We then used in-house sequencing tools for framework, CDR and germline assignment. In this method, clones are defined as unique combinations of V-Gene, J-Gene and CDR1, CDR2 and CDR3. Clonotypes are defined as clones that share the same CDR3 and lineages are clonotypes that share similar CDR3 sequences, specifically having the same length and a positive BLOSUM62 score. Only complete V-domain reads, i.e., first position of FW1 through last position of FW4, were retained for further analysis. A Phred threshold of 25 was applied to bases that define the codons at V-domain positions, and any sequence was eliminated if it contained a base with a quality falling below that threshold, while requiring the definition of a germline identity or a CDR position.³⁸ This has the largest effect on the number of sequences retained, but is critical to maintaining the fidelity of clone, clonotype and lineage assignments. Singlets were removed from all analyses, as was any clone with a frequency <10⁻⁶.

Species accumulation curves

Species are defined as clonotypes with counts above 1. An accumulation curve was generated by taking the clonal frequency distribution ($\sum f_i$, for n clonotypes) of those clonotypes in the sample and simulating their recovery in the full dataset. The accumulation is reported as the percentage of clonotypes recovered as a function of the percentage of the dataset sampled.

hIL-21 R -MJ4-2 structure

Protein purification and complex formation

The anti-IL-21 R- MJ4-2 antibody was expressed in 6 L of media and purified on 6 ml of Mab Capture Protein A resin (Thermo Fisher). The Fab fragment was generated through digestion of this purified protein with immobilized papain (Thermo Scientific) for 24 hours at 37°C. The Fc domain was removed through binding the digest to 1 ml Mab Capture Protein A resin (Thermo Fisher) resulting in a Fab only fraction. The IL-21 R was purified on 10 ml of ProBond Ni resin (Thermo Fisher) followed by a second purification using a Superdex 200 (Thermo Fisher) preparative SEC column in Tris-buffered saline (TBS). Purified Anti IL-21- MJ4-2 was complexed with the IL-21 R at a 1:1.2 Fab: receptor molar ratio and the complex purified on a Superdex 200 preparative SEC column equilibrated in TBS. The protein was concentrated to 14 mg/ml and set up in crystallography trays under a variety of conditions.

Resolution

MJ4-2 Fab in complex with IL-21 R was crystallized at 18°C from a solution containing 10% PEG MME 5000 and 0.1 M MES, pH 6.5. The crystals had symmetry consistent with orthorhombic space group C2221 and contained one protein complex in the asymmetric unit cell. A data set to a 2.2 Å resolution was collected from a single frozen crystal at IMCA beamline 17-ID at the Argonne National Laboratory (APS). The data were processed and scaled using HKL2000. The final data set was 88% complete with average redundancy of 4.6 and with Rmerge of 10%. The structure was solved by molecular replacement with PHASER using the Fab model prepared from our structures of the previously obtained anti-human IL-21 R clones (data not published) and using the IL-21 R model from our crystal structure obtained for IL-21 R bound to IL-21 (data not published). Several iterative rounds of model manual adjustment and model rebuilding using COOT followed by crystallographic refinement using Phenix yielded the final refined model of the complex with a crystallographic Rwork of 21.5% and Rfree of 24.0%. Statistics for data collection and refinement are shown in Table S5. The structure is deposited in the RCSB Protein Data Bank (PDB ID code 7KQ7).

Design of rational mutants

Using the starting crystal structure, we identified mutations that had reduced charge (Arg/Lys/His to Neutral/Negative), increased positive charge (Neutral to Glu/Asp) or had reduced hydrophobicity that would be tolerated. To do this we

predicted the binding affinity and stability for all potential point mutants using Discovery Studio 4.5 and FoldX. Mutations that had a $\Delta\Delta G$ of <-1 kcal/mol were considered tolerated.

Discovery studio

From Discovery Studio, initial pdb formatted structures were converted to .dsv formatted structures by applying the “Prepare Protein” protocol of Discovery Studio 4.5 (Accelrys Inc.) The protocol prepares the parent structure by correcting nonstandard atom names, selecting single conformations for sites with alternative conformations by taking the position with highest occupancy, and adding missing side-chain atoms. Subsequently, side chains for residues with missing atoms were optimized, waters removed, missing loops modeled, and protonation states of titratable side chains were predicted using a pH of 7.4. This protocol was run separately for the VH/VL+IL-21 R chain pdb file and the VH/VL chain pdb file. The change in binding affinity upon mutation was calculated by applying the “Calculate Mutation Energy (Binding)” protocol from Discovery Studio 4.5 using a VH/VL+IL-21 R chain .dsv formatted structure. The default parameters were used with the exception of the nonpolar Surface Coefficient, which was set to 0.007. Here the ligand chain was set to be the IL-21 R Chain. In addition to affinity, changes in stability of the Fab upon mutation were calculated using the VH/VL chain .dsv file. Here, we applied the Calculate Mutation Energy (Stability)” protocol from Discovery Studio 4.5. The default parameters were used with the exception of the nonpolar Surface Coefficient, which was set to 0.007. The same sites as those used for the binding prediction were calculated.

Fold X

Stability and affinity calculations were performed using previously described scripts and methods.³⁹ Similarly, the FoldX 4 program was implemented as previously described.^{40,41} Briefly the VH/VL+IL-21 R chain pdb file was run through the repair protocol, and then each mutation was run through the Mutate, analyze and stability protocols. Here the VH and VL chains were defined as the complex and the IL-21 R chain was defined as the ligand. All 20 amino acid single mutants in all sites of the VH and VL chain were calculated. This was referenced to the wild type (WT) position at each site. The change in binding affinity and stability was extracted by taking the difference between a mutation and the WT residue at each site.

Linear regression analysis

Sequences of the variable regions of 180 IgGs with associated DNA- and insulin-binding ELISA scores, AC-SINs scores and fold change in IC50 as determined by HTRF, were one-hot encoded into $1 \times 20 \times N$ -dimensional vector, where N is the number of sites in the VH plus VL, and each site is 20-column vector that only contains a 1 where the letter corresponds to that amino acid. For each assay, a $20 \times N \times 1$ vector of regression coefficients for each amino acid at each site was determined by using the “REGRESS” multiple linear regression using least-squares method in Matlab R2010b.²⁴ The assay results were taken as the difference from the parental value except the IC50, which was treated as the log of the fold-change

from the parental. To predict the values of sequences outside the training set, these new sequences were one-hot encoded and multiplied by the vector of regression coefficients.

Acknowledgments

The authors would like to thank the Pfizer Postdoctoral Program which provided funding and support for this work. The authors acknowledge Macy Jin for the derivation of the original rat MJ4-2 antibody. The authors also thank Edward Lavallie, Maya Arai, Adam Root, Jason Wade and Sarah Hanscom for reagent generation and protein expression and purification support. For biophysical assessment of proteins, the authors thank Amy Tam. For assistance on methods for co-complex formation, the authors thank Wayne Stochaj.

Abbreviations

AC-SINS	affinity-capture self-interaction nanoparticle spectroscopy
BCA assay	bicinchoninic acid assay
CDRs	complementarity-determining regions
DNA	deoxyribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
FACS	fluorescence activated cell sorting
HEK-293	human embryonic kidney –293
HTRF	homogeneous time resolved fluorescence
IC50	half maximal inhibitory concentration
IgG	immunoglobulin class G
mAbs	monoclonal antibodies
NGS	next-generation sequencing
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pI	isoelectric point
PK	pharmacokinetics
scFv	single-chain variable fragment
SMP	soluble membrane preparation
SOE-PCR	splicing by overhang extension polymerase chain reaction
VH	variable heavy
VL	variable light
WT	wild type

Disclosure of potential conflicts of interest

All authors are past or present employees of Pfizer.

ORCID

Orla Cunningham  <http://orcid.org/0000-0002-3888-9541>

References

1. Bhaskar V, Goldfine ID, Bedinger DH, Lau A, Kuan HF, Gross LM, Handa M, Maddux BA, Watson SR, Zhu S, et al. A fully human, allosteric monoclonal antibody that activates the insulin receptor and improves glycemic control. *Diabetes*. 2012;61(5):1263–71. doi:10.2337/db11-1578.
2. Lim AP, Chan CE, Wong SK, Chan AH, Ooi EE, Hanson BJ. Neutralizing human monoclonal antibody against H5N1 influenza HA selected from a Fab-phage display library. *Virology*. 2008;5:130. doi:10.1186/1743-422X-5-130.
3. Kirkham PM, Neri D, Winter G. Towards the design of an antibody that recognises a given protein epitope. *J Mol Biol*. 1999;285:909–15. doi:10.1006/jmbi.1998.2336.

4. Parsons HL, Earnshaw JC, Wilton J, Johnson KS, Schueler PA, Mahoney W, McCafferty J. Directing phage selections towards specific epitopes. *Protein Eng.* 1996;9:1043–49. doi:10.1093/protein/9.11.1043.
5. Fennell BJ, McDonnell B, Tam AS, Chang L, Steven J, Broadbent ID, Gao H, Kieras E, Alley J, Luxenberg D, et al. CDR-restricted engineering of native human scFvs creates highly stable and soluble bifunctional antibodies for subcutaneous delivery. *mAbs.* 2013;5:882–95. doi:10.4161/mabs.26201.
6. Park H, Kim D, Son E, Shin S, Sa JK, Kim SH, Yoon Y, Nam D-H. Antitumor activity, pharmacokinetics, tumor-homing effect, and hepatotoxicity of a species cross-reactive c-Met antibody. *Biochem Biophys Res Commun.* 2017;494:409–15. doi:10.1016/j.bbrc.2017.09.061.
7. Pershad K, Pavlovic JD, Graslund S, Nilsson P, Colwill K, Karatt-Vellatt A, Schofield DJ, Dyson MR, Pawson T, Kay BK, et al. Generating a panel of highly specific antibodies to 20 human SH2 domains by phage display. *Protein Eng Des Sel.* 2010;23(4):279–88. doi:10.1093/protein/gzq003.
8. Datta-Mannan A, Thangaraju A, Leung D, Tang Y, Witcher DR, 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.* 2015;7(3):483–93. doi:10.1080/19420862.2015.1016696.
9. Julian MC, Li L, Garde S, Wilen R, Tessier PM. Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability. *Sci Rep.* 2017;7:45259. doi:10.1038/srep45259.
10. Wu SJ, Luo J, O'Neil KT, Kang J, Lacy ER, Canziani G, Baker A, Huang M, Tang QM, Raju TS, et al. Structure-based engineering of a monoclonal antibody for improved solubility. *Protein Eng Des Sel.* 2010;23:643–51. doi:10.1093/protein/gzq037.
11. 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.* 2008;377:1518–28. doi:10.1016/j.jmb.2008.01.093.
12. Schoch A, Kettenberger H, Mundigl O, Winter G, Engert J, Heinrich J, Emrich T. Charge-mediated influence of the antibody variable domain on FcRn-dependent pharmacokinetics. *Proc Natl Acad Sci U S A.* 2015;112:5997–6002. doi:10.1073/pnas.1408766112.
13. Li B, Tesar D, Boswell CA, Cahaya HS, Wong A, Zhang J, Meng YG, Eigenbrot C, Pantua H, Diao J, et al. Framework selection can influence pharmacokinetics of a humanized therapeutic antibody through differences in molecule charge. *MAbs.* 2014;6(5):1255–64. doi:10.4161/mabs.29809.
14. Sampei Z, Igawa T, Soeda T, Okuyama-Nishida Y, Moriyama C, Wakabayashi T, Tanaka E, Muto A, Kojima T, Kitazawa T, et al. Identification and multidimensional optimization of an asymmetric bispecific IgG antibody mimicking the function of factor VIII cofactor activity. *PLoS One.* 2013;8(2):e57479. doi:10.1371/journal.pone.0057479.
15. 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.* 2010;23(5):385–92. doi:10.1093/protein/gzq009.
16. 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.* 2012;4(6):753–60. doi:10.4161/mabs.22189.
17. 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.* 2015;7(4):770–77. doi:10.1080/19420862.2015.1043503.
18. Avery LB, Wade J, Wang M, Tam A, King A, Piche-Nicholas N, Kavosi MS, Penn S, Cirelli D, Kurz JC, et al. Establishing in vitro in vivo correlations to screen monoclonal antibodies for physicochemical properties related to favorable human pharmacokinetics. *MAbs.* 2018;10(2):244–55. doi:10.1080/19420862.2017.1417718.
19. Liu Y, Caffry I, Wu J, Geng SB, Jain T, Sun T, Reid F, Cao Y, Estep P, Yu Y, et al. High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. *mAbs.* 2014;6(2):483–92. doi:10.4161/mabs.27431.
20. Kelly RL, Le D, Zhao J, Wittrup KD. Reduction of nonspecificity motifs in synthetic antibody libraries. *J Mol Biol.* 2018;430:119–30. doi:10.1016/j.jmb.2017.11.008.
21. Hamming OJ, Kang L, Svensson A, Karlsen JL, Rahbek-Nielsen H, Paludan SR, Hjorth SA, Bondensgaard K, Hartmann R. Crystal structure of interleukin-21 receptor (IL-21R) bound to IL-21 reveals that sugar chain interacting with WSXWS motif is integral part of IL-21R. *J Biol Chem.* 2012;287:9454–60. doi:10.1074/jbc.M111.311084.
22. Vugmeyster Y, Allen S, Szklut P, Bree A, Ryan M, Ma M, Spaulding V, Young D, Guay H, Bloom L, et al. Correlation of pharmacodynamic activity, pharmacokinetics, and anti-product antibody responses to anti-IL-21R antibody therapeutics following IV administration to cynomolgus monkeys. *J Transl Med.* 2010;8(1):41. doi:10.1186/1479-5876-8-41.
23. Vugmeyster Y, Guay H, Szklut P, Qian MD, Jin M, Widom A, Spaulding V, Bennett F, Lowe L, Andreyeva T, et al. In vitro potency, pharmacokinetic profiles and pharmacological activity of optimized anti-IL-21R antibodies in a mouse model of lupus. *MAbs.* 2010;2(3):335–46. doi:10.4161/mabs.2.3.11850.
24. Matlab R2010b. Natick (MA): Mathworks Inc; 2010. p. 2010.
25. Xu Y, Roach W, Sun T, Jain T, Prinz B, Yu T-Y, Torrey J, Thomas J, Bobrowicz P, Vasquez 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.* 2013;26(10):663–70. doi:10.1093/protein/gzt047.
26. Lou J, Marzari R, Verzillo V, Ferrero F, Pak D, Sheng M, Yang C, Sblattero D, Bradbury A. Antibodies in haystacks: how selection strategy influences the outcome of selection from molecular diversity libraries. *J Immunol Methods.* 2001;253(1–2):233–42. doi:10.1016/S0022-1759(01)00385-4.
27. Sharma VK, Patapoff TW, Kabakoff B, Pai S, Hilario E, Zhang B, Li C, Borisov O, Kelley RF, Chorny I, et al. In silico selection of therapeutic antibodies for development: viscosity, clearance, and chemical stability. *Proc Natl Acad Sci U S A.* 2014;111(52):18601–06. doi:10.1073/pnas.1421779112.
28. Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol.* 2015;427:478–90. doi:10.1016/j.jmb.2014.09.026.
29. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology.* 2018;154:394–406. doi:10.1111/imm.12889.
30. Jawa V, Cousens LP, Awwad M, Wakshull E, Kropshofer H, De Groot AS. T-cell dependent immunogenicity of protein therapeutics: preclinical assessment and mitigation. *Clin Immunol.* 2013;149:534–55. doi:10.1016/j.clim.2013.09.006.
31. Mason DM, Friedensohn S, Weber CR, Jordi C, Wagner B, Meng S, Gainza P, Correia BE, Reddy ST. Deep learning enables therapeutic antibody optimization in mammalian cells by deciphering high-dimensional protein sequence space. *bioRxiv.* 2019;617860.
32. Finlay WJ, Cunningham O, Lambert MA, Darmanin-Sheehan A, Liu X, Fennell BJ, Mahon CM, Cummins E, Wade JM, O'Sullivan CM, et al. Affinity maturation of a humanized rat antibody for anti-RAGE therapy: comprehensive mutagenesis reveals a high level of mutational plasticity both inside and outside the complementarity-determining regions. *J Mol Biol.* 2009;388:541–58. doi:10.1016/j.jmb.2009.03.019.
33. Cummins E, Luxenberg DP, McAleese F, Widom A, Fennell BJ, Darmanin-Sheehan A, Whitters MJ, Bloom L, Gill D, Cunningham O, et al. A simple high-throughput purification method for hit identification in protein screening. *J Immunol Methods.* 2008;339(1):38–46. doi:10.1016/j.jim.2008.07.016.

34. Pini A, Viti F, Santucci A, Carnemolla B, Zardi L, Neri P, Neri D. Design and use of a phage display library. Human antibodies with subnanomolar affinity against a marker of angiogenesis eluted from a two-dimensional gel. *J Biol Chem.* 1998;273:21769–76. doi:10.1074/jbc.273.34.21769.
35. Barbas CF 3rd, Bain JD, Hoekstra DM, Lerner RA. Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc Natl Acad Sci U S A.* 1992;89:4457–61. doi:10.1073/pnas.89.10.4457.
36. Finlay WJ, Bloom L, Cunningham O. Optimized generation of high-affinity, high-specificity single-chain Fv antibodies from multi-antigen immunized chickens. *Methods Mol Biol.* 2011;681:383–401.
37. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27:2957–63. doi:10.1093/bioinformatics/btr507.
38. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, Bolotin DA, Staroverov DB, Putintseva EV, Plevova K, et al. Towards error-free profiling of immune repertoires. *Nat Methods.* 2014;11:653–55. doi:10.1038/nmeth.2960.
39. Sirin S, Apgar JR, Bennett EM, Keating AE. AB-bind: antibody binding mutational database for computational affinity predictions. *Protein Sci.* 2016;25:393–409. doi:10.1002/pro.2829.
40. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol.* 2002;320:369–87. doi:10.1016/S0022-2836(02)00442-4.
41. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res.* 2005;33:W382–8. doi:10.1093/nar/gki387.