

CDR<sub>x</sub>Ab: Antibody Small-Molecule Conjugates  
with Computationally Designed  
Target-Binding Synergy

Thesis by  
Jingzhou Wang

In Partial Fulfillment of the Requirements for  
the degree of  
Doctor of Philosophy

The Caltech logo, featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2021  
(Defended May 13<sup>th</sup>, 2021)

© 2021

Jingzhou Wang  
ORCID: 0000-0002-2850-330X

## ACKNOWLEDGEMENTS

It has been a wonderfully exciting experience to learn and practice science as a Caltech graduate student. I owe most of the experience to my advisor, Prof. Stephen L. Mayo. Thank you for your endless support, encouragement, and guidance. Thank you for the insightful discussions, and for never turning me away when I showed up unexpectedly in your office with either frustrating or exciting news. Your mentorship is what truly supported me throughout my PhD career.

I want to thank Barry Olafson and Paul Chang from Protabit. Thank you for sharing your expertise and for guiding me through a field in which I had almost no experience. Collaborating with you on the projects throughout the years is an experience that I will always cherish. I want to thank Jost Vielmetter from PEC. You are the best biochemist I know and discussing with you is always an enlightening experience. Thank you for teaching me on SPR and all others supports. My research project would never go this far without your help. I also want to thank Justin Chartron from Protabit. Although we have only been collaborating since last year, working with you on some of the most urgent and difficult problems has been a great scientific journey.

I would like to thank all current and past Mayo Lab members throughout the years, including Seth Lieblich for the initial discussion on the conjugation project, Sarah Gillespie for showing me around the lab when I first joined, Aiden Aceves for the discussions and collaborations throughout the years, and Shan Huang for not making me work alone in late night. I would like to especially thank our lab manager Monica Breckow. You are our lab superhero and the best co-worker I could ever hope for. Your help was indispensable in helping my projects move forward.

I want to thank my other committee members: Prof. Pamela Bjorkman, Prof. Shu-ou Shan, and Prof. Frances Arnold. I feel extremely lucky to have the very experts on the crucial aspects of my project in my thesis committee. Thank you for your comments and suggestions throughout the years. Every discussion with you enabled me to think deeper about my work.

In addition, I would like to thank Annie Moradian, Mike Sweredowski, and Caltech PEL in general for teaching me about protein mass spectrometry, Mike Anaya for helping me with FPLC experiments, and Jennifer Keeffe, Andrew Flyak, and Beth Huey-Tubman for allowing me to use a variety of instruments. You made me truly appreciate the collaborative and supportive nature of the Caltech community.

Finally, I want to thank my family for the unconditional support. We have been separated by a Pacific Ocean of distance for most of the time since I was 18, but you'll always be the most important part of my life.

## ABSTRACT

Antibody-drug conjugates (ADCs), or chimeric modalities in general, combine the advantages and offset the flaws of their constituent parts to achieve a broader target space than traditional approaches of pharmaceutical development. My project combines the concept of ADCs with the full atomic simulation capability of computational protein design to define a new class of molecular recognition agents: CDR-extended antibodies, abbreviated as CDRxAbs. A CDRxAb incorporates a small-molecule binding event into *de novo* designed antibody/target interactions, creating antibody small-molecule conjugates that bind tighter against the target of the small molecule than the small molecule itself. In a proof-of-concept study using monomeric streptavidin/biotin pairs at either a nanomolar or micromolar-level affinity, nanobody-biotin conjugates were efficiently designed to exhibit >20-fold affinity improvement against the model protein targets, with stepwise optimization of binding kinetics and the overall stability. A yeast display-based workflow was subsequently developed to further improve the off rate of the best designed conjugate by another 6 folds. By fully incorporating the chemical space of immunoglobulins into the optimization of small molecule binding events, the workflow explored in this work could be potentially used as a generalizable new method to optimize small molecule-based therapeutics, by exploring a previously uncharted chemical space and the related target space. Chapter 1 reviews background information to justify the proposed CDRxAb molecular construct. Chapter 2 documents the detailed computational design process that generated the 10 conjugates, of which the characterization and discussion are elaborated in Chapter 3. Appendix I documents a slightly related ongoing work that uses computational design to improve existing antibody therapeutics.

## PUBLISHED CONTENT AND CONTRIBUTIONS

J. Wang, A.J. Aceves, S.L. Mayo, CDRxAb: Antibody Small-Molecule Conjugates with Computationally Designed Target-Binding Synergy (In Submission)

Contribution: J. Wang participated in the conceptualization of the project and designed the study. J. Wang performed all computational work except for MD simulation and rotamer library generation/measurement. J. Wang performed all experiments, analyzed and interpreted all computational and experimental data, and prepared the first draft of the manuscript, which was reformatted into the majority of Chapter 2 and 3 of this thesis.

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract .....	v
Published Content and Contributions.....	vi
Table of Contents.....	vii
List of Illustrations and/or Tables.....	viii
Chapter I: Introduction .....	1
Section A: the advantages and flaws of small molecules and monoclonal antibodies as therapeutic agents.....	1
Section B: antibody-drug conjugates (ADCs) and a proposed new format of molecular recognition agent .....	5
Section C: review of experimental and computational methods that may realize the new format of ADC.....	8
Section D: study design and summary of research outcome.....	18
Bibliography.....	13
Chapter II: The Computational Design Process of Generating Co-Binding Nanobody-Biotin Conjugates.....	23
Introduction .....	24
Section A: the design process of 4NBX.B-derived nanobody-biotin conjugates.....	24
Section B: the design process of 2X89.A-derived nanobody-biotin conjugates.....	35
Section C: nanobody sequences that are used in this study for docking and binding pose selection .....	39
Bibliography.....	52
Chapter III: Experimental Validation, Evolution, and Discussion of Designed Nanobody-Biotin Conjugates.....	53
Introduction .....	54
Results and Discussion.....	56
Conclusion.....	66
Bibliography.....	94
Appendix I: Computational Design and Experimental Screening of Neutralizing Antibody Variants against SARS-CoV-2 .....	98
Introduction .....	99
Results and Discussion.....	100
Materials and Methods.....	107
Bibliography.....	119

## LIST OF ILLUSTRATIONS AND/OR TABLES

<i>Number</i>	<i>Page</i>
Chapter 1	
1.1 A proposed new method of optimizing small molecule-based binders .....	14
1.2 A schematic representation of molecular architecture of IgG antibodies ..	15
1.3 Alignment of WT streptavidin and monomeric mSA streptavidin .....	16
1.4 A schematic representation of VHH single domain antibody fragments (nanobodies) in comparison with full length IgG.....	17
Chapter 2	
2.1 The schematic representation of the envisioned workflow .....	40
2.2 CDR conformational sampling by loop modeling and by adapting from naturally occurring structures.....	41
2.3 Selecting docked poses against streptavidin S45A/T90A/D180A by three filtering steps to identify potentially realizable binding poses .....	42
2.4 The final seven poses that passed the filtering steps .....	43
2.5 The final prepared 4NBX.B-biotin103 structure in complex with the target protein recapitulates the original 4NBX.B binding mode .....	44
2.6 Alignment results for prepared structures of 4NBX.B-biotin103 in complex with streptavidin S45A/T90A/D180A and mSA.....	45
2.7 Summary of first round of sequence design calculation of 4NBX.B-biotin103 .....	46
2.8 Pose identification and conjugate structure optimization of 2X89.A-CCAA-biotin57 against mSA .....	47
2.9 A rudimentary sequence design pipeline that performs CDR and framework design in a step wise manner.....	48
Table S1 Top 20 sequence outputs from CDR design of site 31, 32, 104, and 105 on 4NBX.B-biotin103 WT.....	49

Table S2 Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v186 .....	50
Table S3 Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v149 .....	51
Chapter 3	
3.1 Computationally designed nanobody-biotin conjugates bind stronger than biotin itself against mSA streptavidin .....	76
3.2 CDR sequence design enhanced the mSA-binding affinity and kinetics of 4NBX.B-biotin103 .....	78
3.3 CDR sequence design followed by framework design monomerically stabilized the designed conjugates without imposing affinity penalty .....	79
3.4 MD simulation reveals design flaws and validates design success .....	81
3.5 Flow cytometry analysis of v186_Fr WT displayed on yeast surface against WT and mSA streptavidin .....	83
3.6 Isolation and validation of a evolved variant of v186_Fr .....	84
3.6 Summary of the design results for mSA-targeting CDRxAbs and the overall design workflow .....	85
S1 Biotin conjugation was performed by biotin C2 maleimide with mutated cysteine residues .....	86
S2 An initial MD simulation that inspired us to move towards framework re-design. ....	87
S3 Additional Supporting SEC traces .....	88
S4 Summary of MD simulations performed for 4NBX.B-biotin103 v186 and v186_Fr against mSA <sub>WT</sub> .....	89
S5 The rest of SPR measurements used to estimate affinity and kinetic profiles of designed conjugates .....	90
S6 Intact-protein mass spectrometry (MS) confirmed mono-conjugated materials .....	91

S7 Designing and testing of a nanobody scaffold obtained by directly docking against mSA.....	92
Table S1 amino acid sequences for the protein templates used in this study...	93

## Appendix I

1 Single-site saturation analysis results of C002 CDR sites against WT SARS-CoV2 RBD .....	109
2 Analysis of C002 combinatorial design results against WT SARS-CoV-2 RBD .....	110
3 Single-site saturation analysis results of C002 CDR sites against B.1.351 SARS-CoV2 RBD .....	112
4 Analysis of C002 combinatorial design results against B.1.351 SARS-CoV-2 RBD.....	113
5 Single-site saturation analysis results of C118 CDR sites against WT SARS-CoV2 RBD .....	115
6 Analysis of C118 combinatorial design results against WT SARS-CoV-2 RBD .....	116
7 Designed C002 variants were isolated by FACS to confirm design effectiveness.....	118

## Chapter 1

### INTRODUCTION

#### ***Section A: the advantages and flaws of small molecules and monoclonal antibodies as therapeutic agents***

Most pharmaceutical mechanism of action involves a tight and specific molecular recognition event [1]. The binder molecule interacts with its target, most frequently a protein, to inhibit or activate certain biological processes that would change the course of disease [1].

Pharmaceutical development primarily focuses on two types of binders: synthetic small molecules and monoclonal antibodies (mAbs) [1]. Small molecules further represent over 90% of all marketed drugs, and also the great majority of molecules that have ever entered the drug development pipeline [2]. Attributed to their versatility and over a century of usage, small molecules have been developed against proteins that differ widely by function and localization, resulting in a rich repertoire of molecules with exciting chemical and biological properties. Unfortunately, only a very small number of molecules from that repertoire are qualified to be used in treatments. From 2000 to 2014, less than 10% of small molecule drug candidates that entered clinical trials were eventually approved [3], while clinical stage development usually has higher success rate than pre-clinical stage [4].

Although multiple reasons beyond the level of binder-target interactions play vital roles in drug failure, such as unrepresentative pre-clinical models and incomprehensive understanding of disease pathology [5], limitations in developing small molecule-based binders do appear as a bottleneck. Previous analysis indicated that most drug development programs were terminated due to failed identification of promising lead molecules [6], while intrinsic difficulties also exist to prevent small molecules from being engineered to bind tightly against arbitrary targets. In general, small molecule drugs conform to the “rule of five”, which was first described by Lipinski *et al.* and indicates that in order to maintain

good solubility and permeability, small molecules need to obey certain limits in size, lipophilicity, and the number of N and O atoms [2,7]. Taking the Lipinski rules and other limits into consideration, including that small molecule targets need to have concave binding pockets to likely establish tight interactions, Hopkins *et al.* estimated that a limited ~10-14% of human proteome are suitable small molecule targets [2]. Indeed, proteins outside of the estimated “druggable” realm have been less tractable by small molecules, exemplified by the challenges of developing potent small molecule inhibitors of protein-protein interactions (PPIs), which are generally mediated by flatter surface [8].

The small molecule target space is also characterized by its crowdedness [9]. Gao *et al.* surveyed ~20,000 binding pockets and concluded that the corresponding structural space can be categorized into ~1,000 representative shapes [9]. The high similarity among small molecule binding sites likely contribute to the observation that small molecule drugs often exhibit low binding specificity and promiscuous target selectivity [10]. Although multi-specific binders can have positive therapeutic effects through beneficial polypharmacology [11], low specificity has been frequently associated with safety concerns, which indeed are a major contributor (~17%) to clinical-stage drug failure [5,10,12]. Meanwhile, a limited and crowded target space may further contribute to the increasingly higher cost and lower success rate of drug development [13]. A survey of existing small molecule GPCR drugs showed that 10.3 molecules in average were approved to target an identical protein, suggesting a significant degree of overlap in past drug development programs [14]. Investing resources into similar mechanism of action may lead to the difficulty of showing benefits in later pharmaceuticals, the so-called “better-than-Beetles problem” [13]. In fact, failed demonstration of efficacy is the most frequent reason behind failed clinical trials [5].

To counter the above-mentioned challenges, pharmaceutical development has increasingly focused on novel drug targets in recent years [15]. In 2019, first-in-class drugs with novel mechanism of action constituted 42% of new approvals, although orphan drugs that were developed against rare diseases made the main contribution [16]. In concordance with this trend, it would be fundamentally beneficial to develop more efficient ways to explore new

targets, especially through methods that allow functional binders to be more easily developed against challenging proteins. One related effort in small molecule drug development is searching unconventional or uncharted chemical space to bring in new functional properties. Several successful small molecule PPI inhibitors were discovered among candidates that were unconventionally large and rich in hydrophobic rings to gain structural rigidity and broader contacting area, although at a cost of suboptimal pharmacokinetics [17,19]. A similar trend is behind the PROTACs, which are also chemically novel small molecules that use bispecific interactions to recruit protease activity to compensate for the relatively weak and transient engagement with difficult targets [18]. These examples suggest that chemical space expansion is a promising way to lead to target space expansion, by bringing in new properties that would be more suitable for certain functions. This idea is also reflected in the method proposed in the next section.

Monoclonal antibodies (mAbs) harbor unique properties that allow them to perform better than small molecules in many areas. Through a binding mechanism facilitated by much larger contacting surface, mAbs usually demonstrate superior binding specificity and overall better biocompatibility [10]. The different binding properties have led mAbs to be a more frequently chosen modality for some difficult targets, such as PPIs [20]. Meanwhile, benefiting from the large size and FcRn-mediated recycling mechanism, mAbs can have a circulation half-life up to weeks [21]. By contrast, small molecules are often cleared from body within hours and require frequent dosage [22]. In addition, the spanned binding interface may also allow mAbs to be better active site inhibitors, through the associated improvement of binding specificity [23].

Despite the advantages and higher clinical stage success rates than small molecules in recent years [3], approved mAbs has minimally overlapped with small molecules in target space. In addition to antibodies' intrinsic inability to target intracellular proteins, current mAbs and small molecules also target vastly different groups of membrane proteins [2,14,24]. For example, GPCRs are the biggest class of membrane proteins and a major type of small molecule targets (~25%) [2,14], but only two antibody drugs against GPCRs

have been approved [25], although GPCRs as drug targets remain popular and under-explored [14,26]. Another major type of small molecule targets, ion channels, still does not have any mAb passing phase I clinical trials [25]. The discrepancies among target preference are largely due to the lack of similarity between how the two modalities are developed. For small molecules, cell-based phenotypic assays are routinely used in early screens [27], while for mAbs immunization is still the predominant method to identify early hits [28], and of course is incompatible with established phenotypic screening methods both by format and throughput. Therefore, whether mAbs can be reliably generated largely depends on whether the target protein can be produced in large quantity and remain stable [28,29]. As the result, currently multi-pass transmembrane proteins such as GPCRs and ion channels are difficult mAb targets [28-30].

The same issue lies behind the challenge of developing site-specific antibodies, which frequently require the desired epitope to remain in stable and native-like conformation as isolated constructs [31,32]. By contrast, developing site-specific small molecules has benefited from the advancements of virtual screening and structure-based design in both hits discovery and hit-to-lead optimization stages [33-35]. Whereas for mAbs, such computational methods are not as robust, and practically have not outperformed immunization and display based approaches [36,83]. Furthermore, eliciting human mAbs against binding pockets seems to be difficult, also evidenced by examples of isolating very tight and specific antibodies that only partially block active sites of interest [23,37].

In summary, small molecules and mAbs each have distinctive advantages and disadvantages, while the respective targets and development processes are in minimal communication. Exploring avenues of crosstalk between the two major modalities might lead to new mechanism of action that offers new solutions to existing challenges. This effort is so far best demonstrated by a chimeric modality called antibody-drug conjugates (ADCs).

## ***Section B: antibody-drug conjugates (ADCs) and a proposed new format of molecular recognition agent***

ADCs combine the advantages and offset the flaws of existing small molecule and mAb drugs in order to achieve synergistic therapeutic effects [38]. The small molecule conjugation is achieved by functionalizing certain chemical groups, frequently primary amino or sulfhydryl moieties, of the antibody component [39]. The conjugates first bind to the antibody target, which is usually a cell type-specific membrane receptor, and then trigger receptor-induced endocytosis [38,40]. The antibody component is then digested in cytosol to release the conjugated small molecule, which is usually a potent toxin that does not discriminate cell types and has very low therapeutic index on its own [38,40]. Through antibody conjugation, the toxin can therefore induce cytotoxicity in a much more selective manner and demonstrate enhanced therapeutic benefits of drugging the corresponding target [38,40].

To date, ADCs are primarily used as a controlled-release method by combining small molecules and mAbs that have been separately developed against different targets [38,40]. While in action, the direct binding activity of each component is also not affected by the other [40,41]. Since conjugation does not change the fundamental mechanism of action for the constituent parts, current ADC research largely focuses on controlling the conjugation sites, quantity of payloads, and release mechanism through improved linker chemistry and the involvement of non-canonical amino acids [42], whereas the molecular-level interactions among the small molecule, linker, and antibody side chains are under-explored. However, these interactions do appear to be associated with various pharmacologically relevant properties of both the small molecule and antibody components [43-48]. Antibodies with conjugated small molecules frequently display higher aggregation propensity, which can be tuned by varying linker composition and conjugation sites/quantity [43,44]. Su *et al.* from Genentech showed that antibody conjugation could modulate the metabolic stability of small molecules through various steric shielding effect surrounding conjugation sites [45]. Another Genentech group also reported that the

chemical composition of neighboring side chains can directly affect the reactivity of conjugated chemical moieties [46]. Small molecule drugs are often hydrophobic, and some studies enhanced water solubility of small molecules through rationally designed peptide-drug conjugation, a concept similar to ADCs [47]. The physicochemical properties of conjugated small molecules in ADCs can also be altered by different linker choices [48]. Finally, small molecules usually demonstrate longer circulation half life upon antibody conjugation, as ADCs in general show antibody-like pharmacokinetics [41].

While the above examples showed the potential of using antibody conjugation to modulate physicochemical behaviors of small molecules and *vice versa*, a recent study of designing ADCs to bind to an identical target explored a deeper level of cooperation between antibodies and small molecules [37]. Cheng *et al.* from Amgen used a co-crystal structure of a small molecule drug sitagliptin and a separately developed antibody in complex with the shared target protein DPP-IV to rationally design ADCs that synergistically inhibit this enzyme [37,49]. The antibody was isolated to bind tightly and specifically against DPP-IV, but only partially blocked the active site of interest, and was therefore not functionally sound [37,49]. Sitagliptin is a marketed DPP-IV inhibitor but has limited efficacy and undesired side effects [37]. The authors used molecular dynamics (MD) simulation to search antibody residue positions and optimal lengths of PEG linkers to connect sitagliptin with minimum energetic penalty [37]. The resulted conjugates showed 13 to 32-fold improvements in  $IC_{50}$  than sitagliptin itself [37]. In this work, antibody interaction is used to enhance the target-inhibition effect of small molecules, showing that ADC technology could potentially optimize the direct mechanism of action of small molecule drugs.

Based on the promise from the above-mentioned studies, this thesis proposes that antibody conjugation could be used as a general methodology to modulate the binding behavior and other pharmacologically relevant properties of small molecule-based binders. Specifically, instead of combining both existing small molecules and antibodies, the proposed format of ADC focuses on existing small molecules only and uses structure-based protein engineering methods to select an appropriate antibody scaffold/sequence, whose

conjugation with the small molecule will create a chimeric modality that binds better against the small molecule target. Besides the binding behavior, other physicochemical properties of the conjugates should also be directly modulated by the sequence of the antibody component. As the result, the chemical space of antibodies can be directly brought into the development process of small molecule drugs.

This new format of ADC echoes with some existing patterns of small molecule drug development. Crystallography is playing an increasingly important role in both hits discovery and hit-to-lead optimization steps, generating an abundance of co-crystal structures for protein targets and small molecule binders [33-35]. Structure-based hits discovery works by soaking protein crystals with small molecule fragments, and the identity of bound molecules can then be revealed by crystallography, leading to the isolation of binders up to low mM level affinity [33-35]. Crystallography is also involved in structure-based hit-to-lead optimization steps, providing co-crystal structures with  $\mu\text{M}$  to nM level small molecule binders [33-35]. By using these structures, the proposed format of antibody conjugation can participate in various stages of molecular discovery, starting from small molecules with mM,  $\mu\text{M}$ , and nM level affinities (Figure 1.1). Potentially, an increased success rate can be attained through the resulted chemical space expansion. This new format of ADC may also be used to rescue failed small molecule drugs. Exploiting previous drug molecules has been an ongoing area of research, although mainly focusing on repurposing approved drugs instead of resurrecting failed small molecules [50]. Again, through an expanded chemical space, structure-based engineering of antibody conjugation may also bring drug-like properties to previous failed small molecule binders, leading to more efficient utilization of available molecular repertoires.

***Section C: review of experimental and computational methods that may realize the new format of ADC***

Despite the projected benefits of this potential modality, the main question is whether existing technical capacities are enough to bring this new ADC format to reality. The technical obstacles are obvious. First, immunization cannot be used to isolate appropriate antibody sequences for conjugation, because the small molecules of interest simply do not exist in canonical antibody sequences. Display-based approaches are widely used for antibody engineering, but current ADC conjugation chemistry primarily utilizes functional groups that are ubiquitous among protein families [39]. As a result, no matter whether yeast or phage display is used to engineer the antibody component, upon small molecule conjugation, there will likely be significant amount of background labeling that masks the desired binding events. Nevertheless, very recently Islam *et al.* reported a non-canonical amino acid (NcAA) based yeast display system that allows appropriate differentiation of displayed conjugates/target interactions [51], indicating that NcAA based protein engineering workflow might be a good direction to pursue. Integration of NcAA has been better explored in peptide engineering, since synthetic amino acid building blocks can be introduced by peptide-building chemistry [52]. Gates *et al.* engineered ~30aa long “xenoproteins” with non-natural side chains in randomized regions, and used MS/MS sequencing to identify hit binders against targets [52]. However, peptide synthesis is limited by length and a typical antibody domain with >100 amino acids is a challenge for current technical capacity [52]. Moreover, MS/MS sequencing also requires certain restrictions on the length and complexity of randomized region to unambiguously report sequence identity [52].

Structure-based computational design methods may offer the most appropriate solutions to the initial engineering of the new synergistically-binding ADC format. Computational protein design (CPD) works by searching the sequence and conformational space of polypeptides to identify 1) structures that lead to the desired functions and 2) amino acid sequences that support the structures, through atomic-level simulations and energy-based

scoring schemes [53,54]. The most prominent benefits of CPD would be allowing an enormous amount of potential sequence choices to be screened without experimental investment [53].

Because engineering the new ADC format will involve creating new antibody/target interface, CPD methods for protein-protein interaction (PPI) would be most relevant. The initial successes of PPI design focused on modeling mutations in existing binding interface, creating mutants with altered binding specificity [55-57,63]. For antibody engineering, the similar approaches were used in nM to pM affinity maturation of existing antibody/target complexes [58]. Lippow *et al.* simulated single amino acid mutations in two antibody/target complexes, tested the single mutations that were predicted to be beneficial, combined the experimentally confirmed mutations, and finally obtained variants with respective ~10 and ~100 folds affinity improvement [58]. Based on the experience of interface redesign, a series of scoring criteria, including overall interaction energy score, hydrophobicity, shape complementarity, certain structural features, and specific types of force such as ionic interactions, were proposed to aid the assessment of mutation choices [58-62]. Debating over which preference should be given over the metrics has been going on in later redesign and *de novo* design works, as design protocols using similar metrics have succeeded and failed in different studies [64,65,70,71,77]. However, there is a more universal agreement over what are the energetic penalties that would jeopardize design success, including buried un-paired charges, solvent-inaccessible un-satisfied H-bond donors/acceptors, exposure of large hydrophobic groups, and cavities [65-69].

The *de novo* PPI designs have largely focused on interactions facilities by ordered secondary structure elements [55,64,67,70,71-73]. In a pioneering study, Huang *et al.* designed monomeric GB1 protein into a heterodimer by a dock-and-design approach, which has become the general framework of *de novo* binding interface design [55]. Dock-and-design first optimizes the binding geometry between interacting components, and then performs sequence optimization of the docked interface [55,64,70]. The designed GB1 dimer only formed transient interactions, likely caused by fluctuation between

intramolecular and intermolecular interactions [55]. In a later study, dock-and-design resulted in the successful generation of a homodimer with certain protocol changes, which included utilizing a pre-engineered scaffold of hyper-stability and docking alanine-mutated surfaces [70]. Additional design successes were reported with modified docking protocols, mainly by including biases suggested by real-world structural evidence. Fleishmen *et al.* designed small proteins against influenza hemagglutinin at high nM/ $\mu$ M affinity by first identifying interaction hot-spots, which are sporadic residues that interact favorably with the target surface, and then docked candidate scaffolds (fetched from PDB) on the surface, with a bias to support the hot-spot residues [71]. Candidate scaffolds can also be generated by recombining structural fragments from PDB [68,72-74], allowing novel and realizable conformational space to be sampled for challenging design tasks. For example, Bale *et al.* docked recombinant scaffolds with geometric constraints, and then sequence designed the modeled interface to create stable and conformation specific icosahedrons [73]. Biases can also be introduced in the sequence design step to generate binders with interesting properties. Taking advantage of the structural rigidity of alpha helical bundles, Boyken *et al.* created the HBNet protocol to favor sequence choices that promote H-bond network formation and designed highly specific helix-helix interacting pairs, which were used to create interaction-based logic gates in follow-up studies [67,75].

Inaccuracies in energy evaluation and insufficient sampling of conformational space are correlated to non-ideal performances of designed proteins [55,64,71,76]. Therefore, directed evolution of designed protein binders has been an effective way to improve affinity [70,71,77]. By contrast, a recent study docked and designed recombinant scaffolds, expanded the screening size to  $10^4$  sequences, and isolated binders at low nM affinity, suggesting that experimentally testing an expanded and sequence-diversified pool of proteins may compensate for design inaccuracies [72]. In addition, advancements in deep sequencing and high throughput screening have allowed sequence-affinity landscapes to be measured and utilized to improve subsequent design efforts by machine-learning [78,79]. For instance, Jenson *et al.* measured the binding affinity of a designed peptide library against three Bcl-2 family proteins, and used the measured data to direct the design of

another library, from which highly specific and tight peptides were isolated against the individual Bcl-2 targets [78]. The above successes demonstrated that a myriad of functional computational methods, which incorporate real-world data and advanced understanding of structure-function relationship, is available for various PPI engineering problems.

*De novo* antibody design is much harder than designing interactions mediated by rigid structural elements. Antibodies interact with targets through complementarity determining regions (CDRs), which are hyper-variable structured loops supported by a relatively conserved  $\beta$ -barrel framework (Figure 1.2) [80]. In mAb, three CDR loops are on each of the heavy and light chains [80]. Loops are much more conformationally flexible than helices and sheets [36,81-83], and often adopt a dynamic ensemble of conformational states that deviate from the single state observed in crystal structures [81,82]. Nevertheless, modeling of antibody structures from sequences has been greatly facilitated by the identification of canonical conformations for most CDR loops [83,84,86,87]. However, the canonical structures of heavy chain CDR3 (CDRH3) loops have not been reliably reported [36,83], possibly due to their much longer length and unique interaction features [83,85]. Facing these challenges, *de novo* prediction of CDR loop structures and subsequent modeling of CDR/target interactions remain an ongoing area of research [36,83].

Due to the difficulties of modeling CDR loops from scratch, successful *de novo* antibody designs reported to date usually used intact CDR conformations from existing structures [66,77,89]. Nevertheless, PDB harbors a vast and diverse collection of antibody/target structures, and therefore provides a functionally proven sequence and conformational space for design calculations to sample from. Nimrod *et al.* docked intact antibody structures from PDB onto IL-17A, repurposed the antibody scaffolds by modeling mutations in the docked interface, and constructed a designed library of  $\sim 10^5$  diversity, from which a binder was isolated at 80 nM affinity against IL-17A after humanization [88]. A similar dock-and-design approach was used by Baran *et al.*, while the antibody scaffolds were recombined from segmented fragments of existing PDB structures, although native CDR loops remained intact in these fragments [66]. Successful designs were reported from  $\sim 10^3$

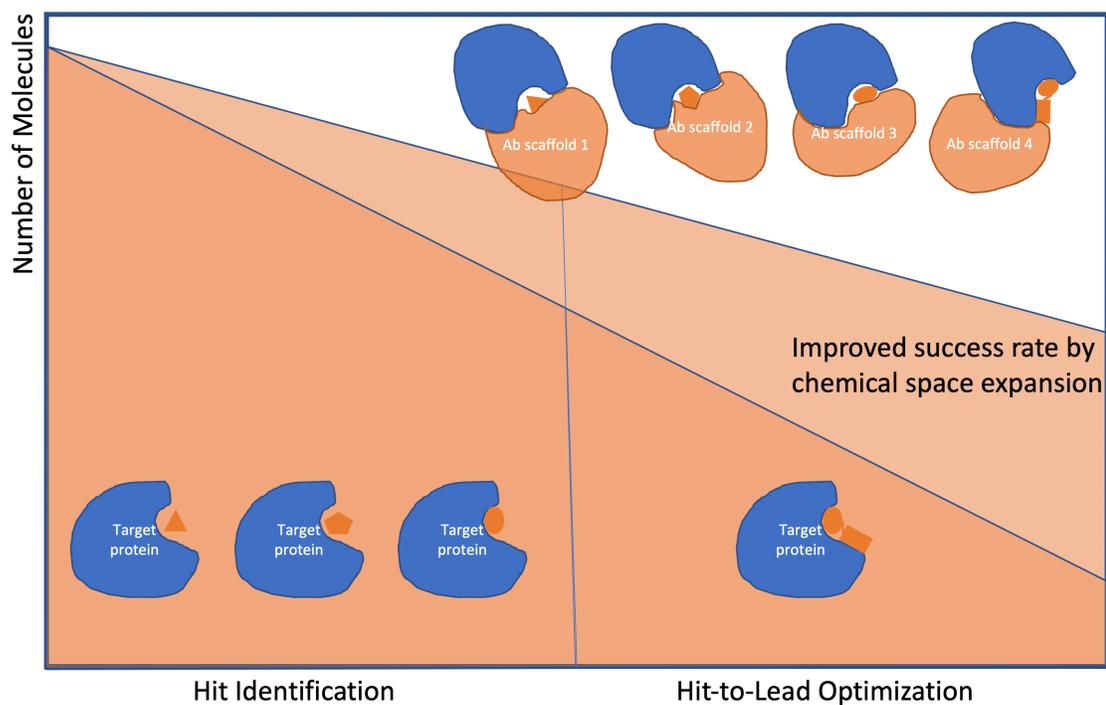
screened sequences, although directed evolution was performed to bring affinity below 100 nM [66]. A similar observation was reported by Dang *et al.* in designing an antibody-mimetic binder, as directed evolution rescued a handful of non-binding designs to pM affinity [77]. Importantly, Baran *et al.* reported that sequence restrictions derived by multi-sequence alignments, in their case PSSM score, increased the success rates of generating stable and functional antibodies [66]. One of the highest success rates was reported by Liu *et al.*, who used hot-spot biased docking to identify binding conformations and isolated a ~100 nM binder from 10 tested sequences [89]. In their work, hot-spot residues were fetched from a naturally occurring binding interface between Keap1 and Nrf2, and intact antibody scaffolds from PDB were docked against the Nrf2 binding surface, with a bias to support the hot-spot residues [89]. This high success rate is likely associated with the usage of naturally occurring interactions in the designed Ab/target interface [88,89], as other hot-spot centric designs that *de novo* modeled the hotspot residues usually required directed evolution or larger screening size to achieve high affinity [71,72]. Different from designs that exploit existing CDR conformations, successful antibody designs by constructing novel CDR structures were mainly reported against disordered peptides [90,91]. For example, Sormanni *et al.* designed antibodies against  $\alpha$ -synuclein and A $\beta$ 42 by piecing together small PDB fragments to build complementary paratope on a CDR loop [90]. Importantly, an existing antibody with insertion tolerant CDR was used as the starting scaffold in their study [90].

For this thesis and the design problem of interest, an antibody sequence should be designed against the target of a small molecule, so that upon conjugation with the small molecule, the antibody should interact with the target protein in a way to assist the small molecule binding event. Although this design scenario is not directly related to published design efforts, the successes discussed above showed that designing antibody/target interactions is a tractable problem. Therefore, by thorough analysis of the proposed design problem and careful selection of applicable design methods, constructing a functional design workflow should be attainable.

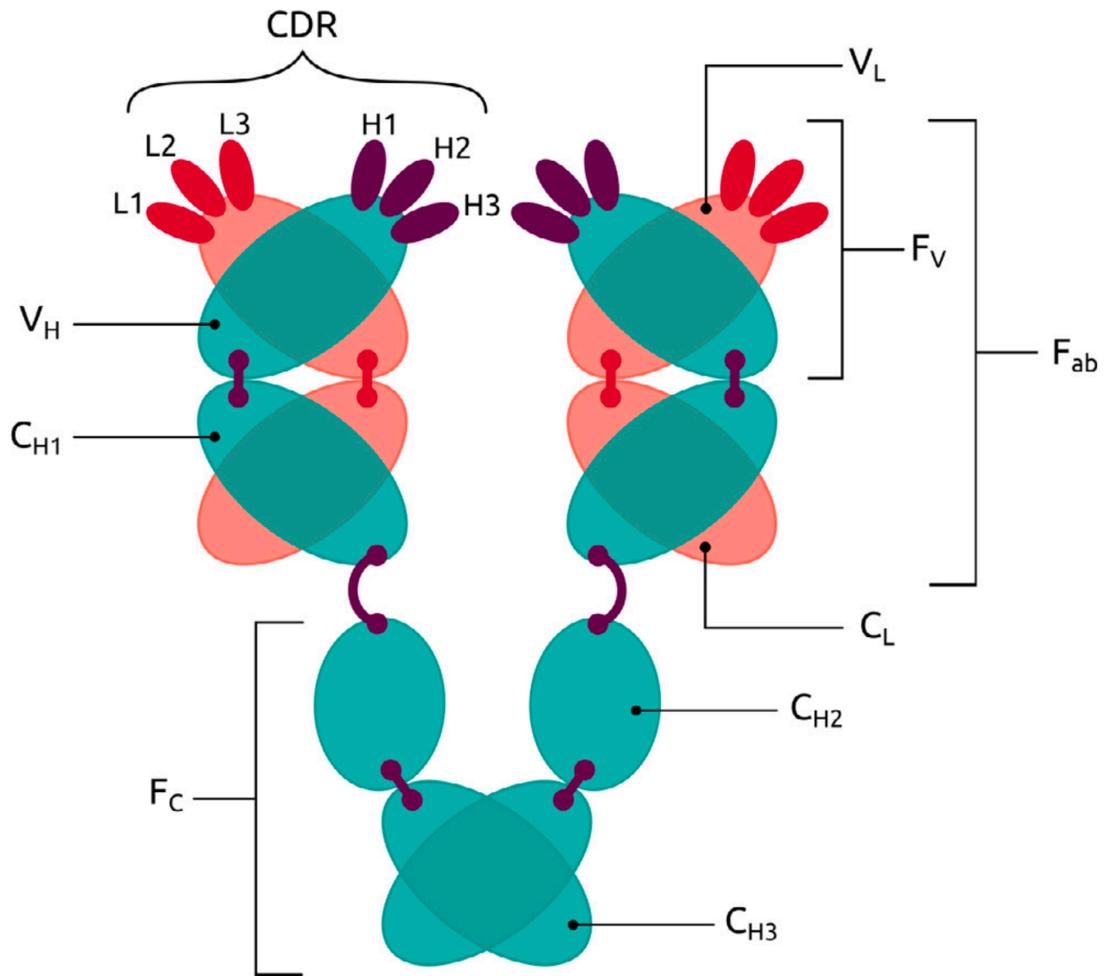
### ***Section D: study design and summary of research outcome***

In this thesis, a proof-of-concept study was conducted to build a computational design workflow for the proposed new format of ADC and demonstrate the workflow on a model system. Specifically, monomeric streptavidin and biotin were chosen as the model target and small molecule. This interaction pair was chosen for the following reasons. First, streptavidin/biotin interactions are extensively studied with reliable co-crystal structures for computational design [92]. Second, throughout the past two decades, all monomeric versions of streptavidin showed dramatically worsened binding affinity and dissociation rate [93-95]. The best monomeric streptavidin reported to date, mSA, has a  $10^5$ -fold reduction in affinity [93]. However, structural alignment showed that the binding pocket interactions in mSA/biotin almost overlap with WT streptavidin/biotin except for a flexible loop, indicating difficulties to further improve the small molecule/target interactions (Figure 1.3). Therefore, monomeric streptavidin and biotin are suitable as a protein and small molecule pair that is functionally sub-optimal and difficult to be improved by traditional methods.

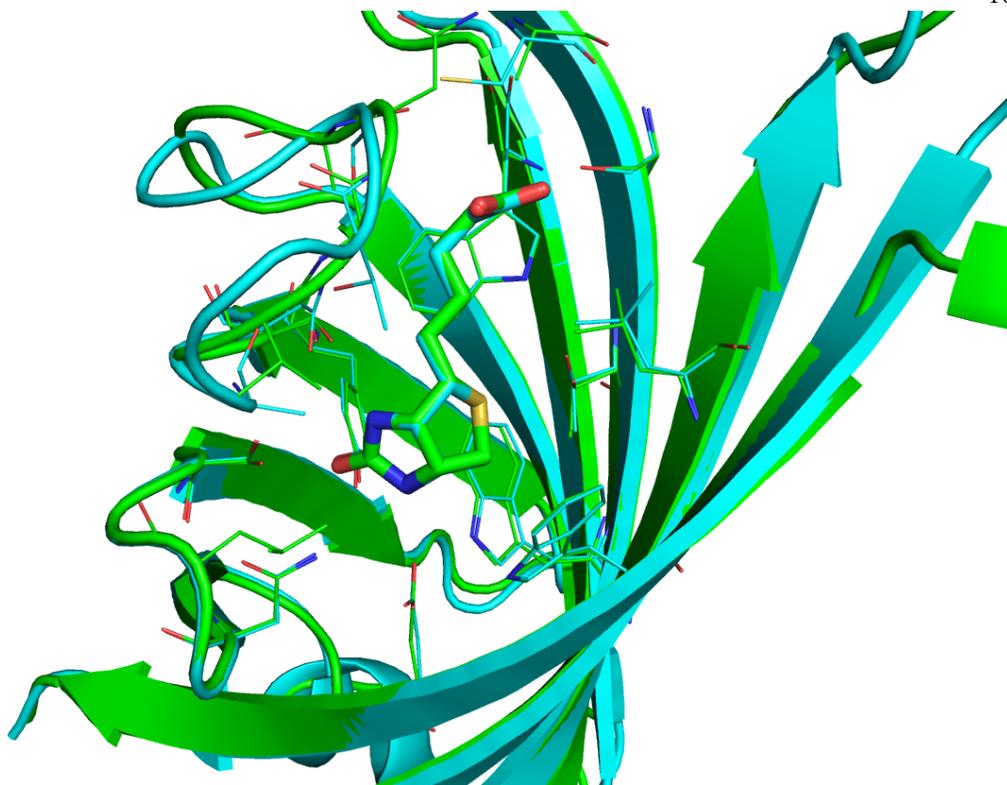
In this work, llama-derived single-domain antibody scaffolds (nanobodies) were chosen as the antibody component, for the easier computational and experimental manipulation in the initial method development stage (Figure 1.4) [96]. Nanobody-biotin conjugates were computationally designed to bind to mSA with >20-fold affinity improvement than biotin itself, through distinctively characterized kinetic enhancements. Designed conjugates were tested on protein/small-molecule pairs with nM and  $\mu$ M level initial affinities to demonstrate broad applicability. The aggregation profiles of the designed conjugates can also be directly improved by sequence design. Finally, a yeast-display based workflow was constructed to improve the designed conjugates, and a directed evolution trial successfully slowed the dissociation rate by another 6 folds. Overall, inspired by ADCs, this work defined a novel format of molecular recognition agent that brings in a previously irrelevant chemical space to optimize small molecule-based binders, an effort that echoes many needs from the pharmaceutical development process.



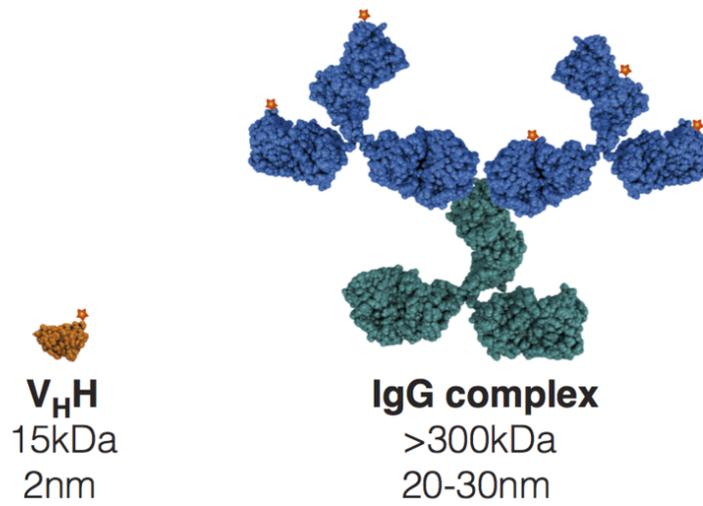
**Figure 1.1 A proposed new method of optimizing small molecule-based binders.** Taking advantage of the increased availability of high-resolution crystal structures of protein/small-molecule complexes in hit discovery and hit-to-lead optimization stages, antibody (Ab) scaffolds conjugating to small molecules in various stages of molecular discovery are proposed to be engineered in a structure-based manner, and the physicochemical property of the conjugated small molecules can be optimized by the antibody sequence. Ideally, the resulted chemical space expansion may lead to increased success rate of molecular discovery for pharmaceutical development.



**Figure 1.2** A schematic representation of molecular architecture of IgG antibodies (adapted from Roy *et al.*, 2017)[80]: The overall molecular composition of IgG antibodies is shown here. Each antibody molecule contains two heavy chains ( $V_H$ ,  $C_{H1}$ ,  $C_{H2}$ , and  $C_{H3}$ ) and two light chains ( $V_L$  and  $C_L$ ). On top of the hetero-dimeric  $F_{ab}$  fragments are the variable domains that interact with targets by six CDR loops L1, L2, L3, H1, H2, and H3.



**Figure 1.3 Alignment of WT streptavidin (PDB ID 1MK5) and monomeric mSA streptavidin (PDB ID 4JNJ) [92,93]:** One protomer of WT streptavidin is aligned globally with mSA. Cyan: crystal structure of mSA streptavidin. Green: crystal structure of WT streptavidin. Biotin molecules in two structures are shown as sticks. Residues that are within 4 Å distance from biotin in both structures are selected and shown as lines. Alignment of the selected residues in WT and mSA streptavidin returned an RMSD of 0.2 Å.



**Figure 1.4 A schematic representation of V<sub>H</sub>H single domain antibody fragments (nanobodies) in comparison with full length IgG (adapted from Chromotek website) [97].** Nanobodies are structurally similar to the V<sub>H</sub> domain of full-length IgG, but are functionally stable as single-domain fragments with comparable target-binding strength [96].

## Bibliography

1. D. C. Swinney, Biochemical mechanisms of drug action: what does it take for success? *Nature Reviews Drug Discovery* **3**, 801–808 (2004).
2. A. L. Hopkins, C. R. Groom, The druggable genome. *Nature Reviews Drug Discovery* **1**, 727–730 (2002).
3. K. Smietana, M. Siatkowski, M. Møller, Trends in clinical success rates. *Nature Reviews Drug Discovery* **15**, 379–380 (2016).
4. T. Takebe, R. Imai, S. Ono, The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. *Clin Transl Sci* **11**, 597–606 (2018).
5. D. B. Fogel, Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun* **11**, 156–164 (2018).
6. G. M. Keserü, G. M. Makara, The influence of lead discovery strategies on the properties of drug candidates. *Nature Reviews Drug Discovery* **8**, 203–212 (2009).
7. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3–26 (2001).
8. G. L. Verdine, L. D. Walensky, The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. *Clin Cancer Res* **13**, 7264–7270 (2007).
9. M. Gao, J. Skolnick, A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. *PLOS Computational Biology* **9**, e1003302 (2013).
10. K. Imai, A. Takaoka, Comparing antibody and small-molecule therapies for cancer. *Nature Reviews Cancer* **6**, 714–727 (2006).
11. E. Gilberg, S. Jasial, D. Stumpfe, D. Dimova, J. Bajorath, Highly Promiscuous Small Molecules from Biological Screening Assays Include Many Pan-Assay Interference Compounds but Also Candidates for Polypharmacology. *J. Med. Chem.* **59**, 10285–10290 (2016).
12. T. J. Hwang, *et al.*, Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. *JAMA Intern Med* **176**, 1826–1833 (2016).
13. J. W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* **11**, 191–200 (2012).
14. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, D. E. Gloriam, Trends in GPCR drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery* **16**, 829–842 (2017).
15. M. Hay, D. W. Thomas, J. L. Craighead, C. Economides, J. Rosenthal, Clinical development success rates for investigational drugs. *Nature Biotechnology* **32**, 40–51 (2014).
16. C. for D. E. and Research. “New Drug Therapy Approvals 2019.” *FDA*, 2020, <https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/new-drug-therapy-approvals-2019> (Accessed April 29, 2021).

17. D. Bojadzic, P. Buchwald, Toward Small-Molecule Inhibition of Protein–Protein Interactions: General Aspects and Recent Progress in Targeting Costimulatory and Coinhibitory (Immune Checkpoint) Interactions. *Curr Top Med Chem* **18**, 674–699 (2018).
18. H. Pei, Y. Peng, Q. Zhao, Y. Chen, Small molecule PROTACs: an emerging technology for targeted therapy in drug discovery. *RSC Advances* **9**, 16967–16976 (2019).
19. M. R. Arkin, Y. Tang, J. A. Wells, Small-molecule inhibitors of protein-protein interactions: progressing towards the reality. *Chem Biol* **21**, 1102–1114 (2014).
20. M. Bakail, F. Ochsenbein, Targeting protein–protein interactions, a wide open field for drug design. *Comptes Rendus Chimie* **19**, 19–27 (2016).
21. M. Pyzik, *et al.*, The Neonatal Fc Receptor (FcRn): A Misnomer? *Front. Immunol.* **10** (2019)
22. H. Gunaydin, *et al.*, Strategy for Extending Half-life in Drug Design and Its Significance. *ACS Med Chem Lett* **9**, 528–533 (2018).
23. D. H. Nam, C. Rodriguez, A. G. Remacle, A. Y. Strongin, X. Ge, Active-site MMP-selective antibody inhibitors discovered from convex paratope synthetic libraries. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14970–14975 (2016).
24. R. B. Dodd, T. Wilkinson, D. J. Schofield, Therapeutic Monoclonal Antibodies to Complex Membrane Protein Targets: Antigen Generation and Antibody Discovery Strategies. *BioDrugs* **32**, 339–355 (2018).
25. C. J. Hutchings, Mini-review: antibody therapeutics targeting G protein-coupled receptors and ion channels. *Antibody Therapeutics* **3**, 257–264 (2020).
26. C. J. Hutchings, A review of antibody-based therapeutics targeting G protein-coupled receptors: an update. *Expert Opinion on Biological Therapy* **20**, 925–935 (2020).
27. J. Eder, R. Sedrani, C. Wiesmann, The discovery of first-in-class drugs: origins and evolution. *Nature Reviews Drug Discovery* **13**, 577–587 (2014).
28. Z.-J. Lu, *et al.*, Frontier of therapeutic antibody discovery: The challenges and how to face them. *World J Biol Chem* **3**, 187–196 (2012).
29. R. B. Dodd, T. Wilkinson, D. J. Schofield, Therapeutic Monoclonal Antibodies to Complex Membrane Protein Targets: Antigen Generation and Antibody Discovery Strategies. *BioDrugs* **32**, 339–355 (2018).
30. C. J. Hutchings, M. Koglin, F. H. Marshall, Therapeutic antibodies directed at G protein-coupled receptors. *MAbs* **2**, 594–606 (2010).
31. G. Ofek, *et al.*, Elicitation of structure-specific antibodies by epitope scaffolds. *PNAS* **107**, 17880–17887 (2010).
32. B. E. Correia, *et al.*, Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic HIV Vaccine Epitope. *Structure* **18**, 1116–1126 (2010).
33. K. H. Bleicher, H.-J. Böhm, K. Müller, A. I. Alanine, Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, 369–378 (2003).
34. G. M. Keserü, G. M. Makara, Hit discovery and hit-to-lead approaches. *Drug Discovery Today* **11**, 741–748 (2006).
35. L. Maveyraud, L. Mourey, Protein X-ray Crystallography and Drug Discovery. *Molecules* **25** (2020).
36. J. C. Almagro, *et al.*, Second antibody modeling assessment (AMA-II). *Proteins: Structure, Function, and Bioinformatics* **82**, 1553–1562 (2014).

37. A. C. Cheng, *et al.*, Structure-guided Discovery of Dual-recognition Chemibodies. *Scientific Reports* **8**, 7570 (2018).
38. G. D. L. Phillips, *et al.*, Targeting HER2-Positive Breast Cancer with Trastuzumab-DM1, an Antibody–Cytotoxic Drug Conjugate. *Cancer Res* **68**, 9280–9290 (2008).
39. K. Tsuchikama, Z. An, Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein & Cell* **9**, 33–46 (2016).
40. Y. Feng, *et al.*, Conjugates of Small Molecule Drugs with Antibodies and Other Proteins. *Biomedicines* **2**, 1–13 (2014).
41. W. D. Hedrich, T. E. Fandy, H. M. Ashour, H. Wang, H. E. Hassan, Antibody–Drug Conjugates: Pharmacokinetic/Pharmacodynamic Modeling, Preclinical Characterization, Clinical Studies, and Lessons Learned. *Clinical Pharmacokinetics* **57**, 687–703 (2017).
42. H. L. Perez, *et al.*, Antibody–drug conjugates: current status and future directions. *Drug Discovery Today* **19**, 869–881 (2014).
43. B. Frka-Petesic, *et al.*, Aggregation of Antibody Drug Conjugates at Room Temperature: SAXS and Light Scattering Evidence for Colloidal Instability of a Specific Subpopulation. *Langmuir* **32**, 4848–4861 (2016).
44. W. Li, *et al.*, Antibody Aggregation: Insights from Sequence and Structure. *Antibodies* **5**, 19 (2016).
45. D. Su, *et al.*, Modulating Antibody–Drug Conjugate Payload Metabolism by Conjugation Site and Linker Modification. *Bioconjugate Chemistry* **29**, 1155–1167 (2018).
46. B.-Q. Shen, *et al.*, Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nature Biotechnology* **30**, 184–189 (2012).
47. Y. Wang, *et al.*, Peptide–Drug Conjugates as Effective Prodrug Strategies for Targeted Delivery. *Adv Drug Deliv Rev* **110–111**, 112–126 (2017).
48. R. Y. Zhao, *et al.*, Synthesis and Evaluation of Hydrophilic Linkers for Antibody–Maytansinoid Conjugates. *J. Med. Chem.* **54**, 3606–3623 (2011).
49. J. Tang, *et al.*, An Inhibitory Antibody against Dipeptidyl Peptidase IV Improves Glucose Tolerance in Vivo. *J. Biol. Chem.* **288**, 1307–1316 (2013).
50. S. Pushpakom, *et al.*, Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* **18**, 41–58 (2019).
51. M. Islam, *et al.*, Chemical Diversification of Simple Synthetic Antibodies. *ACS Chem. Biol.* **16**, 344–359 (2021).
52. Z. P. Gates, *et al.*, Xenoprotein engineering via synthetic libraries. *PNAS* **115**, E5298–E5306 (2018).
53. B. I. Dahiya, S. L. Mayo, De Novo Protein Design: Fully Automated Sequence Selection. *Science* **278**, 82–87 (1997).
54. B. Kuhlman, *et al.*, Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
55. P.-S. Huang, J. J. Love, S. L. Mayo, A de novo designed protein–protein interface. *Protein Science* **16**, 2770–2774 (2007).
56. J. M. Shifman, S. L. Mayo, Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* **323**, 417–423 (2002).
57. D. N. Bolon, R. A. Grant, T. A. Baker, R. T. Sauer, Specificity versus stability in computational protein design. *PNAS* **102**, 12724–12729 (2005).

58. S. M. Lippow, K. D. Wittrup, B. Tidor, Computational design of antibody-affinity improvement beyond *in vivo* maturation. *Nature Biotechnology* **25**, 1171–1176 (2007).
59. R. Barderas, J. Desmet, P. Timmerman, R. Meloen, J. I. Casal, Affinity maturation of antibodies assisted by *in silico* modeling. *PNAS* **105**, 9029–9034 (2008).
60. S. P. Mahajan, *et al.*, Computational affinity maturation of camelid single-domain intrabodies against the nonamyloid component of alpha-synuclein. *Scientific Reports* **8**, 1–14 (2018).
61. D. W. Sammond, *et al.*, Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* **371**, 1392–1404 (2007).
62. J. Karanicolas, B. Kuhlman, Computational Design of Affinity and Specificity at Protein-Protein Interfaces. *Curr Opin Struct Biol* **19**, 458–463 (2009).
63. J. M. Shifman, S. L. Mayo, Exploring the origins of binding specificity through the computational redesign of calmodulin. *PNAS* **100**, 13274–13279 (2003).
64. E. Procko, *et al.*, Computational Design of a Protein-Based Enzyme Inhibitor. *Journal of Molecular Biology* **425**, 3563–3575 (2013).
65. P. B. Stranges, B. Kuhlman, A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Science* **22**, 74–82 (2013).
66. D. Baran, *et al.*, Principles for computational design of binding antibodies. *Proceedings of the National Academy of Sciences*, 201707171 (2017).
67. S. E. Boyken, *et al.*, De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
68. N. Koga, *et al.*, Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
69. P. J. Fleming, G. D. Rose, Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci* **14**, 1911–1917 (2005).
70. Y. Mou, P.-S. Huang, F.-C. Hsu, S.-J. Huang, S. L. Mayo, Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10714–10719 (2015).
71. S. J. Fleishman, *et al.*, Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* **332**, 816–821 (2011).
72. A. Chevalier, *et al.*, Massively parallel *de novo* protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
73. J. B. Bale, *et al.*, Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
74. T. J. Brunette, *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
75. Z. Chen, *et al.*, Programmable design of orthogonal protein heterodimers. *Nature*, 1 (2018).
76. P. Mitra, D. Shultis, Y. Zhang, EvoDesign: de novo protein design based on structural and evolutionary profiles. *Nucleic Acids Res* **41**, W273–W280 (2013).
77. L. T. Dang, *et al.*, Receptor subtype discrimination using extensive shape complementary designed interfaces. *Nat Struct Mol Biol* **26**, 407–414 (2019).
78. J. M. Jenson, *et al.*, Peptide design by optimization on a data-parameterized protein interaction landscape. *PNAS*, 201812939 (2018).

79. R. M. Adams, T. Mora, A. M. Walczak, J. B. Kinney, Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife* **5**, e23156 (2016).
80. A. Roy, S. Nair, N. Sen, N. Soni, M. S. Madhusudhan, In silico methods for design of biological therapeutics. *Methods* **131**, 33–65 (2017).
81. M. L. Fernández-Quintero, J. Kraml, G. Georges, K. R. Liedl, CDR-H3 loop ensemble in solution – conformational selection upon antibody binding. *MAbs* **11**, 1077–1088 (2019).
82. J. R. Jeliazkov, *et al.*, Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification. *Front Immunol* **9** (2018).
83. S. Fischman, Y. Ofran, Computational design of antibodies. *Current Opinion in Structural Biology* **51**, 156–162 (2018).
84. A. Teplyakov, G. L. Gilliland, Canonical structures of short CDR-L3 in antibodies. *Proteins* **82**, 1668–1673 (2014).
85. C. Regep, G. Georges, J. Shi, B. Popovic, C. M. Deane, The H3 loop of antibodies shows unique structural characteristics. *Proteins* **85**, 1311–1318 (2017).
86. B. Al-Lazikani, A. M. Lesk, C. Chothia, Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* **273**, 927–948 (1997).
87. J. Adolf-Bryfogle, Q. Xu, B. North, A. Lehmann, R. L. Dunbrack, PyIgClassify: a database of antibody CDR structural classifications. *Nucleic Acids Res* **43**, D432-438 (2015).
88. G. Nimrod, *et al.*, Computational Design of Epitope-Specific Functional Antibodies. *Cell Reports* **25**, 2121-2131.e5 (2018).
89. X. Liu, *et al.*, Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping. *Scientific Reports* **7**, 41306 (2017)
90. P. Sormanni, F. A. Aprile, M. Vendruscolo, Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **112**, 9902–9907 (2015).
91. F. A. Aprile, *et al.*, Selective targeting of primary and secondary nucleation pathways in A $\beta$ 2 aggregation using a rational antibody scanning method. *Science Advances* **3**, e1700488 (2017).
92. D. E. Hyre, *et al.*, Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci* **15**, 459–467 (2006).
93. K. H. Lim, H. Huang, A. Pralle, S. Park, Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnology and Bioengineering* **110**, 57–67 (2013).
94. K. H. Lim, H. Huang, A. Pralle, S. Park, Engineered streptavidin monomer and dimer with improved stability and function. *Biochemistry* **50**, 8682–8691 (2011).
95. S.-C. Wu, S.-L. Wong, Engineering Soluble Monomeric Streptavidin with Reversible Biotin Binding Capability. *J. Biol. Chem.* **280**, 23225–23231 (2005).
96. S. Muyldermans, Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry* **82**, 775–797 (2013).
97. “Nanobodies/VHHs” Chromotek,  
<https://www.chromotek.com/technology/recombinant-single-domain-vhh-antibodies/>  
(Accessed April 29, 2021).

*Chapter 2*THE COMPUTATIONAL DESIGN PROCESS OF GENERATING CO-BINDING  
NANOBODY-BIOTIN CONJUGATES

Acknowledgement: Aiden J. Aceves wrote and executed the relevant scripts to construct the rotamer library and generate the raw rotamer virtual screening data.

## ***Introduction***

This chapter documents the detailed design process of 10 nanobody-biotin conjugates, of which the experimental validation and discussion are elaborated in Chapter 3. Starting from two PDB-derived nanobody scaffolds, the first generation of conjugates were designed by a docking and filtering procedure that separately optimizes interface geometry and conjugation plan. Sequence optimization was then performed by a process that iteratively introduces limited number of mutations in variable and conserved regions of the immunoglobulin domain. As the result, from the two starting scaffolds, up to 13 and 19 mutations were respectively introduced in the reported constructs. The design procedures described in this chapter are not restricted to the model system of monomeric streptavidin/biotin. Finally, exploration of various computational methods revealed potential key design principles, from which a general design strategy is proposed for this ADC-inspired chimeric modality.

### ***Section A: the design process of 4NBX.B-derived nanobody-biotin conjugates***

#### **Overall design strategy**

Designing the antibody components of synergistically-binding ADCs will involve creating new antibody/target interface, which is challenging, largely because of the difficulty in predicting the global minimum conformation of antibody CDR loops against a targeted surface, while accurately modeling long structured loops remains a challenge in general [1-4]. To restrict unpredicted CDR conformations that could lead to non-binding designs, we decided to adopt an approach similar to the anchored-design methods [5]. Anchored-design creates new protein-protein interface by first identifying hotspot residues that favorably interact with the target, then designing protein scaffolds to stabilize the anchoring hotspots [5-7]. For synergistically-binding ADCs, the conjugated small molecule can be viewed as a hotspot “residue” that interacts with the target protein. Therefore, to create co-targeting ADCs, the drug can be designed as an anchoring non-natural CDR residue that is strengthened by additional CDR-target interactions, integrating the drug-target interaction

into the antibody-target binding event, and forcing the CDRs to more likely adopt the designed conformation (Figure 2.1).

We therefore finalized the design strategy into the following steps: predicting the optimal CDR binding poses against the target surface, and searching for the ideal conjugation strategy that accommodates both the optimized CDR pose and the target-small molecule interaction.

### Searching optimal streptavidin-binding nanobody CDR conformations by docking and loop modeling

Two monomeric streptavidin constructs were used as target protein during the design process. The initial steps were performed on a modeled streptavidin monomer with triple mutations S45A/T90A/D128A, which were reported to break up the WT tetramer [8]. Due to an instability issue described in Chapter 3, another published monomeric streptavidin construct mSA was used to perform later stages of design.

The monomeric streptavidin model with triple mutations S45A/T90A/D180A was prepared from crystal structure of wild type core tetrameric streptavidin (PDB ID: 1MK5). A single subunit was extracted and standardized by an in-house computational protein design suite TRIAD [9]. S45A/T90A/D128A substitution was then performed by the TRIAD sequence-design module.

To optimize nanobody/streptavidin binding conformations, protein-protein docking followed by loop modeling of nanobody CDRs was performed as the initial attempt. A published nanobody structure (PDB ID: 5VNW, chain C) was used as the starting nanobody scaffold, with all CDR residues replaced to alanine by TRIAD sequence design, in a hope to avoid sequence bias [10]. The nanobody scaffold was docked onto a set of manually selected surface residues surrounding the binding pocket of monomeric core streptavidin model S45A/T90A/D128A (Figure 2.2A). Docking was performed by a

previously developed FFT-based docking program, and top 15 CDR binding poses were kept [11]. CDR loop modeling of each pose was then performed to attempt optimizing CDR conformation against the target by the TRIAD loop modeling module. Only the top solution for each pose was kept, and no solution had reasonable CDR conformations that seemed to be beneficial for a good binding interface (Figure 2.2B). The sequence of the core streptavidin with the triple-alanine mutations is attached below, and the residues selected as docking targets are highlighted:

EAGITGTWYN**QLG**STFIVTAGADGALTGT**YEA**AVG**NAESRY**VLTGRYDSAPAT  
DGSGTALGWTVA**WKN**NYRNA**HS**AATWSGQYVGGAEARINTQ**WLL**TS**GT**TEA  
NAWKSTLVGHAT**FT**TKVK.

#### Searching optimal streptavidin-binding nanobody CDR conformations by docking native CDR sequences and conformations

154 nanobodies that have continuous electron density with diverse target-binding CDR conformations were fetched from publicly available nanobody/target structures on PDB (please also refer to section C below). CDRs of each nanobody were subsequently annotated following the CDR-mapping criteria of nanobodies in a published study (Figure 2.2C) [12], but with softer edge cutoffs so that the sampled nanobody approaching geometries are not too stringent. Using the annotated CDRs, each nanobody was docked against the binding pocket surface of the streptavidin model, and top 15 poses for each docking trial were kept. Initially, we still performed alanine-replacement on all CDR residues before docking. However, several CDR sequence design trials on a docked structure returned mostly small amino acids (data not shown), indicating that the designability might be restricted in alanine-replaced docking complexes. Therefore, we decided to dock the 154 nanobody structures with native CDR sequences.

The TRIAD surface-complementarity module was used to perform interface analysis of naturally occurring nanobody/target structures to report the distribution of interface separation distance, shape complementary, and buried interface area for future design reference (Figure 2.2D) [13]. 2310 docked poses were generated, and then filtered by three selection steps to identify most realizable binding poses (Figure 2.3). Step One selected poses with interface separation distance, shape complementary score, and buried interface area within 1 standard deviation of naturally occurring nanobody/target structures, returning 231 poses. Step Two then selected nanobody poses that use >80% of the residues that participate in the original binding interface to interact with the docking target, and returned 31 poses. This step aimed to identify binding poses that best recapitulate how the chosen nanobody scaffold engages with its original target, so the poses would resemble a naturally occurring binding mode. Identification of interface residues in both original PDB structures and docked structures was performed by a publicly available PyMOL script that selects interface residues by changes of solvent accessibility [14]. Lastly, Step Three selected poses that directly blocked the biotin binding pocket, and returned the final 7 poses after scaffolds with lower-than-3Å resolution were removed (Fig 2.3, 2.4). The degree of binding pocket blockage was reflected by calculating the change of solvent accessible area of the following selected residues (using PyMOL), which are a subset of the docking target residues that closely surround the biotin molecule (highlighted below):

EAGITGTWYN**QLGST**FIVTAGADGALTGTYEAA**VGNA**ESRYVLTGRYDSAPAT  
DGSGTALGWTVA**WKNNYR**NA**HS**AATWSGQYVGGAEARINTQ**WLLT**SGTTEA  
NAWKSTLVGHAT**FTK**VK.

A larger group of residues was selected in the previous docking step in order to minimize artificial factors that may unnaturally restrict binding pose sampling.

### Rotamer library screening of biotin C2 maleimide side chain to determine the optimal conjugation plan

Searching of energetically compatible conjugation strategy against the docked poses was then performed. A rotamer library of the biotin-CH<sub>2</sub>-CH<sub>2</sub>-succinimide-S-CH<sub>3</sub> “side chain” was constructed by OpenBabel [15], and virtually screened against CDR amino acid locations of the 7 poses to find the optimal conjugation sites. In the rotamer library, the biotin portion remains intact, while the CH<sub>2</sub>-CH<sub>2</sub>-succinimide-S-CH<sub>3</sub> portion is diverse in torsion angles. Screening was done by measuring 1.) the distance between the terminal carbon of the rotamer and the C $\beta$  of respective CDR residues, 2.) steric clash between the rotamer and the streptavidin, 3.) steric clash between the rotamer and the nanobody, and 4.) the angle of the rotamer terminal carbon approaching the respective attachment spot. Conjugation geometries that clashed with streptavidin by <1 unit, clashed with the nanobody by <15 units, approached the attachment spot by 100-120 degree, and were <1 Å away from the C $\beta$  of screened conjugation sites were kept. Measurements were performed only against CDR residues that are originally alanine, which is similar to cysteine in size and usually does not perform important structural role, as we hypothesized that making an alanine to cysteine mutation would less likely cause serious structural consequences to the nanobody. Alanine 103 on 4NBX.B was the only conjugation site that passed the above filters, and the rotamer that clashed least with both the streptavidin and nanobody was selected for further processing. To prepare the final conjugation structure, excess atoms were deleted, and a bond was made between the C $\beta$  of the biotin-CH<sub>2</sub>-CH<sub>2</sub>-succinimide-S-CH<sub>3</sub> “side chain” to the C $\alpha$  of 4NBX.B site 103. The conjugated structure of 4NBX.B-biotin against the S45A/T90A/D128A streptavidin was then relaxed by Biograf [16], with force constraints placed to maintain the biotin-streptavidin hydrogen bonds and torsion angles of the biotin aliphatic arm (Figure 2.5A).

Inspecting the modeled structure reveals that 4NBX.B binds in a similar mode as in its original PDB structure (Figure 2.5B-C). In the model, the CDR residues interacting with streptavidin overlap significantly with the CDR residues that interact with the original

4NBX.B target. Specifically, two CDR residues Y112 and R27 participate in H-bond formation both in the original PDB structure and in the model. Therefore, the design goal of finding a binding scaffold that consists of both a naturally occurring nanobody structure and a naturally-resembling binding mode was achieved.

To prepare 4NBX.B-biotin103/mSA structure, the crystal structure of mSA (PDB ID: 4JNJ) was aligned to the modeled triple-mutation streptavidin structure, and 4NBX.B-biotin103 structure was relaxed by Biograf with the same force constraints (Figure 2.6) [16].

Alignment of the final prepared models of 4NBX.B-biotin103 against mSA and S45A/T90A/D128A streptavidin showed a high degree of structural similarity, including the modeled linker connecting the nanobody and biotin components (Figure 2.6). Therefore, we kept the designed 4NBX.B-biotin103 construct and proceeded with sequence optimization against mSA as the target.

### Sequence design optimization of 4NBX.B-biotin103

Because of the chimeric nature of the molecule being designed, high-throughput screening assays are not technically feasible. Therefore, the sequence design process was constructed to maximize the chance that a functional mutant could be identified with minimum experimental testing, instead of finding the best possible mutations. In accordance with this preference, each round of design only aimed to introduce a limited number of mutations, and certain restrictions on sequence choices were implemented.

CDR sequence design was first performed on the prepared 4NBX.B-biotin103/mSA<sub>WT</sub> model. CDR residues were determined following the CDR-mapping criteria in a published study [12]. Specifically, the 4NBX.B model was first aligned with an example nanobody structure that follows the criteria (PDB ID: 5VNW, chain C), and then the corresponding CDR residues on 4NBX.B were selected. Residues 27-34 were selected as CDR1, 47-60 were selected as CDR2, and 98-111 were selected as CDR3. Single-point mutation scan was performed on each CDR location with reduced sets of amino acids that were reported

to be frequently used in each corresponding CDR position [12]. The candidate amino acid choices for each position are as follows:

<b>Residue Position(s)</b>	<b>Amino Acid Choices</b>
27	WT/N/S/T/Y
28/51	WT/I
29	WT/S/F
34	WT/M
47	WT/F/L
48	WT/V
49/60/98	WT/A
50/55	WT/A/G/S/T
52	WT/A/D/G/Q/S/T
54	WT/G
56	WT/I/N/S/T
57/58	WT/T
59	WT/N/Y
60/111	WT/Y
109	WT/F/H/L/Y
30-33/53/99-102/104-108/110	WT/A/R/N/D/Q/E/G/H/I/L/K/F/P/S/T/W/Y/V

Position 103 with the attached biotin “side chain” was left un-designed and the coordinates for all atoms were left unchanged. Rosetta force field with covalent terms was used during the calculations [17]. Biograf-relaxed 4NBX.B-biotin103/mSA<sub>WT</sub> structure was used as structural input. During the design calculations, residues that were within 10 Å from the site under design calculation were allowed to repack. Each rotamer optimization for the site under design calculation was initiated by random rotamer configurations, and then repacked while the C $\alpha$  backbone was allowed to relax through Cartesian minimization to optimize the structures with different sequence choices. The sampled sequences were then ranked by the energy scores of the corresponding modeled structures after iterative rotamer repacking and backbone relaxation. The chemical attributes of the biotin103 “side chain” were generated by TRIAD and then used for calculating the energy scores. 10 runs with

different random seeds were performed for each design calculation, and averaged to reflect the final amino acids preference for each site that underwent single-site mutation calculations. Mutation choices with lower Rosetta energy unit than the WT amino acids were kept as designable mutations (Figure 2.7A). Designability of each site was reflected by the sum of Rosetta energy differences from the WT amino acid choice. As the result, the reported designable sites ranked by designability were as follows: 105, 109, 107, 104, 106, 32, 108, 31, and 56. Those sites were also separately grouped into two bins. Bin 1 contains sites that interact with the original target of 4NBX.B: sites 105, 104, 32, and 31, of which the order was ranked by designability. Bin 2 contains sites that do not interact with the original target of 4NBX.B: sites 109, 107, 106, 108, and 56. The order reported here was also ranked by designability.

Combinatorial designs with different biases towards designable sites were performed in parallel, in order to test which design configuration would return mutations with predicted improvements. Combinatorial design 1 was performed on all the 9 designable sites. Combinatorial design 2 was performed on the designable sites in bin 1 only. Combinatorial design 3 was performed on the designable sites in bin 1 and bin 2, with an exception that for bin 2 sites, only mutation choices that are different from WT amino acid by  $>1$  Rosetta energy units were used. Combinatorial design 4 was performed on the designable sites in bin 1 and bin 2, with an exception that for bin 2 sites, only the top-ranked mutation choice by energy score was used.

Combinatorial designs were performed with the same configurations as single-site mutation designs that were described before, with one difference: the output sequences from the 10 parallel design runs were re-ranked by threading the sequences sampled in each run individually onto the backbone of the input structure, followed by rotamer repacking and backbone Cartesian minimization. The re-ranking step was introduced in consideration of the significantly increased size of sequence space being explored. The TRIAD-modeled structures and Rosetta energy scores of the top 20 sequences of the re-ranked sequences were used to evaluate design results.

H-bonds were used as the criteria to confirm improved intermolecular interaction among the designs. Structures of the top 20 sequences for the 4 combinatorial designs were analyzed by PyMOL to visualize intermolecular H-bonds between mSA<sub>WT</sub> and 4NBX.B-biotin103 variants, and also intramolecular H-bonds within 4NBX.B-biotin103 variants, using a publicly available PyMOL script that relies on the “find\_pairs” command module of PyMOL [18]. The goal was to find sequences with improved overall energy score and new intermolecular H-bonds with mSA<sub>WT</sub>. In addition, an intramolecular H-bond profile comparable to 4NBX.B-biotin103 WT was also aimed to be achieved, as the imbalance between forming new interactions with targets and keeping the structural integrity was a common reason behind the failure of designing protein-protein interactions [19]. All combinatorial designs output sequences with improved energy scores, but only combinatorial design 2 output the top 20 sequences with an average number of intermolecular H-bond higher than that of the 4NBX.B-biotin103 WT against mSA<sub>WT</sub> (Figure 2.7B). The top 20 sequences in combinatorial design 2 also had the highest average number of intramolecular H-bonds in the nanobodies among the 4 designs. Variant v119 was the top-ranked sequence in combinatorial design 2 by energy, and variant v149 had the highest number of predicted intermolecular H-bonds among the top 20 sequences (Table S1).

To improve the stability of v149, we hypothesized that a suitable method would optimize the protein structure while preserving the designed interactions contributed by R31/S32/A104/R105. Therefore, we devised a sequential design workflow that creates stepwise local structural optimizations that compensate for the mutations built-up in previous steps. As the result, subsequent rounds of CDR sequence designs were performed on v149.

Each round also consisted of two steps. First, single-point mutation scan was performed on the CDR residues using the identical setup as the first round of design. The calculation results were processed in the same way as the first round of design, with an exception that only mutation choices that were better than WT amino acid by >1 Rosetta energy units

would be kept, in order to reduce the complexity that would build up along the design cycles. Next, skipping the sites that were mutated in previous rounds, four combinatorial designs were performed on the designable sites reported by the single-point mutation scan calculation. Combinatorial design 1 was performed on all designable sites with the reported designable mutation choices. Combinatorial design 2 was performed only on the designable sites in bin 1. Combinatorial design 3 was performed on top 5 designable sites. Combinatorial design 4 was performed on 5 designable sites ranked by designability, but with a bias on sites in bin 1. In other words, sites in bin 2 were not used unless the number of sites in bin 1 was smaller than 5. All combinatorial design calculations were performed and processed under the same setup as the first round of design.

The results of the sequential design calculations are elaborated below. The second round of design was performed using the v149 structure as input. No improvement in the number of intermolecular H-bond formation was observed in the outputs of all 4 combinatorial designs, while the number of intramolecular H-bond was minimally different. Therefore, combinatorial design 4, which was the design calculation that returned the biggest energy score improvement in its top 20 sequences, was chosen. Out of this design calculation, the sequence with the best energy score, v149 plus Y101L/R107F, was selected as the input structure for the third round of design. Combinatorial design 4 was performed with sites 27/59/101/107/110.

After the third round, once again no improvements in the number of intermolecular and intramolecular H-bonds were observed among the 4 combinatorial designs. So the sequence with the best energy score, v186 of combinatorial design 1 (v149 plus Y101L/R107F/R56T/Y106K/D108A/Y110S), was selected. Combinatorial design 1 reported the biggest overall energy score improvement, and was performed with sites 27/29/56/106/108/110. A further round of CDR design was performed on v186 and all combinatorial design results returned sequences with worse energy score than the parent sequence. Therefore, CDR design stopped at v186.

Based on experimental evidence and suggestions from MD simulations that are elaborated in Chapter 3, we hypothesized that only mutating CDRs was not sufficient. Therefore, we proceeded to design the  $\beta$ -barrel framework of v186. Because the framework sequences of nanobody are highly conserved [12], a suitable sets of amino acid choices and locations would be crucial for design success. Since the CDR designs were based on a published summary of nanobody CDR sequence diversity, we referred to the framework sequence used in that study for our design calculation [34]. We aligned 4NBX.B-biotin103 v186 with chain C of 5VNW, identified framework sites where the two nanobodies differ, and performed a combinatorial design with the selected sites being one or the other amino acid choice. Site positions and sequence choices were as follows:

<b>Residue Position(s)</b>	<b>Amino Acid Choices</b>
5	V/G
12	A/V
35	A/G
37	F/Y
40	P/A

The v186 structure was used as input, and the configurations and processing of design calculation were the same as the combinatorial CDR sequence designs described above. The top-ranked sequence by energy score was v186\_Fr (v186 plus A12V/F37Y) (Table S2). To see whether the identified framework mutation would be beneficial for all 4NBX.B derived conjugates, framework design with identical amino acid sites, sequence choices, and calculation configurations was performed using v149 (Table S3). Interestingly, no framework mutations were predicted to improve over the input sequence v149.

In summary, using small-molecule/protein structure as input, 5 nanobody-biotin conjugates were computationally designed by evaluating against a series of energetic and structural features, which are used to select sequences that may realize the desired synergetic binding property. The conjugates are 4NBX.B-biotin103 WT, v119, v149, v186, and v186\_Fr.

Testing whether the designed conjugates behave well in experimental setting would be crucial to evaluate the design quality.

***Section B: the design process of 2X89.A-derived nanobody-biotin conjugates***

One flaw in designing the 4NBX.B-derived conjugates was that the binding scaffold selection step and the sequence design step were performed on two different targets. To test the robustness of the design workflow, especially the reliability of the docking and pose optimization process, mSA was used as the input target structure to design compatible conjugates from scratch.

Docking, pose filtering, rotamer screening, and binding pose generation for 2X89.A-CCAA-biotin<sup>57</sup> WT against mSA<sub>WT</sub>

The 154 nanobody structures from PDB with native CDR sequences were docked against a manually selected set of surface residues around the biotin-binding pocket of mSA, in the same way as described in section A. The amino acids being docked against are highlighted below:

GAEAGITGTWYN**QSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRY**  
 NGTKLEWRVEWNN**STENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPA**  
**TEQGQDTFTKVK.**

Filtering procedures of the docked poses were also identical to the steps elaborated in section A, with an exception that the following highlighted were selected as the target for binding pocket blockage filter:

GAEAGITGTWYN**QSGSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRYN**  
 GTKLEWRVEWNN**STENCHSRTEWRGQYQGGAEARINTQWNLTYEGGSGPATE**  
**QGQDTFTKVK.**

6 poses that respectively used nanobody scaffolds 2X89.A, 3EBA.A, 4LHQ.B, 4OCL.C, 3V0A.C, and 4P2C.G passed the filter series. Visual inspection of the poses revealed that the 4OCL.C and 3V0A.C binding poses showed significant contacts that are mediated by nanobody residues outside of the CDRs, so the corresponding two poses were discarded since these binding modes were potentially unrealistic. As the result, 4 final binding poses were kept for conjugation plan optimization (Figure 2.8A).

The biotin-CH<sub>2</sub>-CH<sub>2</sub>-succinimide-S-CH<sub>3</sub> rotamer library built in section A was used again for rotamer screening. Because the design process of the 4NBX.B-derived conjugates demonstrated that our design capability allowed structural stability of the conjugates to be designed after binding interactions were established, this time we did not put much emphasis on preserving structural integrity in the early stage of design. Therefore, instead of screening only against alanine CDR residues, all CDR residues were screened by measuring 1.) the distance between the terminal carbon of the rotamer and the C $\beta$  of respective CDR residues, 2.) steric clash between the rotamer and the streptavidin, 3.) steric clash between the rotamer and both proteins, and 4.) the angle of the rotamer terminal carbon approaching the respective attachment spot. Conjugation geometries that clashed with streptavidin by <0.5 unit, clashed with both proteins by <10 units, approached the attachment spot by 100-120 degree, and were <2 Å away from the C $\beta$  of screened conjugation sites were kept. Only one rotamer that was screened against I57 of 2X89.A passed the filter. The final conjugation structure was prepared by Biograf, under the same parameters as described in section A [16].

In addition to the introduced cysteine, 2X89.A has a solvent accessible disulfide bond that may result in multi-conjugation. To remove the intra-CDR disulfide bond in the Biograf-relaxed structure, C33A/C104A mutations were introduced by TRIAD sequence design module to create the finalized 2X89.A-CCAA-biotin57/mSA model for further sequence design optimization (Figure 2.8B).

### Sequence design optimization of 2X89.A-CCAA-biotin57

Summarizing the experience from designing the 4NBX.B-biotin103 conjugates, we gained the following insights into the sequence design principles of the conjugates:

- 1.) Performing sequential rounds of design on limited sets of amino acid sites/choices, which are recommended by iterative energetic and structural analysis, allows functionally improved mutant conjugates to be discovered without experimentally screening a large set of sequences.
- 2.) New intermolecular interactions between the nanobody scaffold and the target can be engineered first before further mutations are introduced to optimize the structural integrity of the conjugates.
- 3.) Simply mutating CDR residues is not sufficient for structural optimization of the conjugates, and mutated CDR residues likely need accommodation by introducing mutations in the  $\beta$ -barrel framework region.

Based on the above principles, we established a rudimentary sequence design pipeline, with a primary goal to optimize the overall stability. The detailed procedures are described below (Figure 2.9):

- 1.) The pipeline performs sequential rounds of sequence design that is either restricted on CDR residues or framework residues. The first round of design is performed on CDR residues. Residues that are mutated in previous rounds are kept from mutation in further rounds.
- 2.) H-bonds are still the only intermolecular interactions that are explicitly evaluated after design calculations and biased towards for sequence selection, but other types of interactions can be evaluated if they are deemed to be crucial for specific scenarios.

3.) CDR sequence design follows the procedures of designing the CDR loops of v149, as described in section A, with one exception: besides the four combinatorial designs, it is optional that two additional combinatorial designs can be performed in parallel, respectively on all identified sites and amino acid choices in bin 2, and on the top 5 designable sites with a bias on sites in bin 2. All combinatorial designs are evaluated together as described in section A.

4.) Framework design follows the procedures of designing the framework of v186, as described in section A.

5.) To evaluate the design results of CDR design, the following steps are used:

Sequences that showed worse energy score than the immediate parent sequence are discarded. Sequences with the number of intermolecular H-bonds lower than the immediate parent sequence are discarded. If no sequences survived the above two filters, perform framework design on the immediate parent sequence. For sequences pass the filters, the sequence that has the highest number of intermolecular H-bonds and the best energy score among sequences that share the same number of intermolecular H-bonds is kept as input for the next round of CDR design.

6.) To evaluate the design results of framework design, the following steps are used:

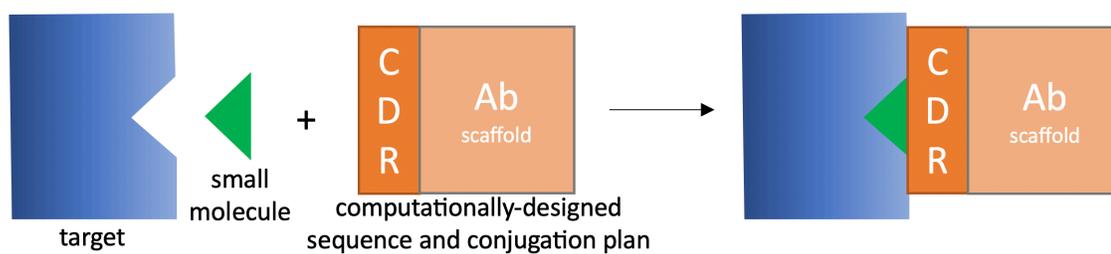
Sequences that showed worse energy score than the immediate parent sequence are discarded. Sequences with the number of intermolecular H-bonds lower than the immediate parent sequence are discarded. Sequences with the number of nanobody intramolecular H-bonds lower than the immediate parent by  $>1$  are discarded. If no sequences survived the above three filters, the design fails. For sequences pass the filters, the sequence that has the highest energy score is kept as input for the next round of CDR design.

Following the pipeline, 7 rounds of sequence designs were performed in total. The originally designed conjugate 2X89.A-CCAA-Biotin57 and the outputs after rounds 3, 5, 6, and 7, which are variants v37, v42, v20, and v5, were selected for experimental testing.

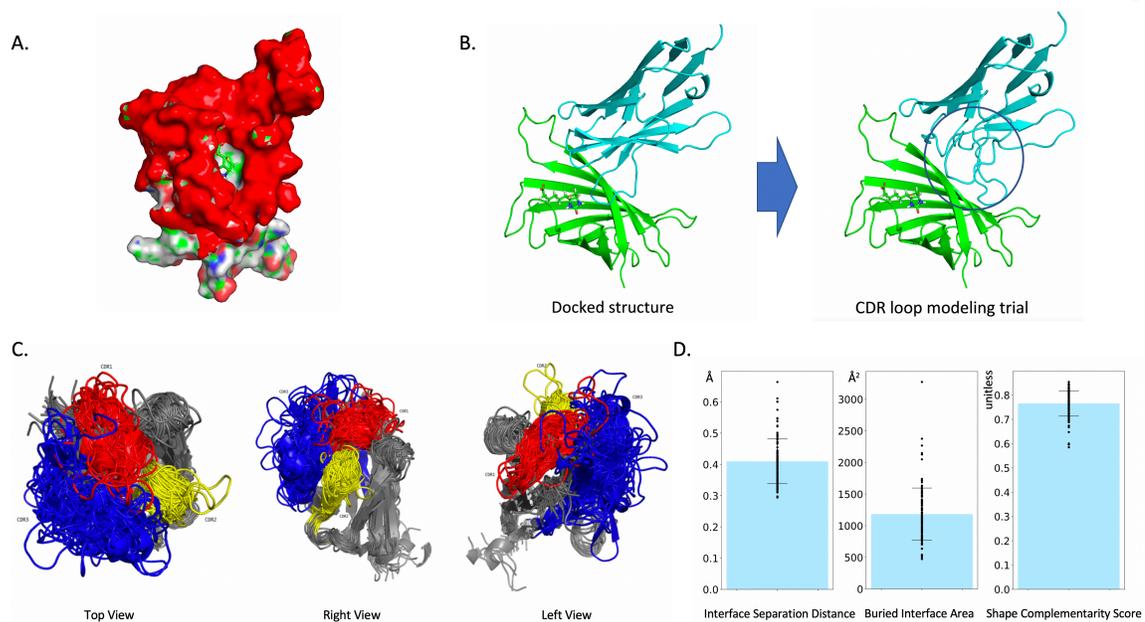
***Section C: nanobody sequences that are used in this study for docking and binding pose selection***

Note: Sequences are represented by [PDB ID].[chain name]

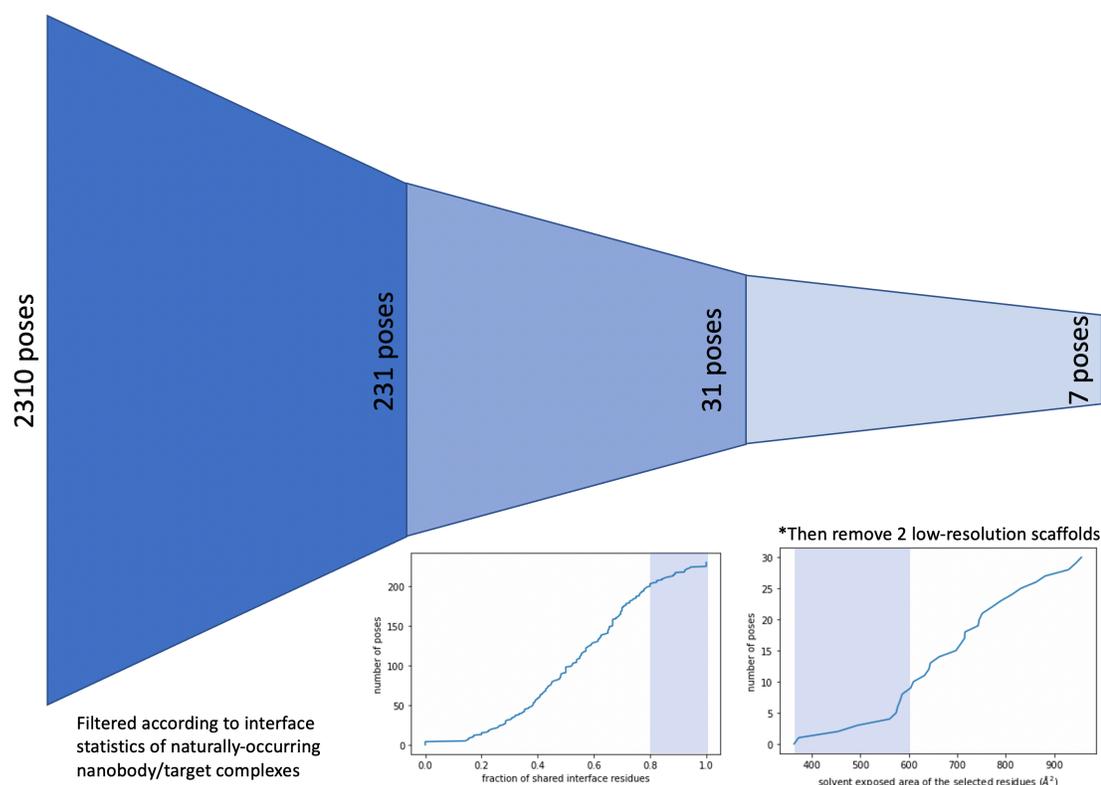
1BZQ.K, 1JTT.A, 1KXQ.E, 1OP9.A, 1ZVH.A, 1ZVY.A, 2X89.A, 3EBA.A, 3JBC.7, 3JBE.7, 3JBF.7, 4GRW.E, 4I0C.C, 4W6W.B, 5BOP.A, 5C2U.B, 5FOJ.A, 5M13.B, 5TJW.K, 2XT1.B, 3K74.B, 3SN6.N, 4C57.C, 4EIZ.C, 4FHB.D, 4GRW.F, 4HEM.E, 4KML.B, 4LGP.B, 4N9O.B, 4NBY.B, 4NBZ.B, 4TVS.b, 5C3L.D, 5F1K.C, 5F1O.B, 5H8O.A, 5IMK.B, 5J57.B, 5JA8.B, 5JMO.C, 5KU2.7, 5LHN.B, 5NBD.C, 5O03.C, 5USF.C, 5UZ7.N, 5VXL.B, 5VXM.B, 4Z9K.B, 5IVN.A, 3QXT.A, 4DK3.A, 4GFT.B, 4KRL.B, 4LHQ.B, 4N1H.B, 4S10.A, 5OJM.K, 5UKB.a, 5VXK.B, 1I3U.A, 1RJC.A, 4GRW.H, 4IOS.D, 4OCL.C, 5BOZ.G, 5JDS.B, 5M2M.D, 1KXV.C, 1QD0, 1RI8.A, 1ZV5.A, 3JBD.7, 3K3Q.A, 3QXV.A, 3RJQ.B, 3STB.A, 4AQ1.B, 4CDG.C, 4I13.B, 4QO1.A, 4W6X.B, 4WEM.B, 4XT1.C, 5E0Q.A, 5G5R.B, 5GXB.B, 5LHR.B, 5LWF.C, 5O8F.K, 5OCL.B, 4C58.B, 2XXM.B, 3JBG.7, 4X7F.C, 5F21.B, 5F7K.C, 5HVG.B, 4LHJ.B, 4M3K.B, 5HVF.B, 5MJE.B, 1G6V.K, 2X6M.A, 3K1K.C, 3V0A.C, 4U3X.A, 5E5M.B, 5HGG.S, 5HM1.A, 5J56.B, 5JA9.A, 5MWN.N, 5O2U.B, 5OVW.G, 5TOK.D, 1KXT.B, 3CFI.C, 3K81.A, 4EIG.B, 4HEP.G, 4LGR.B, 4MQS.B, 4NBX.B, 4NC2.B, 4P2C.G, 4W6Y.B, 4YGA.B, 5DFZ.E, 5IP4.A, 5J1S.C, 5JQH.C, 5KTZ.7, 5KU0.7, 5KWL.7, 5L21.B, 5MY6.B, 5O02.C, 5O04.E, 5O0W.E, 5OCL.A, 5OMN.C, 5TOJ.D, 5UK4.a, 5VXJ.B, 4X7C.C, 3EZJ.B, 5F7L.B, 5M30.D, 5M94.B, 3P0G.B, 4LDE.B, 2BSE.D



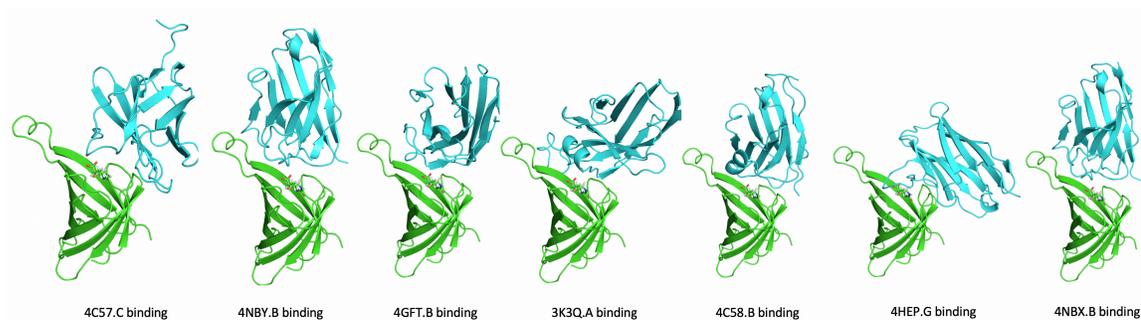
**Figure 2.1 The schematic representation of the envisioned workflow:** Given the availability of a small molecule and its target, the sequence of a complementary immunoglobulin domain and a conjugation plan with the small molecule are computationally determined to create conjugates that synergistically bind to the target.



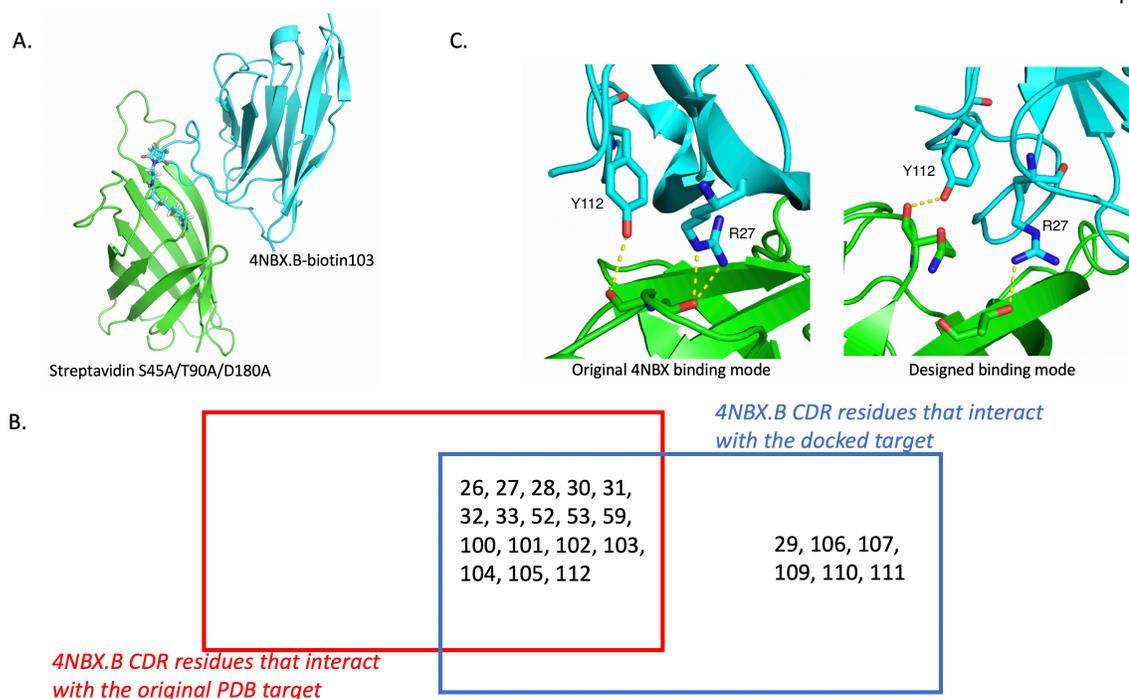
**Figure 2.2 CDR conformational sampling by loop modeling and by adapting from naturally occurring structures:** (A) One streptavidin protomer is shown with residues selected for docking colored red. The biotin molecule is shown as sticks. (B) Loop modeling on docked nanobody binding poses failed to return reasonable loop configurations. Streptavidin S45A/T90A/D180A is colored green, and nanobody scaffold is colored cyan. (C) Whole structural alignment of 154 curated PDB nanobody scaffolds with diverse CDR conformations and sequences. Red: CDR1. Yellow: CDR2. Blue: CDR3. (D) Interface statistics of naturally occurring nanobody-target complexes. Error bars represent standard deviations.



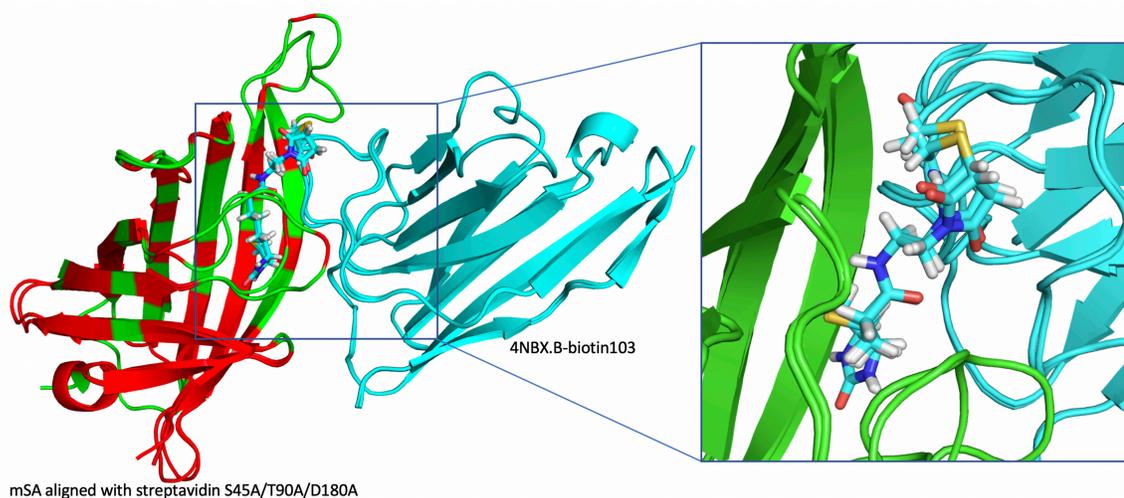
**Figure 2.3 Selecting docked poses against streptavidin S45A/T90A/D180A by three filtering steps to identify potentially realizable binding poses:** The first step selected poses with interface separation distance, shape complementarity, and buried surface area within 1 standard deviation of analyzed naturally occurring nanobody/target complexes. The second step selected poses that used >80% of interface-participating residues in the original PDB structure against the docked target. A CDF graph is shown here for reference. The third step then selected poses with lower than 600 square-angstrom of solvent exposed area for a limited set of residues surrounding the biotin binding site, followed by removing two poses with lower than 3-angstrom resolution. A CDF graph is shown here for reference.



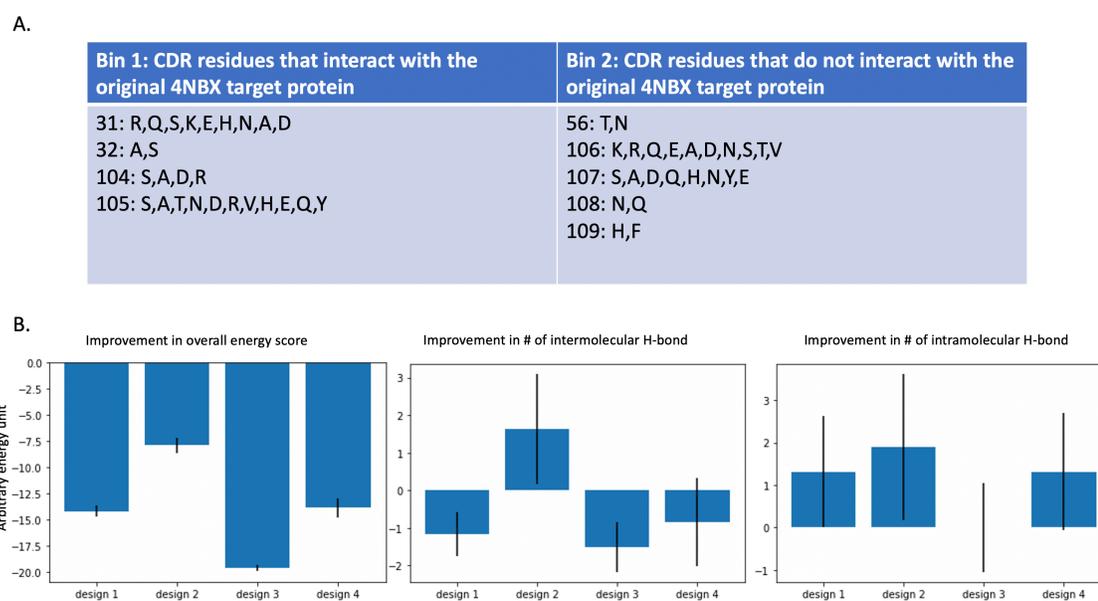
**Figure 2.4** The final seven poses that passed the filtering steps: Cyan: nanobody scaffolds. Green: streptavidin protomer modeled with S45A/T90A/D180A triple mutations. The biotin molecules are shown as sticks.



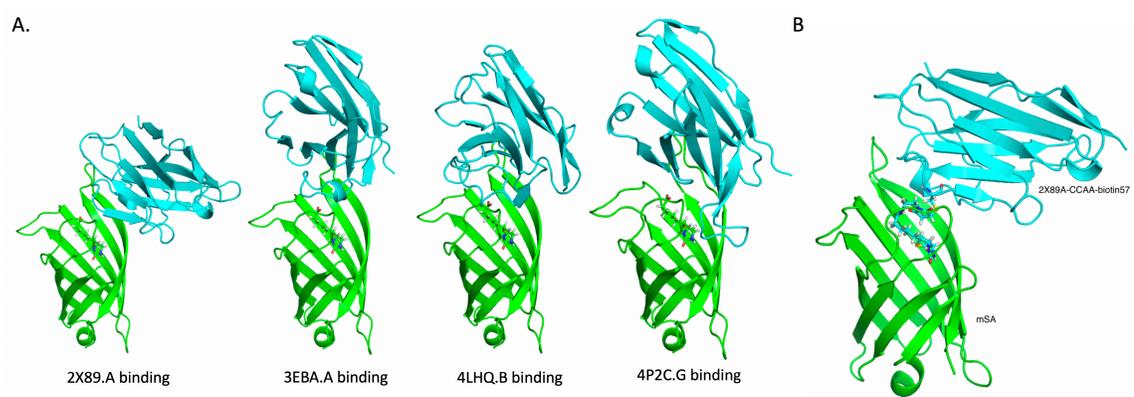
**Figure 2.5 The final prepared 4NBX.B-biotin103 structure in complex with the target protein recapitulates the original 4NBX.B binding mode:** (A) Prepared structure of 4NBX.B-biotin103 in complex to streptavidin S45A/T90A/D180A. Streptavidin S45A/T90A/D180A is colored green, and nanobody scaffold is colored cyan. Biotin103 sidechain is shown as sticks. (B) The overlap of CDR residues that participate in interface formation in the original 4NBX PDB structure and the final modeled structure against the streptavidin target. (C) The H-bond forming potential of Y112 and R27 in 4NBX.B nanobody was predicted to be recapitulated in the designed binding pose with the streptavidin model. Streptavidin S45A/T90A/D180A is colored green, and nanobody scaffold is colored cyan. Y112 and R27 together with their predicted H-bond partners are shown as sticks.



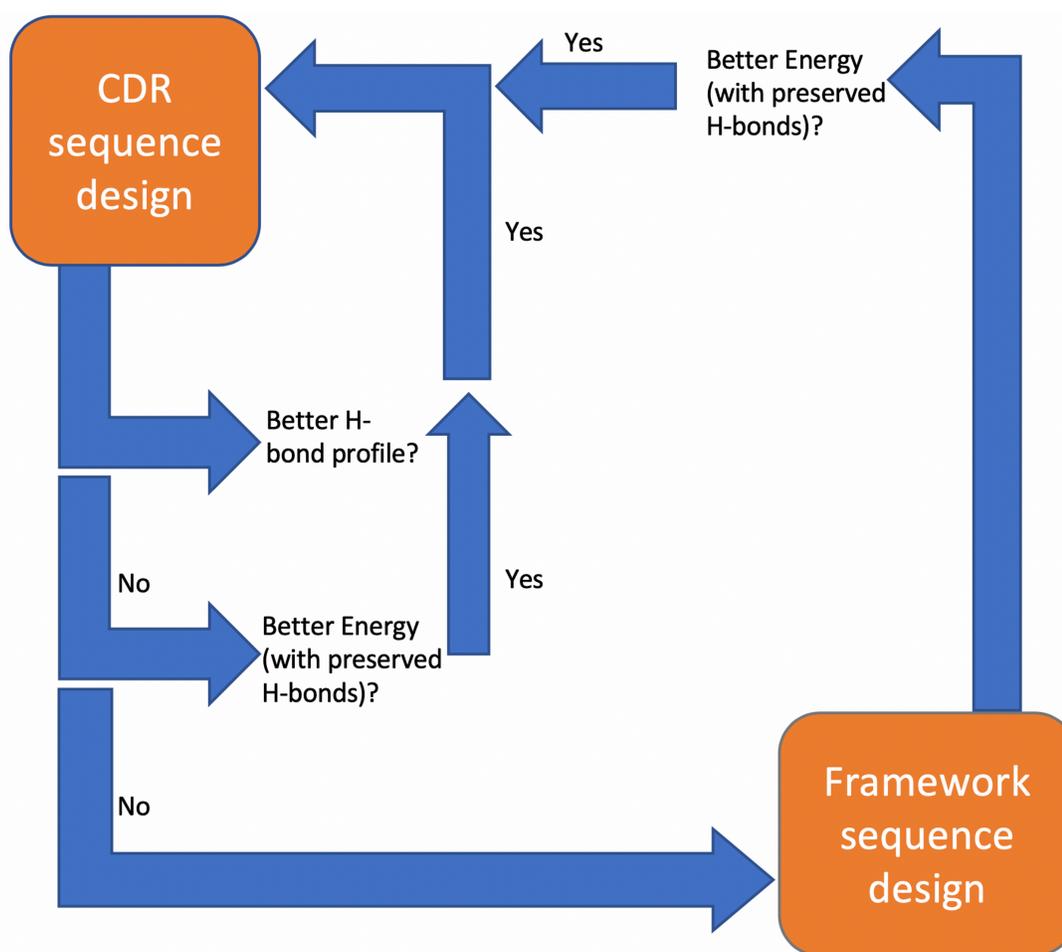
**Figure 2.6** Alignment results for prepared structures of 4NBX.B-biotin103 in complex with streptavidin S45A/T90A/D180A and mSA: Streptavidin models are colored green, and nanobody scaffolds are colored cyan. Residues that were identified by sequence alignment as pair-wise identical sequences are colored red. Biotin103 side chains from both models are shown as sticks.



**Figure 2.7 Summary of first round of sequence design calculation of 4NBX.B-biotin103:** (A) Site-saturation analysis results. Sites with mutations predicted to be more favorable in overall Rosetta energy score were reported and grouped in two bins, as a reference for combinatorial sequence design calculation. (B) Analysis of the top 20 output sequences from four combinatorial designs using sequence choices reported in A. Differences in overall energy score, number of intermolecular H-bond, and number of intramolecular H-bond compared to the WT sequence were reported as average and standard deviation.



**Figure 2.8 Pose identification and conjugate structure optimization of 2X89.A-CCAA-biotin57 against mSA:** (A) Finalized nanobody scaffolds and binding poses against mSA streptavidin. Cyan: nanobody scaffolds. Green: mSA streptavidin. Biotin molecules are shown as sticks. (B) Designed binding model and conjugation scheme of 2X89.A- CCAA-biotin57 against mSA streptavidin surface. Biotin57 sidechain is shown as sticks.



**Figure 2.9** A rudimentary sequence design pipeline that performs CDR and framework design in a step wise manner.

**Table S1: Top 20 sequence outputs from CDR design of site 31, 32, 104, and 105 on 4NBX.B-biotin103 WT. (Ranked by energy score)**

Ranking	Energy Score	Note	Mutations: chain name and accepted residue
1	-457.59	v119	B 31H+B 32A+B 104S+B 105H
2	-456.83		B 31N+B 32S+B 104S+B 105S
3	-456.57		B 31N+B 32S+B 104S+B 105A
4	-456.56		B 31R+B 32A+B 104D+B 105R
5	-456.50		B 31R+B 32A+B 104S+B 105Y
6	-456.44		B 31N+B 32A+B 104S+B 105S
7	-456.26		B 31Q+B 32A+B 104S+B 105Y
8	-456.13		B 31H+B 32S+B 104S+B 105V
9	-456.11		B 31N+B 32A+B 104S+B 105R
10	-455.81		B 31N+B 32A+B 104S+B 105A
11	-455.77		B 31N+B 32S+B 104S+B 105D
12	-455.67		B 31R+B 32A+B 104D+B 105H
13	-455.60		B 31Q+B 32A+B 104A+B 105R
14	-455.27		B 31R+B 104S+B 105H
15	-455.13		B 31H+B 32S+B 104D+B 105S
16	-455.10		B 31N+B 32S+B 104A+B 105S
17	-455.06	v149	B 31R+B 32S+B 104A+B 105R
18	-454.98		B 31R+B 32S+B 104S+B 105Y
19	-454.94		B 31K+B 32S+B 104A+B 105A
20	-454.91		B 31Q+B 32S+B 104A+B 105A

Note: chain B refers to the nanobody

**Table S2: Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v186. (Ranked by energy score)**

Ranking	Energy Score	Note	Mutations: chain name and accepted residue
1	-500.87	v186 Fr	B 12V+B 37Y
2	-500.36		B 37Y+B 40A
3	-500.24		B 5Q+B 12V+B 37Y
4	-500.11		B 12V+B 35G+B 37Y
5	-499.98		B 12V+B 37Y+B 40A
6	-499.91		B 12V
7	-499.23		B 12V+B 35G
8	-499.15		B 5Q+B 12V+B 35G+B 37Y
9	-499.02		B 12V+B 35G+B 37Y+B 40A
10	-498.98		B 5Q+B 12V+B 35G+B 37Y+B 40A
11	-498.79		B 37Y
12	-498.64		B 12V+B 35G+B 40A
13	-498.55		B 35G+B 37Y
14	-498.54		B 12V+B 40A
15	-498.48		B 5Q+B 12V+B 40A
16	-498.31		B 5Q+B 37Y
17	-498.04		WT
18	-497.76		B 5Q+B 12V
19	-497.68		B 5Q
20	-497.60		B 5Q+B 35G+B 37Y

Note: chain B refers to the nanobody

**Table S3: Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v149. (Ranked by energy score)**

Ranking	Energy Score	Note	Mutations: chain name and accepted residue
1	-482.29		WT
2	-481.68		B_37Y
3	-481.66		B_5Q+B_12V
4	-481.65	A12V/F37Y	B_12V+B_37Y
5	-481.52		B_12V+B_35G
6	-481.51		B_5Q
7	-481.48		B_35G
8	-481.39		B_35G+B_37Y
9	-481.33		B_5Q+B_37Y
10	-481.14		B_40A
11	-481.13		B_12V
12	-480.84		B_35G+B_40A
13	-480.83		B_12V+B_37Y+B_40A
14	-480.51		B_5Q+B_35G+B_37Y
15	-480.44		B_5Q+B_12V+B_40A
16	-480.43		B_12V+B_35G+B_37Y
17	-480.30		B_5Q+B_12V+B_37Y+B_40A
18	-480.19		B_5Q+B_12V+B_35G+B_40A
19	-480.17		B_5Q+B_35G+B_40A
20	-480.17		B_12V+B_35G+B_37Y+B_40A

Note: chain B refers to the nanobody

## Bibliography

1. S. Fischman, Y. Ofran, Computational design of antibodies. *Current Opinion in Structural Biology* **51**, 156–162 (2018).
2. J. C. Almagro, *et al.*, Second antibody modeling assessment (AMA-II). *Proteins: Structure, Function, and Bioinformatics* **82**, 1553–1562 (2014).
3. L. T. Dang, *et al.*, Receptor subtype discrimination using extensive shape complementary designed interfaces. *Nat Struct Mol Biol* **26**, 407–414 (2019).
4. K. Kundert, T. Kortemme, Computational design of structured loops for new protein functions. *Biological Chemistry* **400**, 275–288 (2019).
5. S. M. Lewis, B. A. Kuhlman, Anchored Design of Protein-Protein Interfaces. *PLoS ONE* **6**, e20872 (2011).
6. S. J. Fleishman, *et al.*, Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* **332**, 816–821 (2011).
7. X. Liu, *et al.*, Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping. *Scientific Reports* **7**, 41306 (2017).
8. M. H. Qureshi, J. C. Yeung, S.-C. Wu, S.-L. Wong, Development and Characterization of a Series of Soluble Tetrameric and Monomeric Streptavidin Muteins with Differential Biotin Binding Affinities. *J. Biol. Chem.* **276**, 46422–46428 (2001).
9. *Triad*, Protabit LLC., <https://triad.protabit.com>
10. Y. Mou, P.-S. Huang, F.-C. Hsu, S.-J. Huang, S. L. Mayo, Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10714–10719 (2015).
11. P.-S. Huang, J. J. Love, S. L. Mayo, Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J Comput Chem* **26**, 1222–1232 (2005).
12. C. McMahon, *et al.*, Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nature Structural & Molecular Biology* **25**, 289–296 (2018).
13. M. C. Lawrence, P. M. Colman, Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
14. “InterfaceResidues” *PyMOL Wiki*, <https://pymolwiki.org/index.php/InterfaceResidues> (Accessed May 24, 2018)
15. N. M. O’Boyle, *et al.*, Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **3**, 33 (2011).
16. S. L. Mayo, B. D. Olafson, W. A. Goddard, DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909 (1990).
17. R. F. Alford, *et al.*, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
- 18., “PyMOL Script repository” *Queen’s University Protein Function Discovery and Department of Biomedical and Molecular Sciences Molecular Modeling and Crystallographic Computing Facility*, <http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/> (Accessed November 1, 2018)
19. P.-S. Huang, J. J. Love, S. L. Mayo, A de novo designed protein–protein interface. *Protein Science* **16**, 2770–2774 (2007).

*Chapter 3***EXPERIMENTAL VALIDATION, EVOLUTION, AND DISCUSSION OF  
DESIGNED NANOBODY-BIOTIN CONJUGATES**

Acknowledgement: Aiden J. Aceves performed MD simulation, generated snapshots, and wrote the MD-related paragraph of materials and methods.

## ***Introduction***

Most pharmaceutical mechanisms involve drug-target interactions that are mediated by synthetic small molecules or monoclonal antibodies—the two major drug modalities [1,2]. Despite impressive successes, many biological pathways are still difficult or even impossible to pharmaceutically intervene, often because through existing approaches either the desired interactions are fundamentally difficult to engineer, or the pharmacological trade-offs for establishing the interactions outweigh potential benefits [3-6]. Therefore, new modalities that incorporate new chemistry and new biology are constantly being created to realize a versatile toolkit that more easily tackles certain challenging targets, and also expands the targetable molecular space itself [7]. To create new modalities, one common way is combining existing modalities to consolidate individual advantages and offset individual flaws [7,8]. Antibody-drug conjugates (ADCs), for example, take advantage of the excellent specificity and biological compatibility of monoclonal antibodies to improve therapeutic indices of existing small-molecule drugs [8,9]. Traditionally, the antibody and drug components of ADCs are separately developed and bind to different targets while in action [8,9]. Most current ADCs improve the selectivity of conjugated drugs as they deliver the small molecules into cell targets through specific antibody-induced receptor endocytosis [9,10]. Some ADCs and peptide-drug conjugates were also reported to improve the metabolic stability, circulation half-life, and solubility of linked small molecules through antibody-associated pharmacokinetics, chemical environment around the conjugation sites, and linker design, indicating that protein conjugation could modulate a wide range of small-molecule properties [10-14]. Recently, Cheng *et al.* from Amgen developed ADCs whose antibody and drug components bind to an identical protein target to achieve synergistic binding/inhibition effects [15]. In their study, the co-crystal structure of a small molecule drug sitagliptin, a separately developed antibody 11A19, and the protein target DPP-IV was solved first [16]. Based on the structure, optimal conjugation sites and linker sequences were then searched to create ADCs that exhibited 13 to 32-fold  $IC_{50}$  improvement than sitagliptin alone against the target [15]. Cheng *et al.*'s work suggested that small molecule binding events could be directly

optimized by conjugated antibodies, turning ADC technology into a potential tool to expand the chemical space and therefore target space of molecular recognition agents that involve synthetic small molecules.

Overall, the above discoveries demonstrated the potential for using rationally designed antibody conjugation to optimize the mechanism of action, along with many other pharmacologically relevant properties, of small molecule-based binders. However, to engineer the binding synergy required for this kind of applications, established methods that separately develop and characterize the antibody and small molecule components would be resource intensive, thus limiting the application scenarios. To realize the above-mentioned potential, a workflow that can rapidly determine a compatible antibody sequence and conjugation strategy for a to-be-improved small molecule binding event would be ideal. However, whether such workflow is technically achievable is still a question.

In this study, we explored the feasibility of computationally designing the antibody component of synergistically-binding ADCs. We introduced the concept of CDR-extended antibodies (CDRxAbs), which refer to computationally-designed antibodies whose complementarity-determining regions (CDRs) contain a small molecule ligand that binds to a certain target, with surrounding CDR sequences tailored to strengthen the target-binding interactions. At this initial stage, we focused our design on nanobodies, which are llama-derived single-domain antibody fragments that can function by themselves, with attached Fc domains, or reformatted into IgGs [17-19]. Using a modified streptavidin-biotin interaction pair as model system, we demonstrated that with only the structural knowledge of small-molecule/target interactions, nanobody small-molecule conjugates can be computationally designed to bind tighter against the target than the small molecule itself. Through subsequent sequence design, the affinity, binding kinetics, and overall stability of the conjugates can be improved in a stepwise manner.  $\geq 20$ -fold affinity improvements together with targeted kinetic-tuning can be achieved when the starting small-molecule/target affinity is as weak as 1  $\mu\text{M}$ , or as strong as 7 nM. Finally, a yeast display-

based workflow was constructed and tested on a designed conjugate, which was successfully displayed with good expression and a target-binding signal minimally affected by background labeling. A trial directed evolution experiment was then performed to isolate a variant with another 6-fold improvement in dissociation rate, confirming the effectiveness of this platform.

## ***Results and Discussion***

### Computationally designed nanobody small-molecule conjugation creates tighter binders against the small-molecule target protein

We first asked whether computationally determined nanobody sequences and their designed conjugation to a small molecule can exhibit an enhanced binding affinity to the small-molecule target. For demonstration purpose, we chose monomeric streptavidin as our model target and biotin as our model small molecule. Streptavidin-biotin interactions have been extensively studied with high-resolution crystal structures available for reliable design. Tetrameric streptavidin binds to biotin with almost the highest-possible affinity, but multiple monomeric streptavidin constructs were reported with  $>10^5$ -fold reduced biotin-binding affinity [27,28]. So as a model system, monomeric streptavidin-biotin interaction pairs not only provide room for affinity improvement, but also have a known affinity upper limit, and are thus ideal for method development.

Design calculations were performed to first identify compatible nanobody scaffolds with CDR conformations that likely form favorable interactions with the target surface. To search the optimal CDR binding conformations, we first docked a starting nanobody scaffold onto a monomeric core-streptavidin structure with computationally modeled side chain replacements S45A/T90A/D180A, which were reported to monomerize streptavidin and reduce the biotin-binding affinity to  $1.7 \mu\text{M}$  [27], and then performed loop-modeling on docked poses to attempt optimizing CDR conformations against the target surface. Most of the top loop modeling solutions were not representative of naturally occurring interactions. To sample realistic CDR structures, we instead only searched around

previously observed nanobody CDR binding conformations [29]. We curated nanobody structures with diverse target-binding CDR conformations from PDB, and individually docked them onto the target surface. 2310 docked poses were generated and filtered to potentially identify most realizable binding conformations, returning 7 final binding poses. Optimal conjugation strategy was then searched on the finalized poses. We chose to conjugate biotin onto nanobody CDRs by the cysteine-maleimide chemistry, which is a commonly used conjugation method in ADCs (Figure S1) [30]. Biotin C2 maleimide was chosen to be the conjugation reagent. Optimal nanobody scaffolds and conjugation sites were determined by computationally screening a rotamer library of the cysteine-conjugated side chain on the finalized nanobody-streptavidin poses. The top-ranked conjugation plan was amino acid site 103 of the nanobody scaffold 4NBX.B (chain B of PDB structure 4NBX), which originally binds to a target unrelated to any streptavidin construct [31]. From the relaxed structure of the conjugate named as “4NBX.B-biotin103” in complex with monomeric streptavidin, Y112 and R27 of 4NBX.B are predicted to form hydrogen bonds with the target surface, whereas in the original PDB structure, these two residues also participated in H-bond formation, indicating that the designed pose is closely related to the natural binding mode of 4NBX.B, and potentially stabilized by specific CDR-target interactions upon biotin anchoring.

We then synthesized 4NBX.B with site 103 mutated to cysteine, and performed conjugation with biotin C2 maleimide. We attempted to purify and refold the S45A/T90A/D128A mutant of core-streptavidin to perform binding measurement, but the resulted construct was unstable, as most proteins precipitated during refolding, and the refolded materials also quickly precipitated. Therefore, we aligned another previously-reported monomeric streptavidin construct, mSA, onto the triple-mutation streptavidin model that mSA is homologous to (sequence pairwise identity: 57%, structure RMSD: 0.5Å), and relaxed 4NBX.B-biotin103 against mSA [28]. The 4NBX.B-biotin103/mSA model preserved the rotamer configuration of conjugated biotin against the triple-mutation streptavidin, and H bonds contributed by Y112 and R27 were also recapitulated, and potentially participated in a broader predicted H-bond network that incorporated

biotin/mSA interactions, suggesting that 4NBX.B-C103biotin may bind to mSA with the designed beneficial synergy (Figure 3.1B-C). Indeed, surface plasmon resonance (SPR) binding experiments (Figure 3.1A) confirmed that under 25°C 4NBX.B-biotin103 binds to immobilized mSA with a  $K_D$  of  $1.8\pm 0.1$  nM, and mSA binds to immobilized biotin with a  $K_D$  of  $7.0\pm 0.1$  nM, indicating a moderate 4-fold affinity improvement that is contributed by a higher  $k_a$  (Figure 3.1D-E, 3.7A). Wildtype 4NBX.B did not show binding signal to mSA at concentrations up to 100 nM, indicating that the 4NBX.B-biotin103 conjugate binds to the targeted biotin binding pocket (Figure 1E left panel). The SPR-measured biotin/mSA affinity is similar to previously published fluorescence polarization spectroscopy data, which is  $2.8\pm 0.5$  nM under 4°C and  $5.5\pm 0.2$  nM under 37°C [28]. However, because the data fitting quality of the mSA/biotin binding curves is lower than the 4NBX.B-biotin103 binding curves, to confirm the estimated mSA/biotin affinity, we performed an alternative estimation by binding immobilized mSA to Smt3 SUMO protein that was biotinylated at the N terminus by biotin C2 maleimide. Smt3 SUMO protein has an unstructured N-terminus that we hypothesized would minimize the interaction between the protein components [32]. A similar  $K_D$  is estimated with high data-fitting quality, indicating that the measured biotin/mSA affinity is an accurate SPR estimation (Figure 3.1D right panel).

To know whether computationally designed nanobody conjugation shows improved affinity with weakly-binding small molecules, we created a single mutation S27A on mSA, whose counterpart S45A in wild type streptavidin reduces biotin-binding strength and was predicted by molecular dynamics (MD) simulation to minimally affect the overall structure [33]. On size-exclusion chromatography (SEC), mSA<sub>S27A</sub> is eluted at the same time as mSA<sub>WT</sub> (Figure S3A). SPR estimated that mSA<sub>S27A</sub> binds to biotin with a  $K_D$  of  $1.14\pm 0.02$   $\mu$ M, while 4NBX.B-biotin103 binds to mSA<sub>S27A</sub> with a  $K_D$  of  $245\pm 41$  nM, indicating a similarly moderate 5-fold improvement (Figure 3.1F, 3.7B). Together, the above results showed that based on the sole structural information of a small molecule-target interaction, nanobody conjugation to the small molecule can be designed entirely by computational methods to exhibit an affinity-enhancing synergistic binding effect.

### Sequence design further improves the binding affinity and kinetics for computationally designed conjugates

Next, we performed sequence design on the CDR loops of 4NBX.B-biotin103 to improve its binding affinity against mSA and further validate the accuracy of the modeled binding pose. We *in silico* analyzed each CDR amino acid site for its favorability of accepting mutations, and performed combinatorial designs on the mutable sites. Four combinations with different site-selection biases were tested in parallel, and the residue choices for each site were decided according to a published study on the sequence diversity of nanobody CDR loops [34]. Analysis of design outputs revealed that the design with sites 31, 32, 104, and 105 most frequently returned sequences that were likely to form additional H-bonds with mSA and were also energetically stable. The top-ranked variant by energy, v119 with CDR1 mutations M31H/D32A and CDR3 mutations N104S/W105H, was predicted to form new H-bonds with residue Q108 of mSA by H31 and with E105 of mSA by H105 (Figure 3.2A). The D32A mutation also eliminates a buried and unpaired charged residue that does not participate in extensive H-bond network formation. SPR measured the  $K_D$  of 4NBX.B-biotin103 v119 against mSA<sub>WT</sub> to be  $0.9 \pm 0.2$  nM, indicating a  $\sim 2$ -fold improvement from 4NBX.B-biotin103 WT (Figure 3.2A, 3.7A). However, the  $K_D$  improvement was again mainly contributed by  $k_a$  increase, while the observed  $k_d$  values were only minimally different (Figure 3.2A, 3.7A). To obtain a variant that would more significantly reduce the  $k_d$ , we picked variant v149 that has the highest number of predicted H-bond formation from the top 20 output sequences. 4NBX.B-biotin103 v149 has mutations M31R/D32S/N104A/W105R that were predicted to form more extensive H-bonds with Y96, E105, and Q108 of mSA, with a potential salt bridge between the nanobody R105 and mSA E105 (Figure 3.2B). Interestingly, R-E interactions seemed to be frequently used by nanobodies, further validating this designed interaction [35]. Indeed, compared to v119, SPR measured a  $\sim 2$ -fold slower  $k_d$  and a  $\sim 4$ -fold faster  $k_a$  for v149, which together contribute to the  $K_D$  of  $0.12 \pm 0.01$  nM, indicating a  $>20$ -fold  $K_D$  improvement from biotin/mSA<sub>WT</sub> affinity (Figure 3.2B, 3.7A). However, according to SEC

traces, v149 seemed to be very prone to aggregation, indicating protein instability (Figure 3.3A).

### Sequence design reduces aggregation while preserving the binding strength for the designed conjugates

One SEC, both 4NBX.B-biotin 103 WT and v119 showed single peaks eluted roughly at the same time as wild type 4NBX.B nanobody, indicating stabilized monomer foldedness (Figure 3.3A). The reduced monomer stability of v149 agrees with its predicted lower energy score than v119. Since only four residues were designed, to improve the stability of v149, we hypothesized that further CDR designs would better accommodate the biotin103 side chain and the four H-bond contributing mutations, thus stabilizing the loop and overall structure. We therefore performed two additional rounds of CDR residue mutability analysis followed by in-parallel combinatorial designs on v149 until no further CDR mutations were predicted to be energetically favorable. Mutations accumulated in previous rounds of design were kept intact in subsequent rounds. In top 20 sequences ranked by energy score of both rounds of design, no additional H-bond was predicted to form with mSA, so the sequences with the best energy improvement were selected. The resulted variant, v186, has 6 additional CDR mutations Y101L/R107F/R56T/Y106K/D108A/Y110S on top of v149, and was predicted to preserve the H-bonds contributed by v149 mutations. Indeed, v186 seemed to bind to mSA<sub>WT</sub> with very similar  $K_D$  as v149 (Figure S5). However, SEC traces of v186 showed even worse aggregates formation than v149 (Figure 3.3A).

MD simulations have been successfully applied to reveal the source of unexpected functional properties in designed proteins [36]. To understand the flaws of the structure and inform next design strategy, we perform MD simulation of 4NBX.B-biotin103 v186 in complex with mSA<sub>WT</sub>. From the simulation, we noticed that the CDR3 loop that originally folded over the  $\beta$ -barrel framework region became gradually widened from the initial conformation, and eventually protruded away from the framework (Figure S2). The apparently destabilized loop-framework geometry suggests that the framework sequence

is not fully compatible with the mutated CDR sequences, and needs to be optimized. We therefore performed framework sequence design on v186, and the top-ranked variant v186\_Fr was predicted to form additional H-bonds with CDR3 residues through the F37Y mutation (Figure 3.3C, Top). In addition, the A12V mutation also apparently increases the hydrophobic shielding of the  $\beta$ -barrel core (Figure 3.3C, Bottom). Interestingly, when the same framework sequence design was performed on v149, different from the v186 design, the A12V/F37Y mutations were predicted to be less energetically favorable than the parent v149, suggesting that the v186 mutations were a prerequisite for the A12V/F37Y mutations to be beneficial.

4NBX.B-biotin103 v186\_Fr showed significantly reduced aggregation on SEC. Collected fractions excluding the aggregates peak did not re-aggregate once rerun on SEC (Figure 3.3A, S3B). SPR measured the  $K_D$  of v186\_Fr to be  $0.20 \pm 0.03$  nM, which preserved the >20-fold  $K_D$  improvement from biotin/mSA<sub>WT</sub> (Figure 3.3B top panel, 3.7A). The kinetics profile of v186\_Fr against mSA<sub>WT</sub> was also similar to v149 (Figure 3.3B top panel, 3.7A). When binding to mSA<sub>S27A</sub>, v186\_Fr exhibited  $K_D$  to be  $54 \pm 3$  nM, indicating a ~20-fold  $K_D$  improvement contributed by both improved association rate and dissociate rate (Figure 3.3B bottom panel, 3.7B).

To further investigate the functionally relevant structural features of v186 and v186\_Fr, we performed additional 100 ns MD simulations of v186 and v186\_Fr against mSA<sub>WT</sub> in triplicate. In general, during the simulations both the overall binding geometry of the conjugates and the conformation of the biotin103 side chain remained constant with small structural RMSDs (Figure S4, 3.4B first panel). The 4NBX.B nanobody scaffold has two solvent-inaccessible clusters of hydrophobic residues in the framework, one being the  $\beta$ -barrel core and another shielded by the CDR3 loop (Figure 3.4A). Stable solvent inaccessibility and packing of hydrophobic patches is usually correlated with protein folding stability, which is in turn related to aggregation [9,37]. For the majority of time in the MD simulations, the solvent-accessible area for the two hydrophobic clusters of both v186 and v186\_Fr was distributed around similarly low values, indicating that both variants

should be generally foldable (Figure 3.4A). However, in contrast to v186\_Fr, v186 displayed apparent sub-populations whose hydrophobic core and CDR3-shielded hydrophobic residues were significantly more solvent-accessible, indicating possible structural instability that agrees with the expected stabilization effects of F37Y and A12V in v186\_Fr (Figure 3.4A). Additional analysis of the v186\_Fr/mSA<sub>WT</sub> interface from the simulations indicates high shape complementarity, large buried interface area, and close interface distance that remained generally constant along the timescale, in agreement with the measured sub-nanomolar affinity (Figure 3.4B). Overall, the design calculation, experimental data, and MD simulations are well-correlated with each other.

Affinity and kinetics estimation of 4NBX.B-biotin103 WT, v119, v149, and v186\_Fr were performed in biological triplicates. To make sure the prepared conjugates homogeneously harbor one biotin-maleimide “side chain” per nanobody molecule, we used intact-protein mass spectrometry (MS) to analyze one of the SPR-measured triplicates for each of the above-mentioned nanobody-biotin variants, as we reason one replicate should be representative given the small batch-to-batch variations in measured affinities (Figure 3.7A-B). Deconvolution of MS spectra only returned components with molecular weights (MWs) within 20 Da from the expected values of mono-biotin conjugates, while each conjugated biotin-maleimide “side chain” would add an additional mass of 366 Da, indicating that all tested materials were effectively mono-conjugated with biotin C2 maleimide (Figure S6). Subpopulations with  $\sim\pm 17$  Da from the expected MWs were observed, and could be contributed by ring-open products of succinimides or ion adducts (Figure S6).

Although the affinity and kinetics improvements are well correlated to the designed mutations, confirming whether the predicted interactions were accurately established require structural determination. Crystallization attempts using mSA<sub>WT</sub> in complex with v186\_Fr and v119 only produced crystalline that failed to increase in size. This observation and the fact that the affinities of the designed conjugates in this study are predominantly affected by the biotin-binding affinity potentially indicate that the protein-protein

interactions are relatively flexible and dependent on the biotin anchoring, suggesting that further improvements over the protein-protein interface are possible.

### Constructing and testing a yeast display workflow for directed evolution

Yeast display is a powerful tool to perform affinity maturation on antibody fragments, because of its eukaryotic expression system and compatibility with fluorescence-activated cell sorting (FACS) [20-22]. However, the traditional yeast display format is not compatible with the conjugates introduced in this study. First, antibodies are usually displayed through the Aga1p-Aga2p interaction that contains two disulfide bonds [23], which would be compromised by the maleimide conjugation reaction, as a TCEP reduction step is required to free up cysteine sulfhydryl handle [24]. Second, cysteine residues are ubiquitous on yeast surface and will result in a significant amount of background labeling [25,26], which will likely contribute to nanobody-unspecific binding signals that mask the binding event of interest. To solve the first problem, we employed an alternative published yeast display system that does not rely on disulfides to anchor onto cell surface [34]. To solve the background labeling problem, we chose to display our best designed conjugate v186\_Fr, in a hope that the affinity difference would result in a high signal-to-noise ratio on FACS.

4NBX.B-biotin103 v186\_Fr was successfully displayed on yeast cell surface at high expression level (Figure 3.5). An on-cell maleimide labeling protocol was constructed to conjugate biotin C2 maleimide on the displayed nanobody. Taking advantage of the slow dissociation rate of v186\_Fr, we introduced a wash step after incubating the labeled cell with mSA to further eliminate background binding events. As a control experiment, a parallel group of cells were incubated with WT streptavidin and washed in the identical way. Because WT streptavidin almost irreversibly binds to biotin, it would likely bind to all biotinylated yeast surface protein with similar strength and can therefore reveal the level of unspecific labeling. As expected, cells incubated with WT streptavidin showed almost

overlapping target-binding signal for cells with or without v186\_Fr expression (Figure 3.5A). By contrast, cells displayed with v186\_Fr showed clear mSA binding signal that is proportional to the nanobody expression signal, while cells without expression showed virtually no binding signal against mSA (Figure 3.5B). This observation confirmed that unspecific labeling on yeast cells is significant for cysteine-maleimide reaction, but the designed conjugate v186\_Fr binds to mSA at a strong affinity that allows the target protein to differentiate the conjugate from background binding partners. Therefore, an engineering process that first designs a conjugate and then improves it by directed evolution is a potential viable route.

To test whether v186\_Fr can be further improved by this yeast display platform, we conducted a trial directed evolution experiment by constructing an error-prone PCR library that samples both high and low mutation rates on the full nanobody sequence. To identify variants with further improved dissociate rate, four rounds of off-rate screening was performed on the library, and a variant with six additional mutations G10R/M34T/T58S/A61V/T91R/A108T, named as v186\_Fr.1, was overwhelmingly enriched (Figure 3.6A). v186\_Fr.1 was highly expressed on yeast surface, but had very low yield and significantly aggregated when expressed in *E. Coli*, indicating low stability (Figure 3.6B). This is surprising, since the display level of antibody fragments in yeast display is usually well-correlated to foldedness [20-22]. Nevertheless, after expressing a large quantity of v186\_Fr.1, a sufficient amount of material was collected from the monomeric peak position on SEC to run an SPR experiment (Figure 3.6B). On SPR, v186\_Fr.1 showed a ~6-fold slower  $k_d$  than the parent, in agreement with yeast display result (Figure 3.6B).

Overall, the directed evolution trial demonstrated that the binding property of designed conjugates can be further improved by *in vitro* approaches. However, the disagreement between display level and protein stability may disfavor strong selection stringency and high mutation rates being implemented in evolution efforts. Although the disagreement in stability could be caused by the difference in prokaryotic and eukaryotic expression

systems, the high compatibility with bacterial expression is a unique advantage of nanobodies and should be preserved. Therefore, in future evolution experiments on designed conjugates, it would be favorable to introduce mutations in a more incremented manner with milder selection stringencies.

### Summary and further testing of a computational workflow for creating synergistically-binding nanobody small-molecule conjugates

Based on the above design results, we summarized the design process into the following general workflow (Figure 3.7C): docking a library of nanobody structures with diverse CDR sequences and conformations onto a desired target in complex with the to-be-conjugated small molecule, filtering binding poses to preserve ones that closely resemble the original binding mode of the original nanobody scaffold, screening the rotamer library of the conjugated small molecule onto the poses to identify most tolerable conjugation plan, and finally re-designing the sequences of both the nanobody CDR loops and framework to improve binding affinity, kinetics, and overall stability. Because 4NBX.B was not obtained by directly docking nanobody scaffolds against mSA, we re-performed the docking, filtering, and rotamer screening steps on mSA, and selected a different scaffold, 2X89.A, with biotin conjugated to site 57. Similar to 4NBX.B-biotin103 v186\_Fr, the selected pose of 2X89.A was predicted to interact with mSA<sub>WT</sub> through a R-E interaction, together with other potential intermolecular H-bonds (Figure S7A top panel). Since the original 2X89.A has an additional intra-CDR disulfide bond, to avoid over conjugation, the disulfide bond was replaced by two alanine residues. The resulted final conjugate, 2X89.A-CCAA-biotin57 binds to mSA<sub>WT</sub> with a  $K_D$  of  $0.8 \pm 0.2$  nM, and remarkably, a  $k_d$  that is slightly better than our best designed 4NBX.B variant v186\_Fr (Figure 3.7A, S7A bottom panel). 2X89.A-CCAA-biotin57 aggregated obviously on SEC (Figure S7B). To reduce aggregation, we constructed a rudimentary sequence design pipeline that sways between CDR and framework design based on our previous experience on designing 4NBX.B conjugates, and applied the pipeline on 2X89.A. Top-ranked variants along the six rounds

of CDR designs and one round of framework design showed first worsened then improved aggregation profile after 18 mutations were accumulated (Figure S7C), similar to what we observed in the design process of 4NBX.B conjugates.

### ***Conclusion***

Using mSA/biotin system, we demonstrated for the first time to our knowledge that with the sole structural information of a small molecule binding to its target, a complementary immunoglobulin domain conjugating to the small molecule can be designed entirely by computational methods to bind tighter against the target, further bridging the two worlds of small molecules and biologics. The binding interface for the designed conjugates comprise of both an ultra-deep pocket that is uncommon for antibodies, and broad contacting interface that is uncommon for small molecules [38,39]. Therefore, the chemical space and target space of traditional molecular recognition agents could be expanded in this manner, offering new potential solutions to a wide range of challenges, such as reutilizing failed small molecules or tackling undruggable targets in pharmaceutical development. Our results showed that the affinity, kinetics, and stability of the conjugates can be designed in a step-wise manner, indicating that the development process is highly tunable and multiple physicochemical properties can be simultaneously optimized.

Testing whether the design strategy introduced in this study works for therapeutically relevant targets would be a crucial next step. It will also be beneficial to study whether the workflow works with virtually docked small-molecule/target complexes. In addition, testing whether the workflow can engineer specificity in addition to affinity will be also highly desirable. There are many computational methods that could be used to improve the design strategy. Virtually recombining structural fragments was reported to help affinity maturation of computationally designed antibodies [26,40,41]. Specifically tailored algorithms that put more bias in the formation of hydrogen-bonding networks were also proven to be useful to the affinity and specificity of designed protein/protein interfaces [42].

Advanced loop-modeling methods and ensemble design could also facilitate more accurate assessment of binding poses for the conjugates, and potentially engineer specificities [43,44].

Because the designed conjugates have the CDR loops chemically extended beyond the natural repertoire, we name the computationally designed synergistically-binding antibody small-molecule conjugates to be CDR-extended antibody, abbreviated as CDRxAb.

### ***Materials and Methods***

Computational design workflow for nanobody-biotin conjugates: The detailed description of the computational design workflows for the nanobody-biotin conjugates introduced in this study is in Chapter 2.

Plasmids, expression cell lines, and cloning of protein variants: pRSET-mSA was a gift from Sheldon Park (Addgene plasmid # 39860) [28]. S27A mutation was created by site-directed mutagenesis using commercially available kits (NEB). 4NBX.B\_C103 and 2X89.A\_CCAA\_C57 sequences were directly ordered from IDT, and cloned into pHen6c vector by Gibson assembly using commercially-available reagents (NEB) [45]. The assembled pHen6c vectors harbor a PelB signal sequence before the N terminus of the nanobody sequence, allowing bacterial periplasmic expression [46]. Variants of 4NBX.B\_C103 and 2X89.A\_CCAA\_C57 were created by mutagenic PCR and assembled into pHen6c vector by Gibson assembly using commercially available reagents (NEB) [45]. 4NBX.B WT sequence with C103A mutation was created by site-directed mutagenesis using commercially available kits (NEB). Smt3 SUMO protein with an N-terminal cysteine was created from wild type Smt3 SUMO by mutagenic PCR, and subcloned into pY71A(lc) vector by Gibson assembly using commercially available reagents (NEB) [45]. Yeast display vector pYDS6<sub>49</sub>HM was a kind gift from Andrew Kruse [34]. The DNA sequence of v186\_Fr together with the linearized pYDS6<sub>49</sub>HM was electroporated into yeast strain BJ5465 (ATCC), and then recombined *in vivo* to construct the full plasmid. For

cytoplasmic expression of v186\_Fr.1, the DNA sequence was assembled into pET28b vector by Gibson assembly using commercially available reagents (NEB) [45].

Expression and purification of mSA streptavidin wild type and S27A variant: Expression, purification, and refolding of mSA variants followed published protocols with slight variations [28]. The expression plasmids were first transformed to *E. Coli* BL21-Gold (DE3) chemically competent cells (Agilent), which were then grown overnight in LB with 100 µg/mL of ampicillin (amp 100) at 37°C and 250 rpm. 1 mL of the overnight culture was used to inoculate 300 mL of TB medium (2.3 g KH<sub>2</sub>PO<sub>4</sub>, 16.4 g K<sub>2</sub>HPO<sub>4</sub>, 12 g tryptone, 24 g yeast extract, 4 mL glycerol, dissolved in water to 1 L volume) supplemented with 2 mM MgCl<sub>2</sub>, 0.1% glucose, and amp 100. Inoculation was done at 37°C and 250 rpm until OD<sub>600</sub> hit 1.5-2. Expression was induced by 1 mM IPTG at 28°C and 250 rpm for 18 hours. Cells were then centrifuged by 4500 g for 15 minutes at 4°C, and protein extraction was then performed using 50 mL of chemical lysis buffer. The buffer was composed of 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4), 1x CelLytic B reagent (Sigma), 0.02 mg/mL DNase1, 0.2 mg/mL lysozyme, and 1 mM protease inhibitor AEBSF (Sigma). Pellets were resuspended in lysis buffer and nutated for 4 hours at room temperature. Cell lysate was then centrifuged at 15000 g for 30 minutes at 4°C, and the precipitates were used to refold and purify the protein.

Precipitates were first resuspended into 3 mL of 6 M guanidine hydrochloride in 1x TBS (50 mM Tris, 150 mM NaCl, pH 8.0) and incubated under 37°C for 30 minutes to solubilize the proteins. Un-dissolved materials were cleared by 15000 g centrifugation for 5 minutes at 4°C. Supernatants were then chilled on ice before added drop by drop into 40 mL of pre-chilled refolding buffer (50 mM Tris-HCl, 150 mM NaCl, 0.3 mg/mL D-biotin, 0.2 mg/mL oxidized glutathione, 1 mg/mL reduced glutathione, pH 8.0) while stirring. The refolding buffer with added mSA protein was then allowed to incubate on ice for another 2 hours with stirring before centrifuged by 15000 g for 30 minutes at 4°C to remove insoluble materials. The supernatants were then supplemented with 20 mM imidazole and then loaded onto 1 mL bed volume of Ni-NTA agarose beads (Qiagen) which were pre-washed

with 5 column volumes of 1x PBS. Sample loading was performed by gravity flow. Column was then washed with 10 column volumes of 1x PBS supplemented with 20 mM imidazole before 3 mL of 1x PBS supplemented with 500 mM of imidazole was used to elute the proteins. The entire Ni-NTA purification process was done at 4°C. The 3 mL of purified proteins were then dialyzed against fresh 1 L of 1x PBS (pH 8.0) under 4°C for 3 times using Slide-A-Lyzer 10 kDa molecular-weight cutoff dialysis cassettes (Thermo). Each dialysis step took longer than 4 hours. The dialyzed mSA products were further purified at 4°C by Superdex 75 10/300 GL SEC column (GE) using 1x PBS (pH 7.4) as running buffer, and the fractions corresponding to the monomeric peak were collected for subsequent experiments.

Expression and purification of Smt3 SUMO protein with N-terminal cysteine: The expression plasmids were first transformed into E. Coli BL21-Gold (DE3) chemically competent cells (Agilent), which were then grown overnight in LB with amp 100 and 1% glucose at 37°C and 250 rpm. 1 mL of the overnight culture was then used to inoculate 300 mL of TB medium with 2 mM MgCl<sub>2</sub>, 0.1% glucose, and amp 100 at 37°C and 250 rpm until OD<sub>600</sub> hit 1.5-2. Expression was then induced by 1 mM IPTG at 28°C and 250 rpm for 18 hours. Cells were then pelleted under 4500 g for 15 minutes at 4°C, resuspended in 50 mL of chemical lysis buffer supplemented with 5mM 2-mercaptoethanol (BME), and incubated for 4 hours at room temperature to extract the expressed proteins. Lysate supplemented with 20 mM imidazole was cleared by 15000 g centrifugation at 4°C for 30 minutes, and the supernatant was loaded onto 1 mL bed volume of Ni-NTA agarose beads (Qiagen) pre-washed with 5 column volumes of 1x TBS (pH 7.3). Sample loading was performed by gravity flow. The loaded column was then washed with 5 column volumes of 1x TBS (pH 7.3) supplemented with 20 mM imidazole and 5 mM BME and another 5 column volumes of 1x TBS (pH 7.3) supplemented with 20 mM imidazole. 3 mL of 1x TBS (pH 7.3) supplemented with 500 mM of imidazole was used to elute the proteins. All Ni-NTA purification procedures were performed under 4°C. Purified proteins were then concentrated to ~0.5 mL using Amicon 10 kDa molecular-weight cutoff centrifuge filters (GE), and stored for subsequent experiments.

Cytoplasmic expression and purification of 4NBX.B-biotin103 v186\_Fr.1: v186\_Fr.1 in pET28b was chemically transformed in the Shuffle T7 Express cell line (NEB). The overnight culture preparation, inoculation, induction, and expression procedures were identical to those of Smt3 SUMO with N-terminal cysteine. Biotin conjugation was performed in the same way as the designed conjugates, except that PD-10 column was used for buffer exchange after both the TCEP reduction and maleimide reaction steps.

Periplasmic expression and purification of nanobodies: The pHen6c expression plasmids were first transformed into E. Coli BL21-Gold (DE3) chemically competent cells (Agilent), which were then grown overnight in LB with amp 100 and 1% glucose at 37°C and 250 rpm. 1 mL of the overnight culture was then used to inoculate 300 mL of TB medium with 2 mM MgCl<sub>2</sub>, 0.1% glucose, and amp 100 at 37°C and 250 rpm until OD<sub>600</sub> hit 1.5-2. Expression was then induced by 1 mM IPTG at 28°C and 250 rpm for 18 hours. Cells were then pelleted under 4500 g for 15 minutes at 4°C, and suspended in 12 mL of TES periplasmic extraction buffer (0.2 M Tris, 0.5 mM EDTA, 0.5 M sucrose, pH 8.0), supplemented with 5 mM BME if the nanobody had a cysteine handle for conjugation, before incubated on ice with shaking at 32 rpm for 1 hour [47,48]. 18 mL of 4x diluted TES buffer, supplemented with 5mM BME if the nanobody had a cysteine handle, was then added to the cells which were incubated on ice at 32 rpm for another hour [47,48]. After periplasmic extraction, the cells were pelleted by 15000 g at 4°C for 30 minutes, and the supernatants were supplemented with 20 mM of imidazole before loaded onto 1 mL bed volume of Ni-NTA agarose beads (Qiagen) pre-washed with 5 column volumes of 1x TBS (pH 7.3). Sample loading was performed by gravity flow. The loaded column was then washed with 10 column volumes of 1x TBS (pH 7.3) supplemented with 20 mM of imidazole, before 3 mL of 1x TBS (pH 7.3) supplemented with 500 mM of imidazole was used to elute the proteins. For nanobodies with a cysteine handle, 5 mM BME was added to the first 5 column volumes of wash buffer, and the elution buffer was supplemented with 5 mM TCEP. The elution buffer was incubated with the beads for 30 minutes before eluting the proteins. The entire Ni-NTA purification process was done at 4°C. Purified nanobodies with a cysteine handle in 3 mL of the elution buffer were concentrated to ~0.5 mL using

Amicon 10 kDa molecular-weight cutoff centrifuge filters (GE), and stored for subsequent experiments. 4NBX.B WT was instead further purified by Superdex 75 10/300 GL SEC column (GE) using 1x PBS (pH 7.4) as running buffer, and the fractions corresponding to the monomeric peak were collected for subsequent experiments.

Biotin C2 maleimide conjugation and purification of conjugates: Maleimide labeling on surface cysteines of nanobodies followed a published protocol with some modifications [49]. Purified nanobodies in storage were first incubated with another 5 mM TCEP supplement under 4°C for 2 hours, and then buffer exchanged to 1x TBS (pH 7.3) using HiTrap desalting columns (GE) under room temperature to remove TCEP. Thawed stock solutions (100 mM in DMSO) of biotin C2 maleimide (AnaSpec) were immediately added to the buffer-exchanged nanobodies to 1 mM final concentration before the reaction mixture was nutated under 4°C for 4 hours with tinfoil cover to avoid light contact. Excess maleimide stock solutions were tossed away and not re-frozen for future experiments. Biotin C2 maleimide was in >20-fold molar excess over the nanobody in the reaction mixture. Finished reaction mixture was then filtered by 0.2 µm syringe filters (Thermo) to remove precipitated proteins, and then buffer exchanged to 1x TBS (pH 7.3) using PD-10 desalting columns (GE) under room temperature to remove excess maleimide reagents. The labeled nanobodies were further purified at 4°C by Superdex 75 10/300 GL SEC column (GE) using 1x TBS (pH 7.3) as running buffer, and the fractions corresponding to the monomeric peak were collected for subsequent experiments. Maleimide labeling of Smt3 SUMO protein with N-terminal cysteine followed the identical procedures as above.

Yeast display and off-rate screening: An error-prone PCR library was constructed on the WT v186\_Fr sequence in the display vector. A series of MnCl<sub>2</sub> concentrations from 50 µM to 300 µM were used to introduce both low and high mutation rates in PCR reactions performed in parallel. PCR products were then pooled together and electroporated in BJ5465 yeast with linearized display vector, resulting in ~500,000 transformants. Yeast cells harboring either v186\_Fr WT or the error-prone PCR library were grown in SD-CAA media (Teknova) supplemented with uracil at 30°C/250rpm, until OD600 between 1 and

10 was reached. The saturated growth culture was then pelleted down and resuspended in SG-CAA induction media (Teknova) supplemented with uracil at OD<sub>600</sub>=1, and expressed at 20°C/250rpm for 60 hours. For maleimide labeling reaction, 10<sup>7</sup> expressed cells were first incubated with 5mM of TCEP under 4°C for 30 minutes. Then, the cells were washed with TBS (pH 7.3)/0.1% BSA to remove residual TCEP. The washed cell pellets were then immediately resuspended in 1mL of TBS (pH 7.3)/0.1% BSA and 1mM Biotin 2C maleimide, and the reaction was left to proceed for 4 hours with tin-foil cover under 4°C. Labeled cells were then washed with TBS (pH 7.3)/0.1% BSA to remove excess maleimide reagents.

To compare binding signals between mSA and WT streptavidin, mSA and WT streptavidin at equivalent concentrations of binding sites were respectively incubated with identical amount of yeast cells displaying conjugated v186\_Fr WT. WT streptavidin and mSA are both coupled with AF647 fluorescence dye. After overnight incubation at 4°C, the cells were pelleted down and washed in TBS (pH 7.3)/0.1% BSA for 3.5 hours. Anti-HA antibody with AF488 tag was also incubated in the wash buffer to assess expression level. After the wash step, the cells were further washed to remove the anti-HA antibody. The first two rounds of off-rate screening were performed by incubating the conjugated yeast cells with mSA-AF647, and resuspending the incubated cells in 50 mL of TBS (pH 7.3)/0.1% BSA at 4°C. Cells were harvested after 8 hours and 24 hours respectively for the first and second rounds. The third and fourth rounds of screening were performed by first incubating yeast cells with mSA-AF647, then resuspending the cells with 10 μM of mSA without fluorescence tag supplemented. With anti-HA antibody (AF488). Cells were harvested after 27 hours and 1 week respectively for the third and fourth rounds. Stained yeast cells were analyzed and sorted by a Sony SH800 cell sorter, according to the AF488 and AF647 fluorescence intensity that respectively represents nanobody expression and target binding.

Intact protein mass spectrometry (MS) workflow to analyze conjugation efficiency: HPLC-MSD (HP, Agilent) was used to assess the labeling efficiency of prepared nanobody-biotin conjugates. Conjugates were first dried out using a spin vacuum evaporator, and

resuspended in 0.2% formic acid. A C3 HPLC column was used first to separate the protein sample before MS analysis. Before running samples, the column was first washed with isopropyl alcohol (IPA) to clean the column and also reveal background peaks irrelevant to our samples. A analysis of conjugation efficiency of 4NBX.B-based conjugates was performed by deconvoluting the eluted sample HPLC peak using the following parameters: positive adduct ion +H 1.0079 Da, negative adduct ion -H -1.0079 Da, molecular weight cutoff 5000-80000 Da, maximum charge 90, minimum peaks 5, ion PWHH 0.6 Da, molecular weight agreement 0.05%, noise cutoff 0, abundance cutoff 10%, molecular weight assignment cutoff 40%, and envelope cutoff 50%. Deconvolution was performed in ChemStation (Agilent). For each sample of interest, about 0.1-1  $\mu\text{g}$  of material was used for the above analysis.

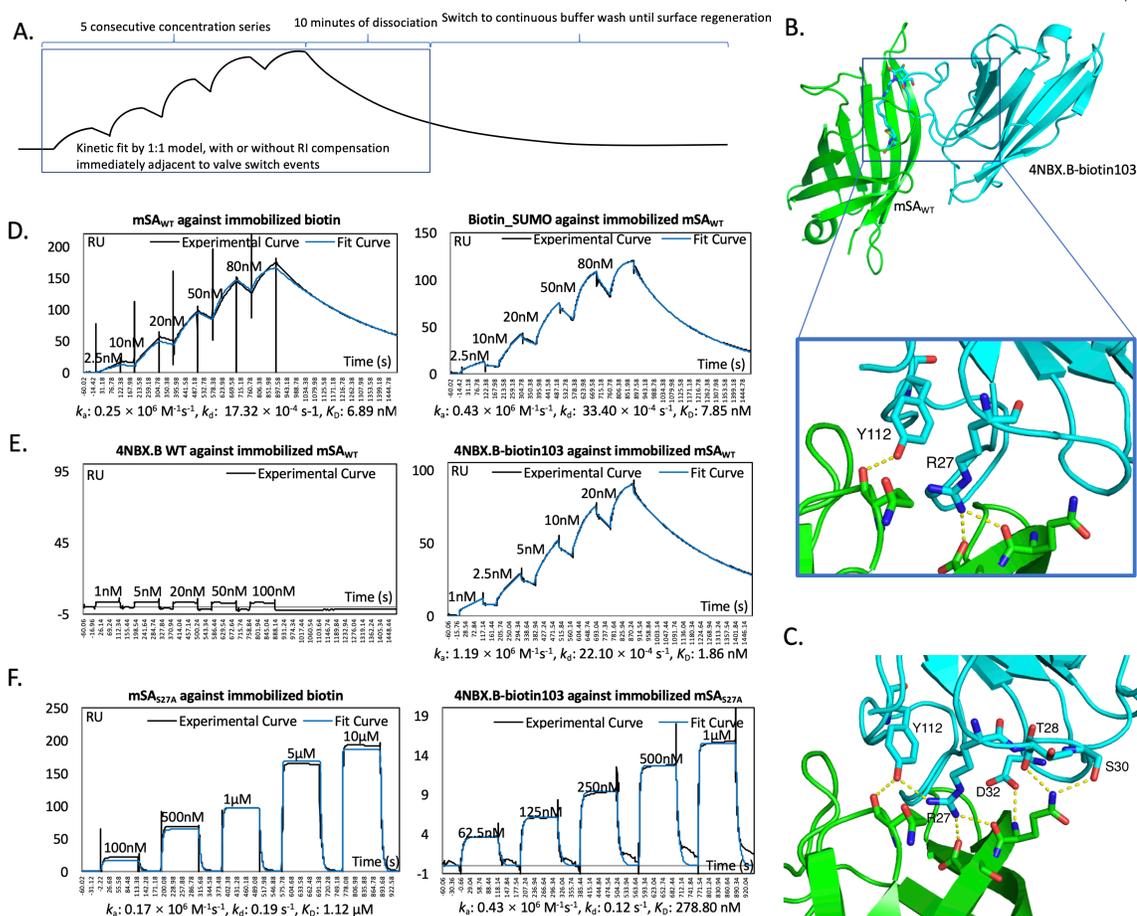
Surface plasmon resonance (SPR) analysis of binding affinity and kinetics: A Biacore T200 instrument (GE) was used to perform SPR analysis. 4NBX.B WT, 4NBX.B-biotin103 conjugates, and 2X89.A-CCAA-biotin57 conjugates were first buffer-exchanged to HBS-EP+ buffer (Teknova) using Amicon 10 kDa molecular-weight cutoff centrifuge filters (GE). The concentrations of the conjugates were then determined by BCA assay using commercially-available kits (Thermo). The calibration curve for BCA assay was prepared using purified 4NBX.B WT, which was also buffer exchanged to HBS-EP+ but had concentrations determined by  $A_{280}$  readings using extinction coefficient  $30035 \text{ M}^{-1}\text{cm}^{-1}$ . For SPR analysis, biotin pentylamine (Thermo), mSA<sub>WT</sub>, and mSA<sub>S27A</sub> were respectively immobilized on CM5 sensor chip (GE) by EDC/NHS amine coupling kit following standard protocol (GE). Binding kinetics were measured by kinetic titration, or single-cycle kinetics (GE). Biotin pentylamine was immobilized at 7.5 mM concentration to reach target surface density of  $\sim 200$  resonance units (RUs) [50]. In order to compare how binding events changed in response to different surface densities, surfaces with three different densities of immobilized mSA<sub>WT</sub> at  $\sim 1000$  RU,  $\sim 2500$  RU, and  $\sim 3000$  RU were respectively prepared under immobilization concentrations 0.1  $\mu\text{M}$ , 0.5  $\mu\text{M}$ , and 1  $\mu\text{M}$ . Immobilization of mSA<sub>S27A</sub> was also performed at 0.01  $\mu\text{M}$  and 0.05  $\mu\text{M}$  concentrations with target surface density of  $\sim 200$  and  $\sim 600$  RU. The fitted affinities and kinetics of

identical conjugates against the different densities were minimally different. Reference channels were either treated with EDC/NHS using blank HBS-EP+ buffer, or 1  $\mu$ M of 4NBX.B WT to assess if the conjugates would self-associate and roughly see if the conjugates un-specifically interacted with proteins not of interest. No visible signal differences against reference channels with or without immobilized 4NBX.B WT were observed for various tested conjugates. All immobilization samples were dissolved in acetate buffer (pH 4.5).

Binding experiments were performed under 25°C. HBS-EP+ was used as running buffer. The flow channels were first incubated in the running buffer before analytes at 5 different concentrations were consecutively injected at 30  $\mu$ L/min flow rate through both the reference channel and the sample channel with immobilized molecules of interest. After injections, the surface-bound analytes were allowed to dissociate for 10 minutes to generate dissociation curves. HBS-EP+ buffer then washed through both reference and sample channels continuously to allow the rest of the bound analytes to dissociate, in order to regenerate the surfaces for next binding experiments. Curve fitting of sensorgrams processed by subtracting the reference channel signal from the sample channel signal was performed in Biacore evaluation software using 1:1 kinetics model. No incompletely subtracted bulk contributions were observed in binding against immobilized biotin. For binding curves against immobilized mSA<sub>WT</sub>, global fitting of bulk shifts was turned on as small bulk shift contributes before and after each injection event were distinctively observed. Bulk shift fitting was turned off in binding curves involving mSA<sub>S27A</sub>, because potential bulk shift signals would be obscured by the kinetics curves with fast dissociation rates and therefore not distinctively visible.

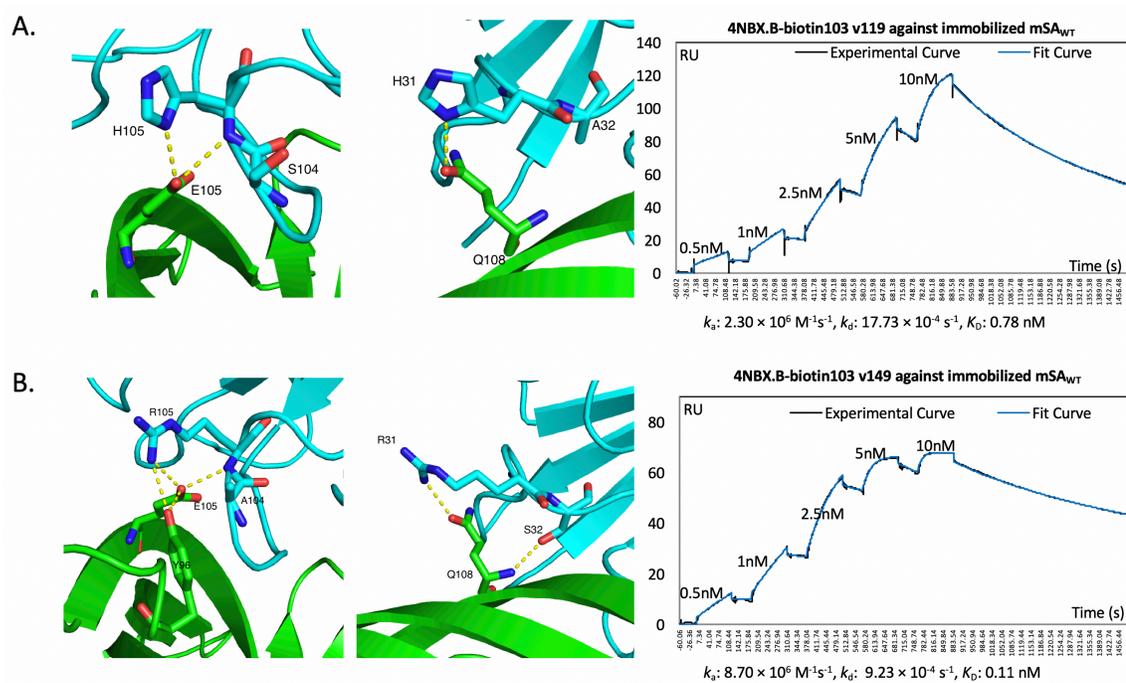
Molecular dynamics (MD) simulation protocols: Molecular Dynamics simulations were carried out using ACEMD (Acellera) [51]. Each system studied was placed in a box with dimensions selected to allow an excess length of 12 angstroms on each side. The system was solvated using the TIP3P water model [52], and ions were added to neutralize the overall charge. The built system was then minimized for 500 steps. Subsequently, a 5 ns

equilibration was completed to allow the system to reach a stationary state, and a 100 ns production run was carried out at 300 degrees K. All experiments utilized the Amber ff14SB force field and a 4 femtosecond timestep [53]. Data from the equilibration run was not included in subsequent analysis, and where replicates were collected no part of the intermediate data was reused. Parameters for the biotin-CH<sub>2</sub>-CH<sub>2</sub>-succinimide-S-CH<sub>3</sub> “side chain” were prepared using Antechamber and utilized RESP charges calculated with Gaussian 09 [54,55]. Calculation of solvent accessible surface area was performed using MDTraj, and hydrogen bonding was assessed using a tcl script written for VMD [56, 57].

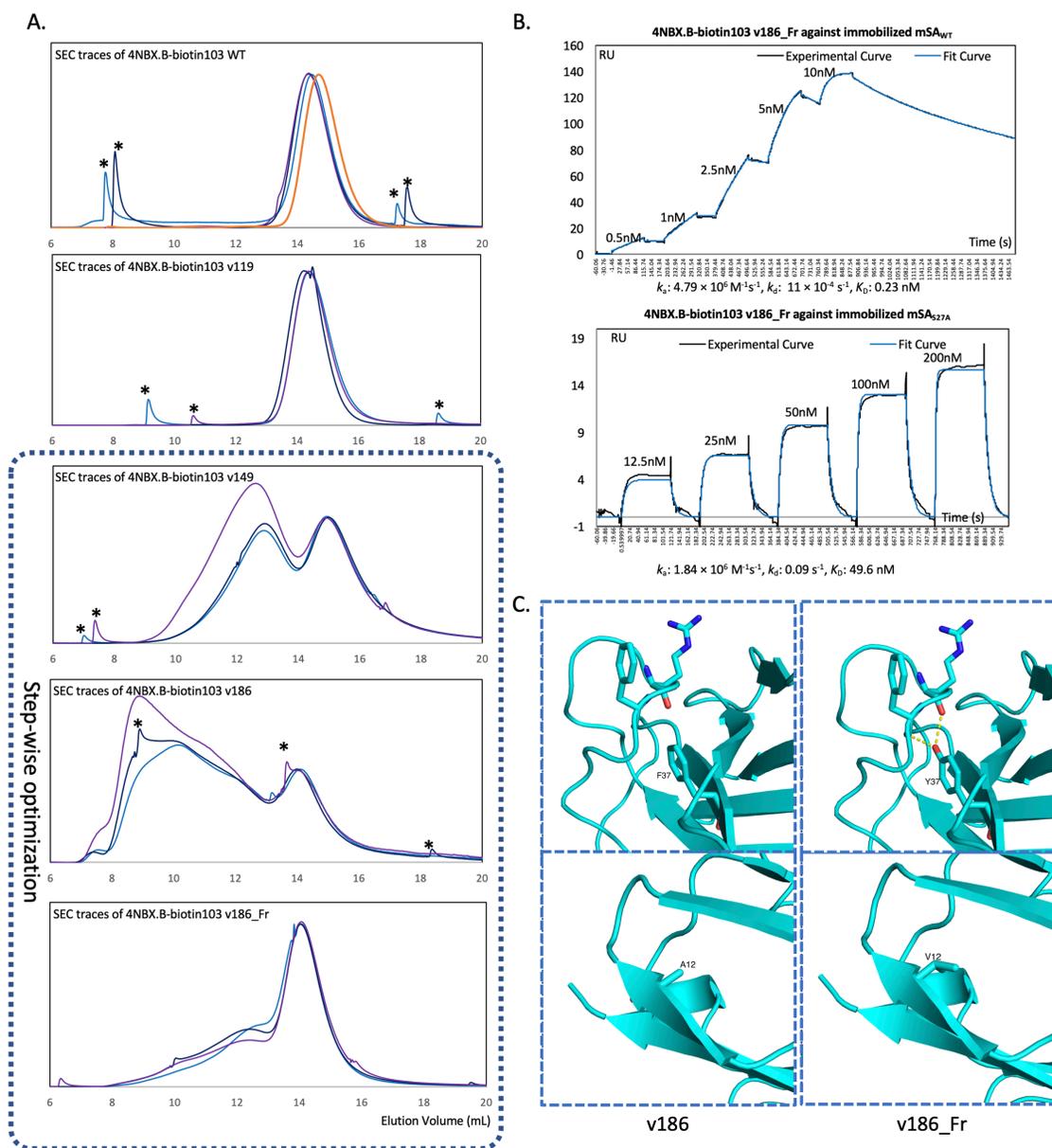


**Figure 3.1 Computationally designed nanobody-biotin conjugates bind stronger than biotin itself against mSA streptavidin:** For SPR-measured  $K_D$ ,  $k_a$ , and  $k_d$  results, data from one of the triplicates is shown here, and data from the other two replicates is in Fig. S5. (A) Schematic representation of SPR binding experiments. Binding kinetics were measured by kinetic titration or single-cycle kinetics. Five consecutive concentrations of the analyte are injected onto the immobilized ligand and then spontaneously dissociate. The kinetic curves are fitted by 1:1 binding model to report  $k_a$  and  $k_d$ , which were then used to calculate  $K_D$ . (B) Finalized model of 4NBX.B-biotin103 in complex with mSA streptavidin. The mSA is colored green, and nanobody scaffold is colored cyan. Biotin103 side chain is shown as sticks, and the H-bond forming potential of Y112 and R27 with mSA residues is also represented. (C) Y112 and R27 are predicted to participate in a broader potential H-bond network that involves biotin/mSA interactions. (D). SPR estimation of mSA/biotin binding

parameters by two methods. (E) SPR measurements determined that 4NBX.B-biotin103 occupies the biotin-binding pocket of mSA with improved affinity and kinetics. (F) SPR measurements determined that 4NBX.B\_biotin103 binds stronger towards a weaker biotin-binding mutant of mSA than biotin itself.

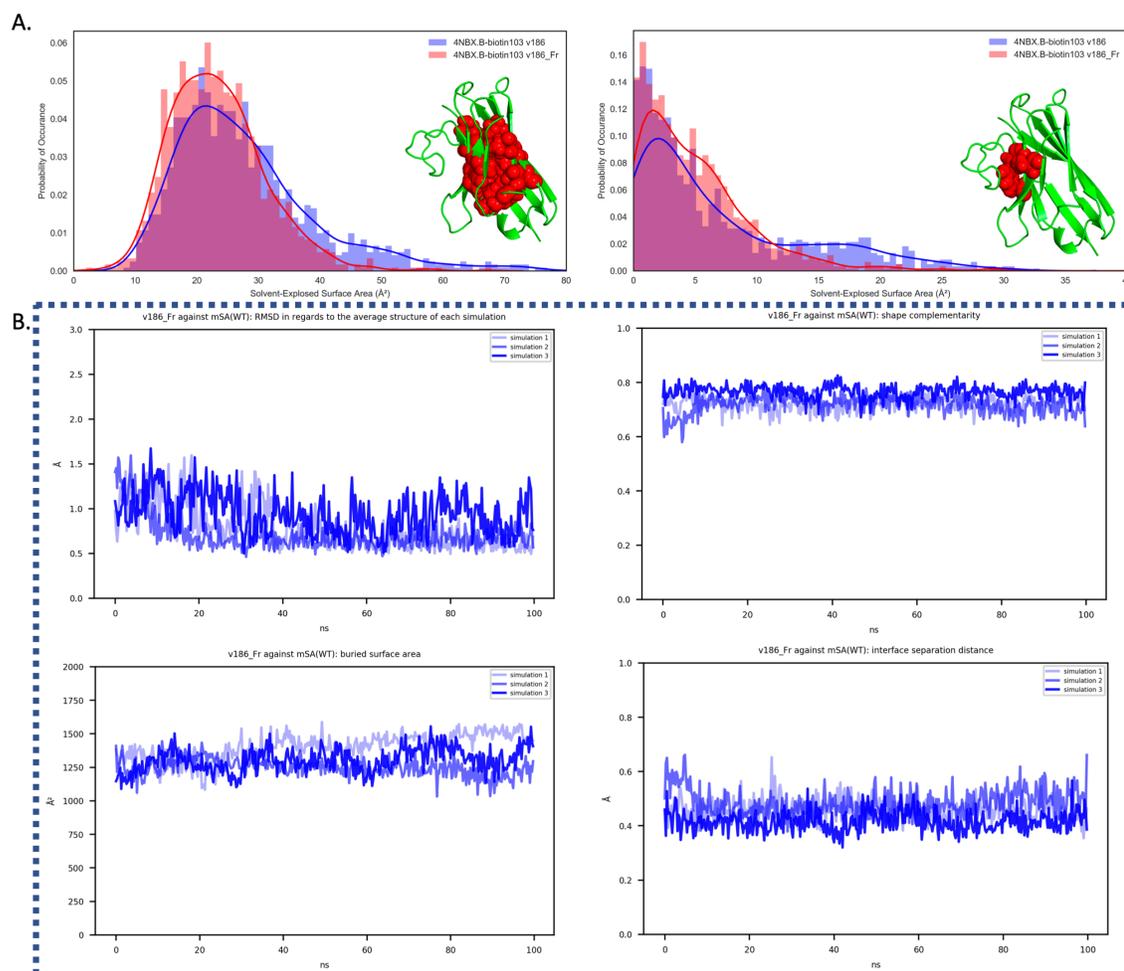


**Figure 3.2 CDR sequence design enhanced the mSA-binding affinity and kinetics of 4NBX.B-biotin103:** In the structural models, the mSA is colored green, and nanobody scaffold is colored cyan. Residues with predicted H-bond formation are shown as sticks. For SPR-measured  $K_D$ ,  $k_a$ , and  $k_d$  results, data from one of the triplicates is shown here, and data from the other two replicates is in Fig. S5. (A) Predicted affinity-contributing mutations and SPR-measured binding profiles of 4NBX.B-biotin103 v119 against mSA<sub>WT</sub>. (B) Predicted affinity-contributing mutations and SPR-measured binding profiles of 4NBX.B-biotin103 v149 against mSA<sub>WT</sub>.



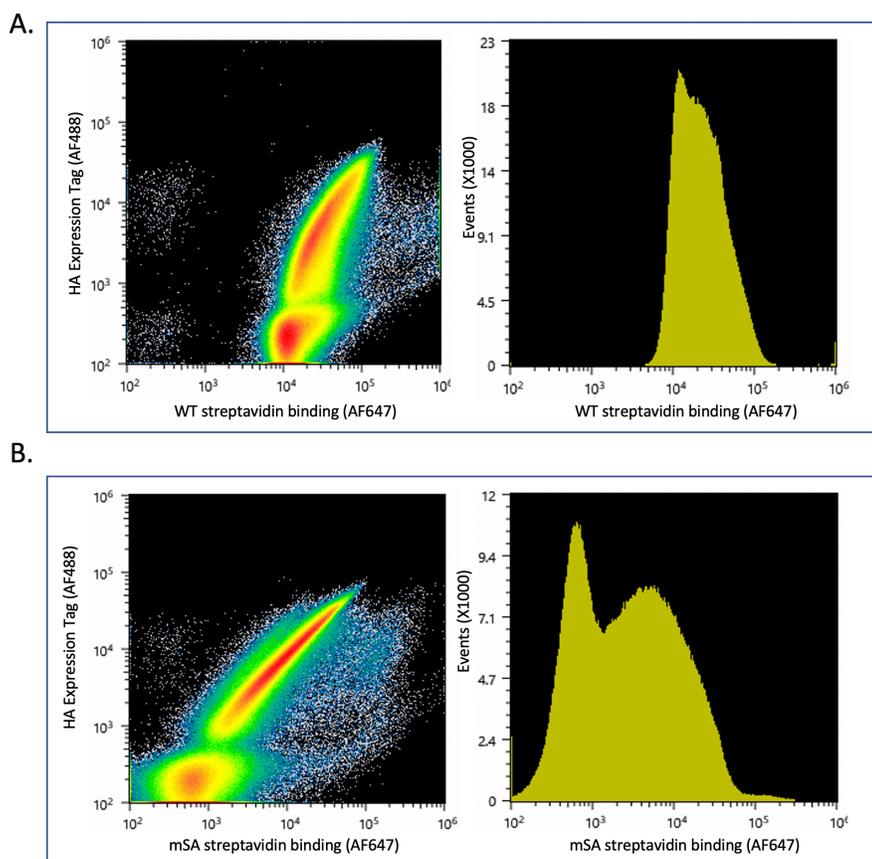
**Figure 3.3 CDR sequence design followed by framework design monomerically stabilized the designed conjugates without imposing affinity penalty:** (A) SEC traces of biological triplicates (colored by blue with different intensity) for designed 4NBX.B-biotin103 conjugates, normalized by monomer peak height for better comparison of aggregates formation. SEC trace of 4NBX.B WT is overlaid with 4NBX.B-biotin103 WT traces and colored orange. \* indicates peaks of sample-irrelevant instrument defect of the overall FPLC, please refer to Figure S3C for more details. SEC trace of 4NBX.B WT

nanobody is overlaid with 4NBX.B-biotin103 WT traces as reference. (B) SPR-measured binding profile of v186\_Fr against mSA<sub>WT</sub> and mSA<sub>S27A</sub> indicates that the improved binding affinity and kinetics in v149 are preserved. Data from one of the triplicates is shown here, and data from the other two replicates is in Fig. S5. (C) Structural representation of nanobody amino acid position 12 and 37 before and after framework redesign. v186 and v186\_Fr are colored cyan, and residues of interest are shown as sticks. Additional H-bonds introduced by F37Y with CDR3 residues are shown as dashes, while the relevant CDR3 residues are also shown in both v186 and v186\_Fr models.

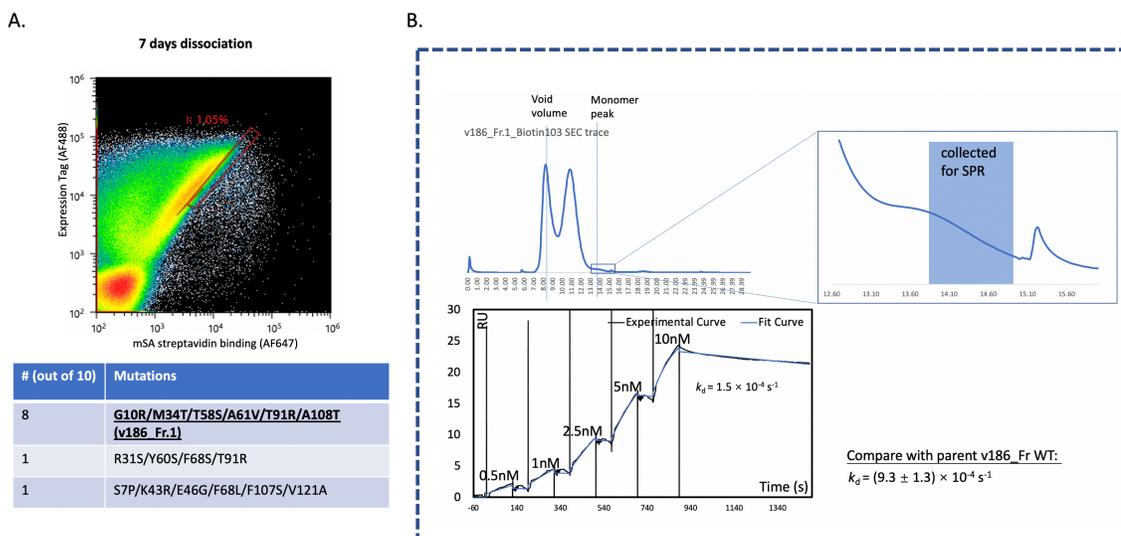


**Figure 3.4 MD simulation reveals design flaws and validates design success:** (A) MD simulation revealed possible origins for the improved monomeric stability of 4NBX.B-biotin103 v186\_Fr compared to v186. Here shows the analysis of the solvent-accessible area for the selected hydrophobic residues of v186 and v186\_Fr from 100ns MD simulations performed in triplicates. The selected residues are presented as spheres in the nanobody models shown on both panels. The observed distributions of the solvent accessible area for the selected residues from the 3X simulations of v186 and v186\_fr are plotted into 80 bins along the x-axis (bars) with respective kernel density estimation (lines). Left panel: analysis of the hydrophobic core residues. Right panel: analysis of the CDR3-shielded residues. (B) Analysis of the interaction interface in the triplicate MD simulations of 4NBX.B-biotin103 v186\_Fr against mSA<sub>WT</sub>. Traces from simulation replicates are plotted on top of each other

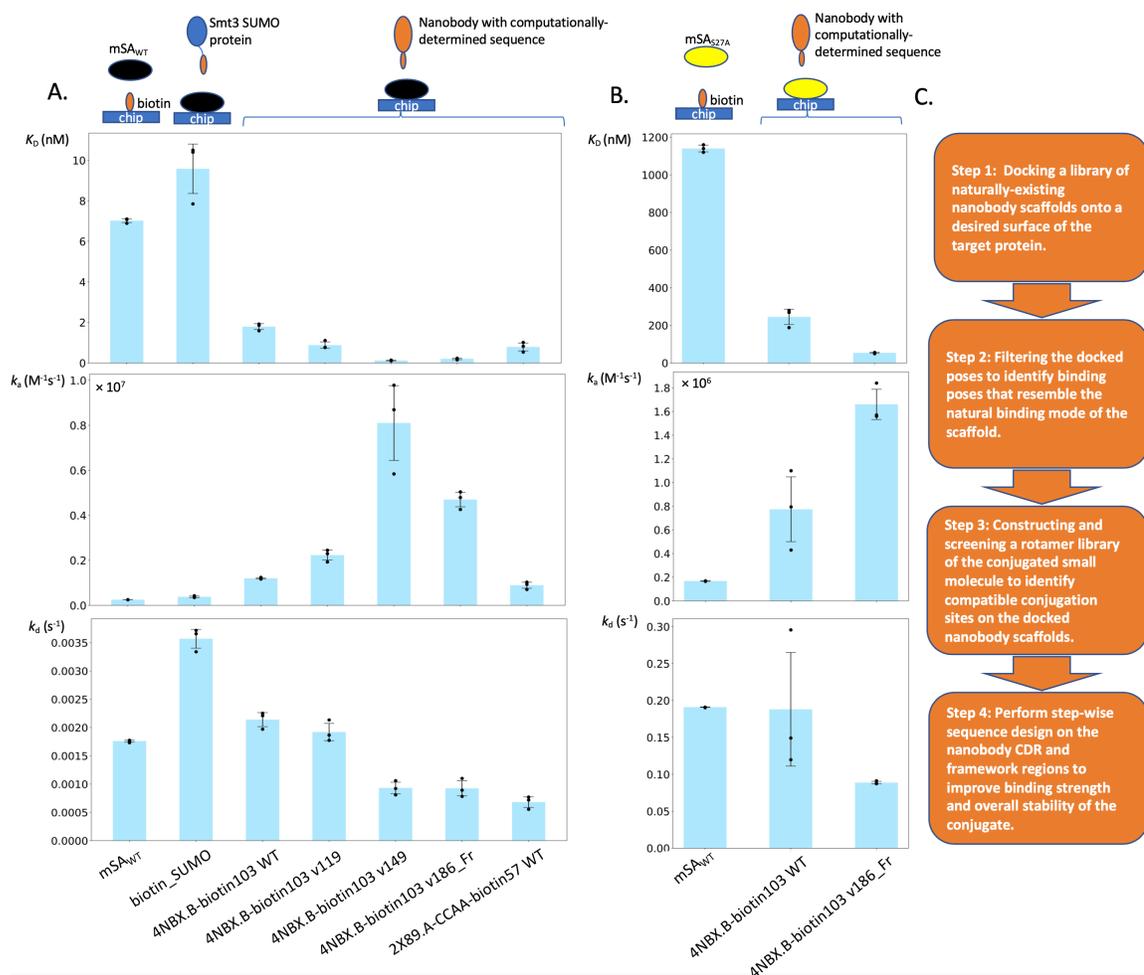
along the 100ns timescales. Changes of whole-structure RMSD, interface shape complementarity, buried surface area, and interface separation distance along the time trajectories are plotted.



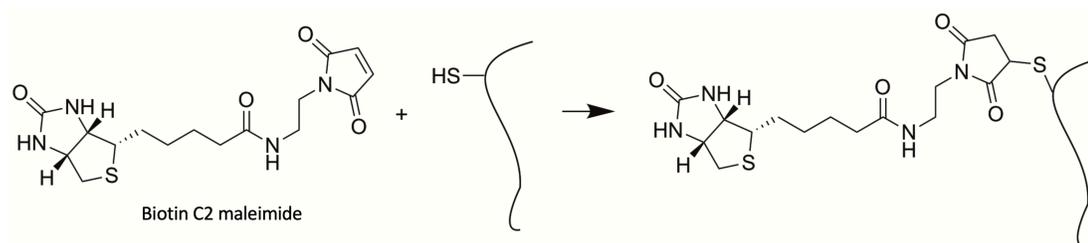
**Figure 3.5** Flow cytometry analysis of v186\_Fr WT displayed on yeast surface against WT and mSA streptavidin: These graphs show v186\_Fr incubated with either (A) AF647 tagged WT streptavidin or (B) AF647 tagged mSA streptavidin, together with AF488 tagged anti-HA antibody in both scenarios to quantify nanobody expression. AF488-AF647 correlation diagram and AF647 only event count diagram are shown for both experiments.



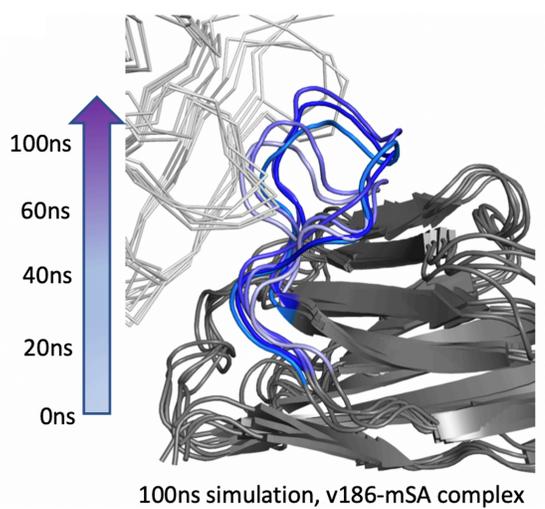
**Figure 3.6 Isolation and validation of a evolved variant of v186\_Fr:** (A) FACS sorting results after 4 rounds of off-rate screening. 10 sequences were isolated from the boxed cells and sequenced. The variant v186\_Fr.1 was predominantly enriched. (B) SEC and SPR measurement of v186\_Fr.1 against WT mSA streptavidin.



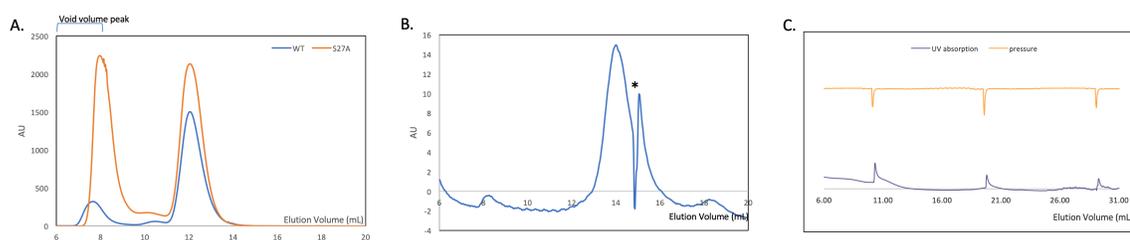
**Figure 3.7 Summary of the design results for mSA-targeting CDRxAbs and the overall design workflow:** (A-B) Summary of SPR-measured binding affinity and kinetics of nanobody-biotin conjugates and controls against mSA<sub>WT</sub> and mSA<sub>S27A</sub>. Experimental designs for each SPR binding experiment are depicted by the cartoon above the lanes. Blue square: SPR cheap for immobilization. Spheres: molecules that are immobilized (attached to chip) or flew through (floating above the chip) during binding experiments. Different molecules are represented by different colors. Individual data points represent measurements from biological triplicates. Error bars represent standard deviation. (C) Summary and proposal of a general design workflow for synergistically-binding nanobody-small molecule conjugates.



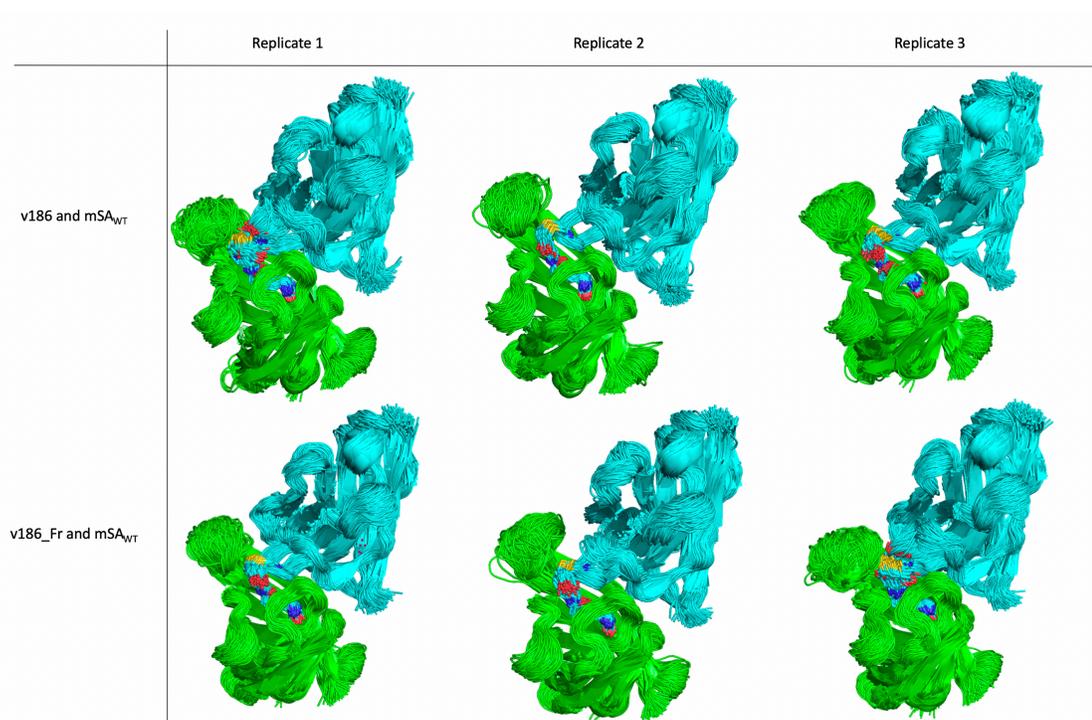
**Figure S1 Biotin conjugation was performed by biotin C2 maleimide with mutated cysteine residues.**



**Figure S2 An initial MD simulation that inspired us to move towards framework re-design:** 5 snapshots from a 100 ns MD simulation of v186/mSA complex. Carton: nanobody. Ribbon: mSA. Blue: CDR3 loop. Snapshots from earlier time points are shown with dimmer saturation.



**Figure S3 Additional Supporting SEC traces:** (A) SEC traces of mSA<sub>WT</sub> and mSA<sub>S27A</sub>. (B) SEC rerun trace of collected monomeric fraction for 4NBX.B-biotin103 v186\_Fr. \* indicates peaks of sample-irrelevant instrument defect of the overall FPLC, please refer to section C for more details. (C) Blank run of the FPLC to reveal the sample-irrelevant periodic peaks that were constantly observed in SEC data.



**Figure S4 Summary of MD simulations performed for 4NBX.B-biotin103 v186 and v186\_Fr against mSA<sub>WT</sub>:** For each simulation, 400 snapshots evenly spaced along the 100ns timescale are aligned together. Green: mSA<sub>WT</sub>. Cyan: nanobody-biotin conjugates. Biotin103 residue is shown as sticks.

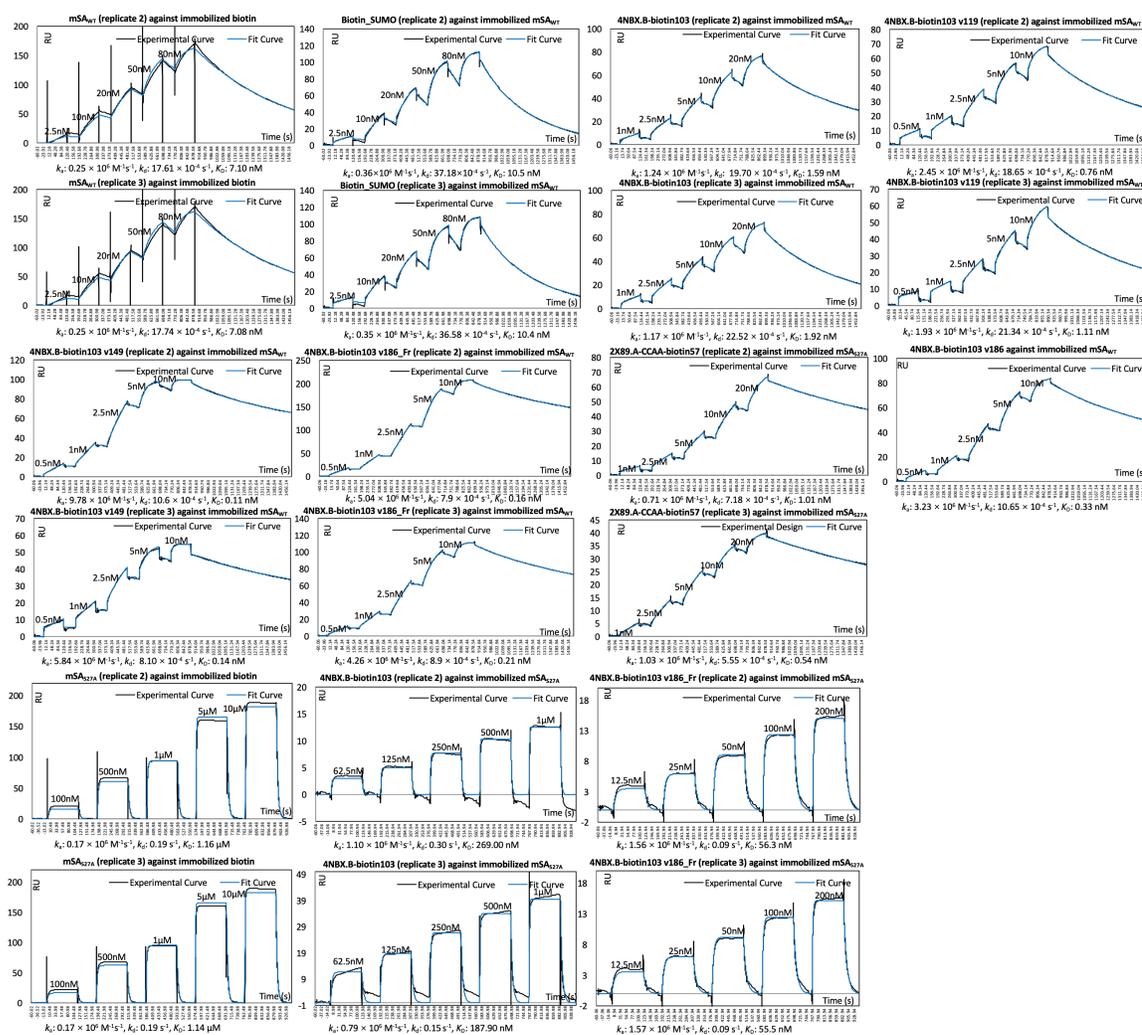
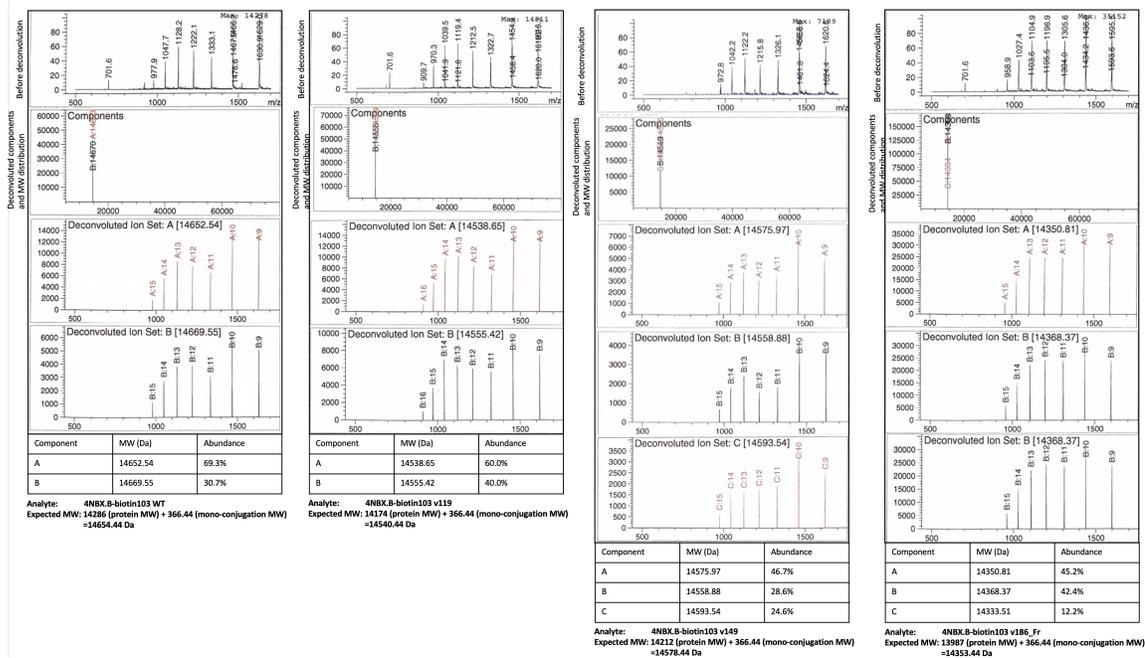
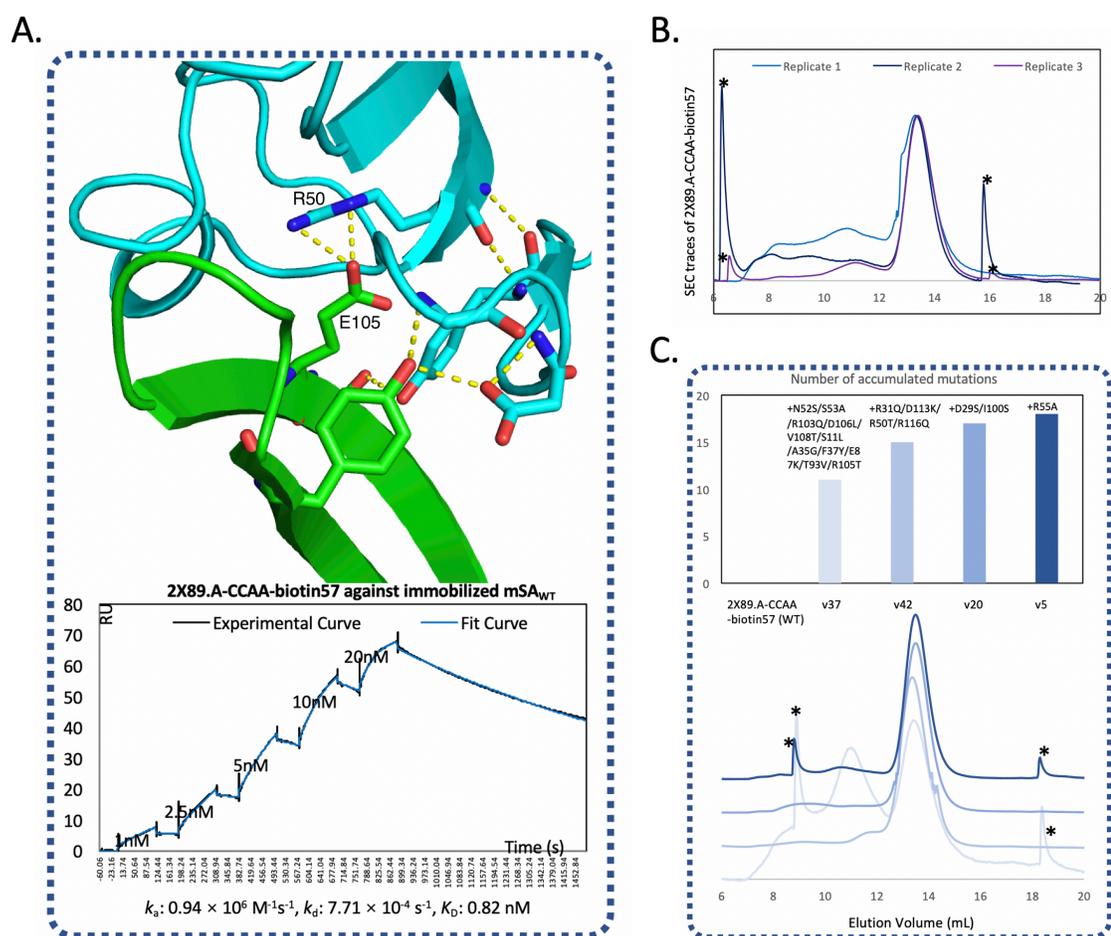


Figure S5 SPR measurements from the intermediate design variant 4NBX.B-biotin103 v186, and from additional biological replicates not shown in the main text but were included for affinity and kinetics estimation.



**Figure S6 Intact-protein mass spectrometry (MS) confirmed mono-conjugated materials: MS deconvolution of nanobody-biotin conjugates only returned MWs within 20 Da from expected MW of mono-conjugated materials.**



**Figure S7 Designing and testing of a nanobody scaffold obtained by directly docking against mSA:** (A) Predicted H-bond formation profile and SPR binding curve for 2X89.A-CCAA-biotin57 WT against immobilized mSA<sub>WT</sub>. Data from one of the triplicates is shown here, and data from the other two replicates is in Fig. S5. (B) Size-exclusion chromatography (SEC) traces of biological triplicates for 2X89.A-CCAA-biotin57, normalized by monomer peak height for better comparison of aggregates formation. \* indicates peaks of sample-irrelevant instrument defect of the overall FPLC, please refer to figure S3C for more details. (C) SEC traces and newly accumulated mutations for sequence-designed variants of 2X89.A-CCAA-biotin57 conjugates. Peaks of the monomeric fractions are normalized to identical heights. \* indicates peaks of sample-irrelevant instrument defect of the overall FPLC, please refer to figure S5C for more details.

**Table S1: amino acid sequences for the protein templates used in this study**

4NBX.B_ A103C	QVQLQESGGGLAQAGGSLRLSCAASGRFTFSMDPMAWFRQPPGKREFV AAGSSTGRITTYADSVKGRFTISRDNKNTVYVYLMNSLKPEDTAVYYC AAAPYGCNWRDEYAYWGQGTQVTVSSHHHHHH
2X89.A_C CAA_157C	QVQLQESGGGSVQAGGSLRLSCAASGYTDSRYAMAWFRQAPGKEREW VARINSGRDCTYYADSVKGRFTFSQDNKNTVYVYLMDSLEPEDTATYY CATDIPLRARDIVAKGGDGFYWGQGTQVTVSSHHHHHH
mSA_WT	HHHHHHSQDLASAEAGITGTWYNQSGSTFTVTAGADGNLTGQYENRAQ GTGCQNSPYTLTGRYNGTKLEWRVEWNNSTENCHSRTEWRGQYQGGA EARINTQWNLTYEKGGSPATEQGQDTFTKVKPSAASGSDYKDDDDK
Smt3 SUMO with N- terminal cysteine	CLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME AFAKRQKEMDSLRFYDGIQADQAPEDLDMEDNDIIEAHREQIGGH HHHHH

Note: the numbering scheme of the above sequences follows 1-2-3-4... The first residue is numbered as 1.

### *Bibliography*

1. D. C. Swinney, Biochemical mechanisms of drug action: what does it take for success? *Nature Reviews Drug Discovery* **3**, 801–808 (2004).
2. J. K. H. Liu, The history of monoclonal antibody development – Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery* **3**, 113–116 (2014).
3. A. L. Hopkins, C. R. Groom, The druggable genome. *Nature Reviews Drug Discovery* **1**, 727–730 (2002).
4. G. L. Verdine, L. D. Walensky, The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. *Clin Cancer Res* **13**, 7264–7270 (2007).
5. M. A. Ayoub, *et al.*, Antibodies targeting G protein-coupled receptors: Recent advances and therapeutic challenges. *MAbs* **9**, 735–741 (2017).
6. R. B. Dodd, T. Wilkinson, D. J. Schofield, Therapeutic Monoclonal Antibodies to Complex Membrane Protein Targets: Antigen Generation and Antibody Discovery Strategies. *BioDrugs* **32**, 339–355 (2018).
7. E. Valeur, *et al.*, New Modalities for Challenging Targets in Drug Discovery. *Angewandte Chemie International Edition* **56**, 10294–10323 (2017).
8. G. D. L. Phillips, *et al.*, Targeting HER2-Positive Breast Cancer with Trastuzumab-DM1, an Antibody–Cytotoxic Drug Conjugate. *Cancer Res* **68**, 9280–9290 (2008).
9. Y. Feng, *et al.*, Conjugates of Small Molecule Drugs with Antibodies and Other Proteins. *Biomedicines* **2**, 1–13 (2014).
10. W. D. Hedrich, T. E. Fandy, H. M. Ashour, H. Wang, H. E. Hassan, Antibody–Drug Conjugates: Pharmacokinetic/Pharmacodynamic Modeling, Preclinical Characterization, Clinical Studies, and Lessons Learned. *Clinical Pharmacokinetics* **57**, 687–703 (2017).
11. D. Su, *et al.*, Modulating Antibody–Drug Conjugate Payload Metabolism by Conjugation Site and Linker Modification. *Bioconjugate Chemistry* **29**, 1155–1167 (2018).
12. B.-Q. Shen, *et al.*, Conjugation site modulates the in vivo stability and therapeutic activity of antibody–drug conjugates. *Nature Biotechnology* **30**, 184–189 (2012).
13. Y. Wang, *et al.*, Peptide–Drug Conjugates as Effective Prodrug Strategies for Targeted Delivery. *Adv Drug Deliv Rev* **110–111**, 112–126 (2017).
14. R. Y. Zhao, *et al.*, Synthesis and Evaluation of Hydrophilic Linkers for Antibody–Maytansinoid Conjugates. *J. Med. Chem.* **54**, 3606–3623 (2011).
15. A. C. Cheng, *et al.*, Structure-guided Discovery of Dual-recognition Chemibodies. *Scientific Reports* **8**, 7570 (2018).
16. J. Tang, *et al.*, An Inhibitory Antibody against Dipeptidyl Peptidase IV Improves Glucose Tolerance in Vivo. *J. Biol. Chem.* **288**, 1307–1316 (2013).
17. S. Muyldermans, Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry* **82**, 775–797 (2013).
18. I. Hmila, *et al.*, VHH, bivalent domains and chimeric Heavy chain-only antibodies with high neutralizing efficacy for scorpion toxin AahI'. *Molecular Immunology* **45**, 3847–3856 (2008).

19. C. I. Webster, *et al.*, Brain penetration, target engagement, and disposition of the blood–brain barrier-crossing bispecific antibody antagonist of metabotropic glutamate receptor type 1. *The FASEB Journal* **30**, 1927–1940 (2016).
20. L. R. Pepper, Y. K. Cho, E. T. Boder, E. V. Shusta, A decade of yeast surface display technology: Where are we now? *Comb Chem High Throughput Screen* **11**, 127–134 (2008).
21. G. M. Cherf, J. R. Cochran, Applications of Yeast Surface Display for Protein Engineering. **1319**, 155–175 (2015).
22. Y. Sun, B. Ban, A. Bradbury, G. A. S. Ansari, D. A. Blake, Combining Yeast Display and Competitive FACS to Select Rare Hapten-Specific Clones from Recombinant Antibody Libraries. *Anal. Chem.* **88**, 9181–9189 (2016).
23. W. Zou, M. Ueda, A. Tanaka, Genetically Controlled Self-Aggregation of Cell-Surface-Engineered Yeast Responding to Glucose Concentration. *Appl Environ Microbiol* **67**, 2083–2087 (2001).
24. E. B. Getz, M. Xiao, T. Chakrabarty, R. Cooke, P. R. Selvin, A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* **273**, 73–80 (1999).
25. S. M. Marino, *et al.*, Characterization of surface-exposed reactive cysteine residues in *Saccharomyces cerevisiae*. *Biochemistry* **49**, 7709–7721 (2010).
26. Y.-M. Go, J. D. Chandler, D. P. Jones, The Cysteine Proteome. *Free Radic Biol Med* **84**, 227–245 (2015).
27. M. H. Qureshi, J. C. Yeung, S.-C. Wu, S.-L. Wong, Development and Characterization of a Series of Soluble Tetrameric and Monomeric Streptavidin Muteins with Differential Biotin Binding Affinities. *J. Biol. Chem.* **276**, 46422–46428 (2001).
28. K. H. Lim, H. Huang, A. Pralle, S. Park, Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnology and Bioengineering* **110**, 57–67 (2013).
29. G. Nimrod, *et al.*, Computational Design of Epitope-Specific Functional Antibodies. *Cell Reports* **25**, 2121–2131.e5 (2018).
30. N. Jain, S. W. Smith, S. Ghone, B. Tomczuk, Current ADC Linker Chemistry. *Pharm Res* **32**, 3526–3540 (2015).
31. T. Murase, *et al.*, Structural basis for antibody recognition in the receptor-binding domains of toxins A and B from *Clostridium difficile*. *J. Biol. Chem.* **289**, 2331–2343 (2014).
32. W. Sheng, X. Liao, Solution structure of a yeast ubiquitin-like protein Smt3: The role of structurally less defined sequences in protein–protein recognitions. *Protein Sci* **11**, 1482–1491 (2002).
33. F. Liu, J. Z. H. Zhang, Y. Mei, The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Sci Rep* **6** (2016).
34. C. McMahon, *et al.*, Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nature Structural & Molecular Biology* **25**, 289–296 (2018).
35. L. S. Mitchell, L. J. Colwell, Analysis of nanobody paratopes reveals greater diversity than classical antibodies. *Protein Eng Des Sel* **31**, 267–275 (2018).

36. H. K. Privett, *et al.*, Iterative approach to computational enzyme design. *PNAS* **109**, 3790–3795 (2012).
37. A. Nisthal, C. Y. Wang, M. L. Ary, S. L. Mayo, Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *PNAS* **116**, 16367–16377 (2019).
38. L. N. Makley, J. E. Gestwicki, Expanding the Number of ‘Druggable’ Targets: Non-Enzymes and Protein-Protein Interactions. *Chemical Biology & Drug Design* **81**, 22–32 (2013).
39. D. H. Nam, C. Rodriguez, A. G. Remacle, A. Y. Strongin, X. Ge, Active-site MMP-selective antibody inhibitors discovered from convex paratope synthetic libraries. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14970–14975 (2016).
40. D. Baran, *et al.*, Principles for computational design of binding antibodies. *Proceedings of the National Academy of Sciences*, 201707171 (2017).
41. C. A. Voigt, C. Martinez, Z.-G. Wang, S. L. Mayo, F. H. Arnold, Protein building blocks preserved by recombination. *Nature Structural Biology* **9**, 553–558 (2002).
42. S. E. Boyken, *et al.*, De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
43. B. D. Allen, A. Nisthal, S. L. Mayo, Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *PNAS* **107**, 19838–19843 (2010).
44. D. J. Mandell, E. A. Coutsiaris, T. Kortemme, Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods* **6**, 551–552 (2009).
45. D. G. Gibson, *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
46. D. Demeestere, *et al.*, Development and Validation of a Small Single-domain Antibody That Effectively Inhibits Matrix Metalloproteinase 8. *Molecular Therapy* **24**, 890–902 (2016).
47. K. Hand, M. C. Wilkinson, J. Madine, Isolation and purification of recombinant immunoglobulin light chain variable domains from the periplasmic space of *Escherichia coli*. *PLOS ONE* **13**, e0206167 (2018).
48. S. B. Hansen, N. S. Laursen, G. R. Andersen, K. R. Andersen, Introducing site-specific cysteines into nanobodies for mercury labelling allows de novo phasing of their crystal structures. *Acta Crystallographica Section D Structural Biology* **73**, 804–813 (2017).
49. T. Pleiner, *et al.*, Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *eLife* **4**, e11349 (2015)
50. M. L. B. Magalhães, *et al.*, Evolved streptavidin mutants reveal key role of loop residue in high-affinity binding. *Protein Sci* **20**, 1145–1154 (2011).
51. S. Doerr, M. J. Harvey, F. Noé, G. De Fabritiis, HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **12**, 1845–1852 (2016).
52. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
53. J. A. Maier, *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

54. J. Wang, W. Wang, P. A. Kollman, D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **25**, 247–260 (2006).
55. Gaussian 09, Revision A.02, M. J. Frisch, *et al.*, Gaussian, Inc., Wallingford CT, 2016.
56. R. T. McGibbon, *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528–1532 (2015).
57. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

*Appendix I***COMPUTATIONAL DESIGN AND EXPERIMENTAL SCREENING OF  
NEUTRALIZING ANTIBODY VARIANTS AGAINST SARS-COV-2**

Contribution: Jingzhou Wang performed combinatorial designs on C002 and C118 against the RBD of both WT and B.1.351 strains of SARS-CoV-2. Jingzhou Wang performed FACS sorting of a MACS-enriched degenerate codon library of C002 against WT RBD, and isolated 9 variants for further characterization.

## ***Introduction***

COVID-19 is an ongoing global pandemic that is caused by the coronavirus SARS-CoV-2 [1,2]. The virus enters human cells by its trimeric spike protein that protrudes from the capsid shell [2,3]. The receptor-binding domain (RBD) of the spike adopts two distinctive conformations [2,3]. Using the “up” conformation, the SARS-CoV-2 RBD interacts with the ACE2 receptor on host cells to start a cascade of biochemical processes that lead to viral entry [2-4]. Upon infection, the immune system produces IgG antibodies to recognize the virus outer protein surface at different epitopes, and a fraction of antibodies that bind to the RBD are able to neutralize the virus by inhibiting RBD-ACE2 engagement [5]. Structural characterization of patient-derived neutralizing antibodies (nAbs) classified four types of neutralization mechanisms [6]. Class I antibodies recognizes the “up” conformation of RBD and directly compete with ACE2 for RBD binding, class II antibodies compete with ACE2 while recognizing multiple RBD conformations, and class III and IV antibodies neutralize the virus by recognizing epitopes that do not overlap with ACE2 binding site [6]. Passive immunization therapies using especially potent nAbs were shown to effectively reduce the hospitalization rates for COVID-19 patients in high-risk groups [7]. Several antibody treatments against SARS-CoV-2 were approved by FDA, including the monoclonal antibody bamlanivimab from Eli Lilly and the antibody cocktail REGEN-COV from Regeneron. Despite the demonstrated benefits, several drawbacks prevent antibody treatments against COVID-19 from broad usage. First, current passive immunization treatments usually require a high dosage around 1 gram per antibody [8,9], which results in a high cost that is incompatible with the current rapid infection rate across the globe. Second, no current antibody treatment has shown effectiveness in hospitalized patients, and therefore demonstrates a narrow therapeutic window with limited usage scenario [8,9]. Third, escape mutations on RBD have been reported to dramatically reduce neutralization efficacy of certain patient-derived nAbs, jeopardizing the long-term effectiveness of existing antibody treatments [10].

We hypothesize that a further improvement in neutralization potency of existing nAbs might lead to more effective passive immunization treatments that alleviate the above-mentioned problems. We chose two previously reported antibodies, which are respectively a highly potent class II antibody C002 [6] that is comparable in potency to approved antibody treatments, and a less potent class IV antibody C118 [11] that recognizes an evolutionarily conserved RBD epitope. Two combinatorial designs were performed respectively on C002 and C118 against WT RBD to identify mutation choices that potentially improve RBD engagement and preserve the existing binding mode. For C002, escape mutation E484K was reported to disable RBD binding and occurs in a reported variant strain B.1.351 [6,10]. Therefore, a second C002 combinatorial design was performed against the B.1.351 RBD. Together, the three design trials output three sets of suggested mutations, which can be used for antibody library construction. Furthermore, the reported mutations for C002 design against WT RBD were combined with other calculation results to finalize a 3-million-member degenerate codon library, from which a variant with 6-fold improvement in potency was isolated.

## ***Results and Discussion***

### Combinatorial design results of C002 against WT RBD

C002/RBD(WT) structure was kindly provided by Christopher O. Barnes. The complex structure was first processed by standardizing the coordinates in Rosetta force field [12] and filling in missing RBD electron density using a complete RBD structure from PDB ID 7BZ5. Interface residues were identified by a publicly available PyMOL script that uses change of solvent accessibility as the assessment criteria [16]. Site-saturation design using the full 20 amino acids was performed on each CDR residue in the binding interface, and the sequence ranks from 10 repeated calculations with different random seeds were averaged to report single mutations with improved overall Rosetta energy score than the WT amino acid (Figure 1). Single mutations were also ranked by improvements in interaction energy score and shape complementarity (Figure 1). The beneficial mutations from the three lists were then combined with different biases to perform 7 combinatorial

designs on the heavy chain and three combinatorial designs on the light chain, all in parallel (Figure 2). Each design calculation was repeated for 10 times with different random seeds, and the sequences sampled in each repeated trial were combined and re-ranked to report the final output sequences, ranked by overall Rosetta energy score.

For each combinatorial design, the top 20 sequences and the corresponding structural models were used to assess design quality by overall energy score, interaction energy, and shape complementarity (Figure 2). Of the 7 designs on the heavy chain (HC), the “all” and “inter\_energy” design calculations showed consistent energy and shape complementarity improvements in the top 20 sequences, and are therefore chosen to construct the final recommended sets of mutations.

For the “all” design calculation, the mutation preference in the top 20 sequences are as follows (times of occurrence are reported in the brackets):

<b>Residue Position (HC)</b>	<b>Sequence Choices</b>
28	WT(0),D(10),Q(10)
32	WT(13),F(7)
33	WT(0), A(1),P(19)
57	WT(11),H(9)
99	WT(0), G(20)
101	WT(0), G(18),S(2)
102	WT(0), D(1),G(17),K(1),N(1)
103	WT(5),D(13),N(2)
104	WT(5),E(15)
106	WT(0), G(6),N(13),S(1)
107	WT(0), F(2),H(1),M(1),N(2),Q(3),R(10),S(1)
109	WT(0), D(2),N(18)

For the “inter\_energy” design calculation, the mutation preference in the top 20 sequences are as follows (times of occurrence are reported in the brackets):

<b>Residue Position (HC)</b>	<b>Sequence Choices</b>
28	WT(1),D(11),Q(8)
32	WT(12),F(8)
33	WT(1),A(6),P(10),S(3)
57	WT(13),H(7)
99	WT(1),G(19)
101	WT(0),G(20)
102	WT(0),S(20)
103	WT(14),N(6)
106	WT(0),G(5),N(15)
107	WT(18),G(2)
109	WT(0),A(1),D(3),G(2),H(8),M(1),N(5)

The final set of HC mutations were recommended by combining the above two lists of sequence choices and forcing WT amino acid choice in each position:

<b>Residue Position (HC)</b>	<b>Sequence Choices</b>
28	T,D,Q
32	Y,F
33	G,A,P,S
57	N,H
99	E,G
101	R,G,S
102	P,D,G,K,N,S
103	S,D,N
104	D,E
106	V,G,N,S
107	V,F,H,M,N,Q,R,S,G
109	V,A,D,G,H,M,N

No obvious improvements were seen in combinatorial designs of the light chain (LC) residues. However, two of the designs “L\_SC” and “L\_int\_skip3292” were chosen for consistently improved interaction energy and shape complementarity.

For the “L\_SC” design calculation, the mutation preference in the top 20 sequences are as follows (times of occurrence are reported in the brackets):

<b>Residue Position (LC)</b>	<b>Sequence Choices</b>
30	WT(10),G(10)
93	WT(5),D(1),G(1),K(1),Q(1),T(11)
94	WT(6),A(2),F(1),H(1),I(2),N(1),S(3),W(3),Y(1)
96	WT(8),F(12)

For the “L\_int\_skip3292” design calculation, the mutation preference in the top 20 sequences are as follows (times of occurrence are reported in the brackets):

<b>Residue Position (LC)</b>	<b>Sequence Choices</b>
30	WT(9),G(10),N(1)
93	WT(4),A(4),D(1),G(2),M(1),N(2),T(6)
94	WT(3),E(2),F(2),H(1),I(3),L(2),M(3),N(1),R(2),W(1)
96	WT(6),F(11),S(3)

The final set of LC mutations were recommended by combining the above two lists of sequence choices, also forcing WT amino acid choice in each position.

<b>Residue Position (LC)</b>	<b>Sequence Choices</b>
30	S,G,N
93	S,A,D,G,M,N,K,Q,T
94	T,A,E,F,H,I,L,M,N,S,R,W,Y
96	R,F,S

### Combinatorial design results of C002 against B.1.351 RBD

Structure of C002/RBD(WT) from PDB ID 7K8S was used, and the B.1.351 mutations E484K and K417N that are close to C002 binding sites were modeled in after standardizing the structure in Rosetta force field [10,12]. C002 residues that are within 6 angstroms away from residue 484 were considered for redesign. Since no obvious interactions were observed between C002 and residue 417 in the published structure, designs were not considered for K417N.

Site-saturation design and analysis were performed on the selected residues in the same manner as the previous C002 design against WT RBD, resulting in three lists of ranked mutation choices for each CDR site of interest (Figure 3). Combinatorial designs were performed using the identified mutation sites and choices, in the same way as the design against WT RBD. Analysis of top 20 sequences from each design calculation indicated various degree of improvements in total energy and interaction energy, but all of the designs resulted in reduction in shape complementarity (Figure 4). Nevertheless, since the starting sets of acceptable mutation choices were small, as only three sites showed mutation choices with favorable overall energy score, all of the four design calculations were combined to recommend the final set of mutation choices. WT amino acid is forced in each position to reduce the chance of non-functional sequences:

<b>Residue Position</b>	<b>Sequence Choices</b>
HC/52A	Y,E,W
HC/56	N,Y
LC/96	R,D,E,F,H,I,M,S,T,V,A

The consistently worse shape complementarity across the designs is discouraging. This result suggests that instead of designing with a single conformational state, a more comprehensive exploration of backbone and sidechain mobility may find better solutions to counter the E484K penalty. However, combining mutations reported by this design effort with favorable mutations identified by other avenues could also be advantageous.

### Combinatorial design results of C118 against WT RBD

C118-RBD(SARS1) structure was kindly provided by Claudia A. Jette. The structure was standardized in the same way as C002, and mutations were introduced to recapitulate sequence differences between SARS1 and SARS2 RBDs that are close to the antibody binding site. C118 interface residues were identified in the same way as C002 against WT RBD. Site-saturation and combinatorial design and analysis were performed in the same way as C002 (Figure 5&6). All of the four combinatorial designs performed in parallel showed consistent improvement in the three evaluation criteria: overall energy, interaction energy, and shape complementarity (Figure 6). Combining the results of the best three design calculations returned the final sequence recommendation, where the WT amino acid choice was forced in each position, like the other design trials:

<b>Residue Position</b>	<b>Sequence Choices</b>
HC/28	T,D,E,N,Q
HC/31	N,H,L,Q,Y,E,F
HC/53	Y,W,F
HC/56	S,T
HC/57	N,H
HC/101	T,I
HC/108	R,G,N
HC/110	D,E,Q,A
LC/51	N,F,Y
LC/56	H,F,Y
LC/57	S,E,L

### FACS sorting of a degenerate codon library that incorporates the C002 combinatorial design results against WT RBD

The recommended mutations of C002 HC against WT SARS-CoV-2 RBD were combined with results of a single/double site(s) mutation search and a PSSM analysis, which were respectively performed by our collaborators Paul Chang and Stephanie Contreras, to mature into the final explicit sequence library:

<b>Residue Position (HC)</b>	<b>Sequence Choices</b>
28	T,D,H,I,P,V
32	Y,F
33	G,S
57	N,H,K,Y
99	E,G,H,S,Y
101	R,G,H,S,Y
102	P,D,G,N,S
103	S,D,N
106	V,S,W,Y
107	V,F,G,H,M,Q,R,S
109	V,A,F,M,W,Y

Our collaborators Paul Chang and Justin Chartron further represented this library into degenerate codons:

<b>Residue Position (HC)</b>	<b>Codon Choice</b>
28	VHC
32	TWC
33	GGA
57	MAC
99	GRA
101	BRC
102	VSC
103	RRC
106	KKG
107	SDC
109	DKS

Justin Chartron then constructed the final degenerate codon library into a yeast display vector as scFv, and performed one round of magnetic-activated cell sorting (MACS) with 50 nM WT RBD.

The MACS-enriched sub-library was then sorted by fluorescence-activated cell sorting (FACS) using WT RBD at 2 nM, which is a ~6-fold lower concentration than the published  $K_D$  of WT C002 (Figure 7A) [6]. Cells with positive expression and RBD signals were collected, re-grown, and sequenced. 9 variants were isolated. Out of the 12 mutations observed in the isolated variants, 10 mutations were in the final explicit sequence library, 1 is introduced by degenerate codon but not recommended by design, and 1 is introduced opportunistically during library construction. Importantly, 8 out of the 12 mutations were predicted by the combinatorial design calculation alone. Together, this result demonstrates the robustness of our sequence design capability on existing antibody/target complexes.

The 9 variants were reformatted into IgGs, which were expressed by our collaborators Annie Lam and Jost Vielmetter together with the WT C002 IgG as control. Finally, our collaborators Priyanthi Gnanapragasam and Jennifer Keeffe tested and analyzed the neutralization potency of the 10 C002 IgG variants against a SARS-CoV-2 pseudovirus. One variant, v4 with mutations T28P/N57H/V107H showed ~6-fold  $IC_{50}$  improvement than WT C002 in three independent technical repeats, and is also roughly one magnitude more potent than the FDA-approved antibodies from Eli Lilly and Regeneron (Figure 7B) [5,13,14].

### ***Materials and Methods***

Computational design: Combinatorial sequence design calculations on C002 and C118 were performed on an in-house protein design suite TRIAD [15]. PyMOL (Schrodinger) was also used to assist structural analysis.

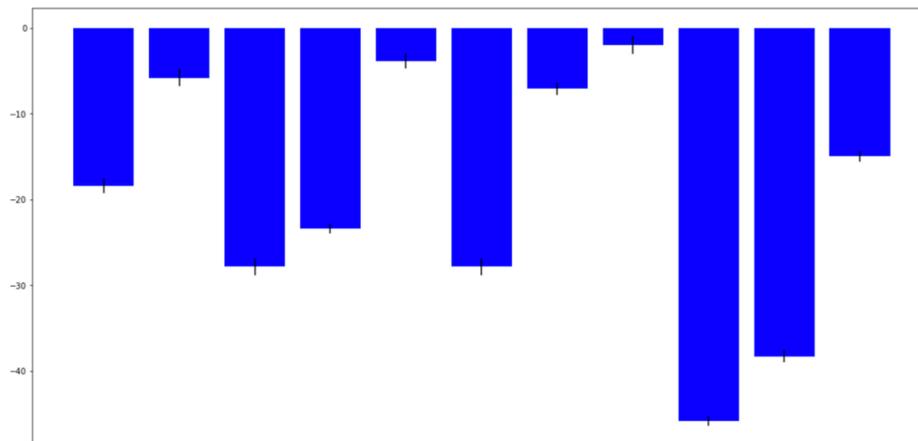
Yeast display and FACS sorting: MACS-enriched EBY100 yeast cells with display vector pCTCON2 containing C002 scFv variants were provided by Justin Chartron. Under 4°C,  $10^7$  cells were first washed by TBS (pH 7.4)/0.1% BSA and incubated with anti-c-myc antibodies (AF488) and 2 nM biotinylated SARS-CoV2 RBD(WT), which was provided

by Christopher O. Barnes. Incubated cells were then pelleted down, stained with streptavidin (AF647) for 5 minutes, and washed once by TBS/BSA buffer at 4°C before sorted by FACS. FACS was performed by a Sony SH800 cell sorter. Antibody expression was reflected by AF488, and RBD binding was reflected by AF647. Cells with double-positive fluorescence signal were isolated and regrown in SD-CAA media (Teknova). Isolated scFv sequences were then miniprepmed by commercially available kit (Zymo) and identified by sanger sequencing.

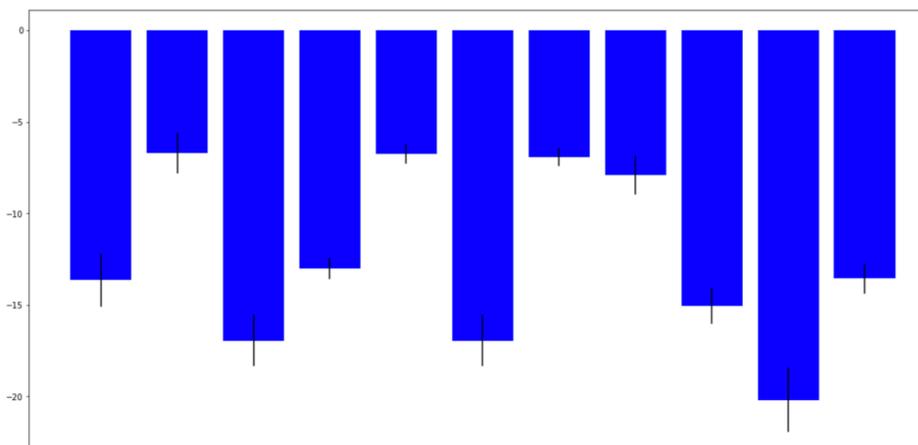
1. Mutation choices with improved overall energy score	2. Mutation choices with improved interaction energy	3. Mutation choices with improved shape complementarity
<pre> ["-aas H102 Ser,Gly,Arg,Cys,Lys,Asp,Glu,Leu,Gln,Met,His,Ala,Asn,Tyr,WT", "-aas H101 Gly,Ser,Asp,Ala,Thr,Gln,Glu,His,Cys,Asn,WT", "-aas H107 Arg,Lys,Gln,Met,His,Thr,Ser,Glu,Asn,Asp,Ala,Gly,Trp,Phe,Cys,Ile,WT", "-aas H109 Asn,Asp,Gly,Ala,Ser,His,Met,Glu,WT", "-aas H106 Gly,Ser,Ala,Asp,Lys,Asn,Met,Glu,WT", "-aas H33 Pro,Ala,Ser,WT", "-aas H99 Gly,WT", "-aas H28 Gln,Asp,WT", "-aas H103 Asn,Asp,WT", "-aas H32 Phe,WT", "-aas H104 Glu,WT", "-aas H57 His,WT", "-aas H26 WT", "-aas H27 WT", "-aas H20 WT", "-aas H11 WT", "-aas H32 WT", "-aas H32 WT", "-aas H50 WT", "-aas H52 WT", "-aas H53 WT", "-aas H54 WT", "-aas H59 WT", "-aas L91 WT", "-aas L92 WT", "-aas L93 WT", "-aas L94 WT", "-aas L96 WT", "-aas H109 WT", "-aas H56 WT"] </pre>	<pre> [H28: M,E,R,N,K,W,C,I,L,A,S,D,V,Q,H,P H107: G H109: A,G,S,M,H,Q,P,K,R,Y,W,E,D,T,N,C,F,I H26: A,S,N,K,D,P,L,Y,H,V,M,Q,R,W,C,T,I H108: A,D,Y,N,M,L,F,P H27: T,H,S,P,L,Q,K,R,W H101: T,A,Q,E,W,Y,C,N,G,M,F,V,P H106: N,Q,L,F,E,G,T,H,Y,W H33: S,N,D,T,Q,Y,M,R,L,W,E,C,A,V,P,K,I,H,F H56: P,A,F,D,H,G,Y,C H52: D,A,H,R,S,V,P,M,N,K,E,T,F,Q,I,G,C H59: I,H,T,K,P,S,W,N,F,E H50: N,K,F,Q,C,E,M,R,Y H57: D,A,C,G,E,P,M,Q,L,Y,R,T,H,I,V,F,W H103: F,T,E,W,A,G,Q,L H104: T,K,R,P,A,M,N,L,F H102: S H99: T,S,K,R,W,M,P,F,Q,L,G,D,I,Y H52: N H54: A,M,Y,T,E,K,C,P H53: A,W,N,R,P,I,Q,D,G,T] </pre>	<pre> [H28: M,E,R,N,K,W,C,I,L,A,S,D,V,H,Y H107: M,S,P,G,I,L,E,N,A,K,C H109: A,G,S,M,H,Q,P,K,R,Y,W,E,D,T,N,C,F,I H26: A,S,N,E,P,Y,H,Q,R,W,I H108: M,P H27: T,H,S,L,N,K H101: T,Q,M,P H106: A,N,M,I,L,F,E,S,R,G,T,Y H33: S,Y,M,R,E,H H56: N,E,R,M,P,A,F,W,V,L,H,G,Y,C H52: D,A,H,R,S,V,P,M,N,K,L,T,Q,I,G,C H59: K,S,W,M,F,L,A,D H50: S,T,G,D,N,K,H,Q,C,E,M,R,Y,W H57: D,A,C,G,E,P,M,Q,L,Y,R,T,H,I,V,F H103: M,Y,F,T,E,A,V,H,D,W,N,G,Q,L,I H104: T H102: L,Q,S H99: F,I,G,I H52: H H54: A,V,Y,V,G,N H53: A,W,N,P,Q,D,T H31: K,F,L,A,Y] </pre>

**Figure 1 Single-site saturation analysis results of C002 CDR sites against WT SARS-CoV2 RBD:** Mutations with predicted improvements in overall Rosetta energy score, interaction energy, and shape complementarity than WT sequence choices were reported for each position. “H” refers to the heavy chain, “L” refers to the light chain, and “-aas” refers to amino acid site positions.

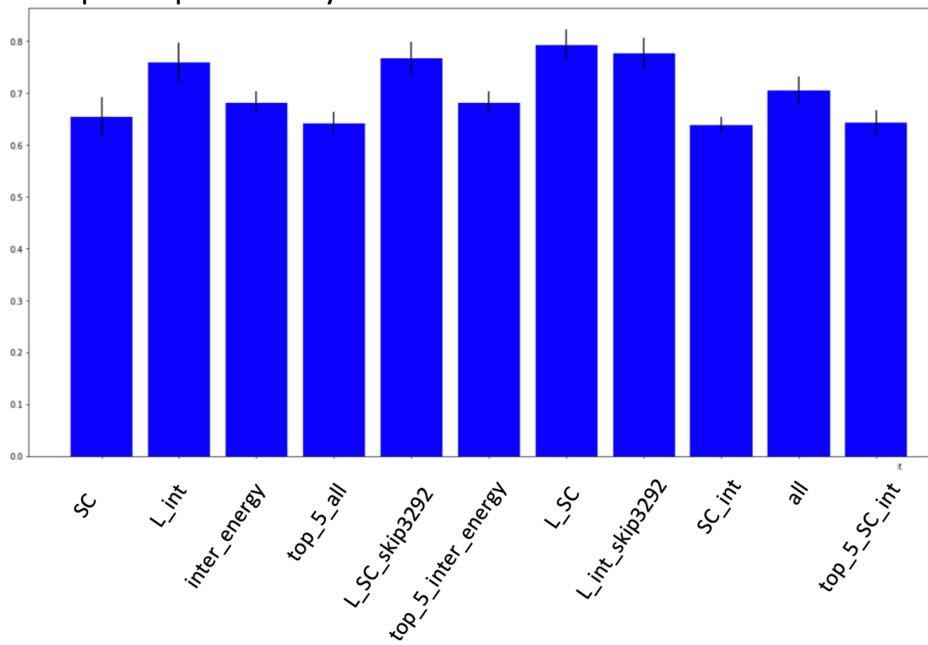
Overall energy score improvement  
(in  $\Delta$ Rosetta energy score compared to WT sequence)



Interaction energy (in Rosetta energy score)



Shape complementarity

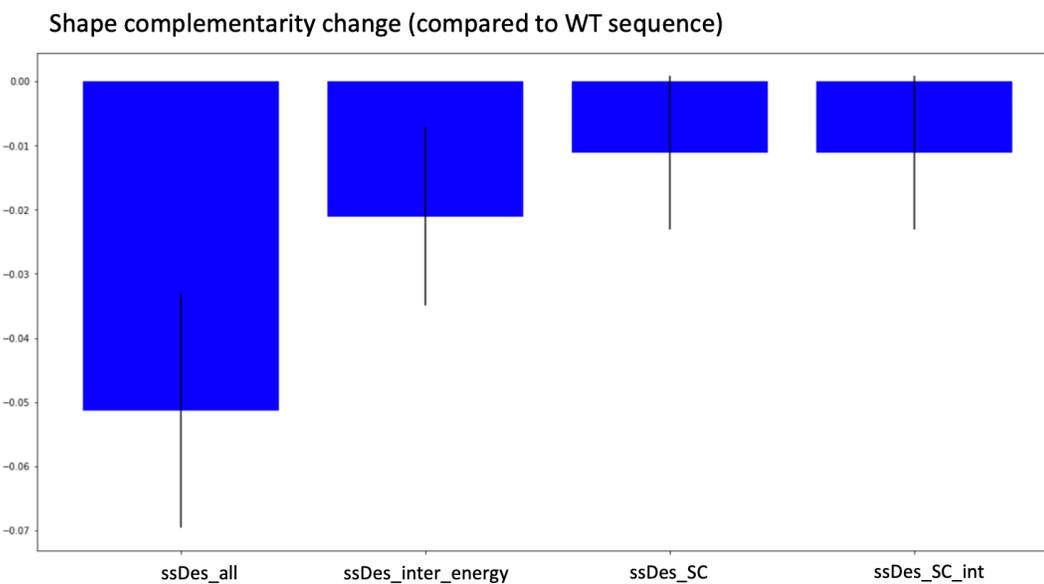
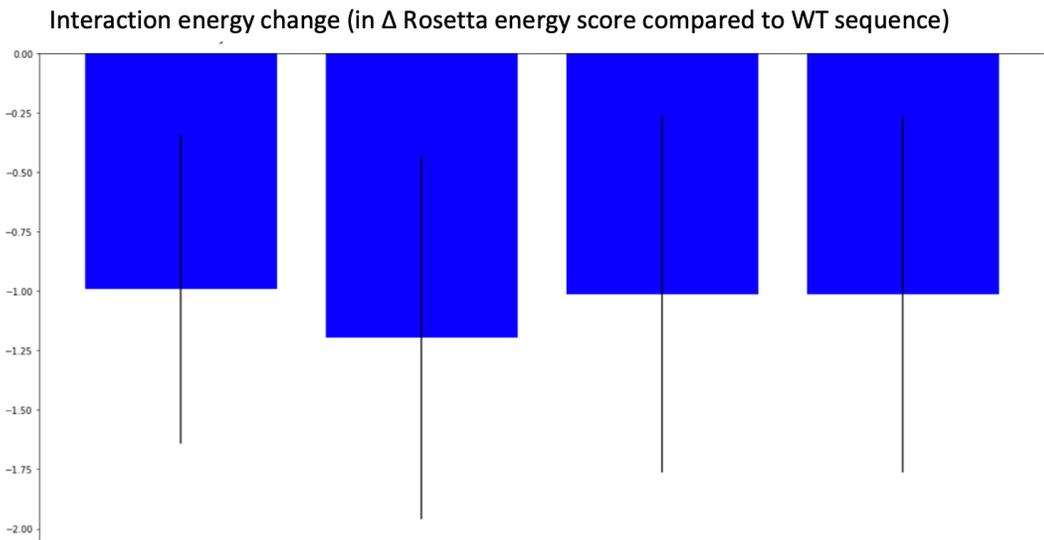
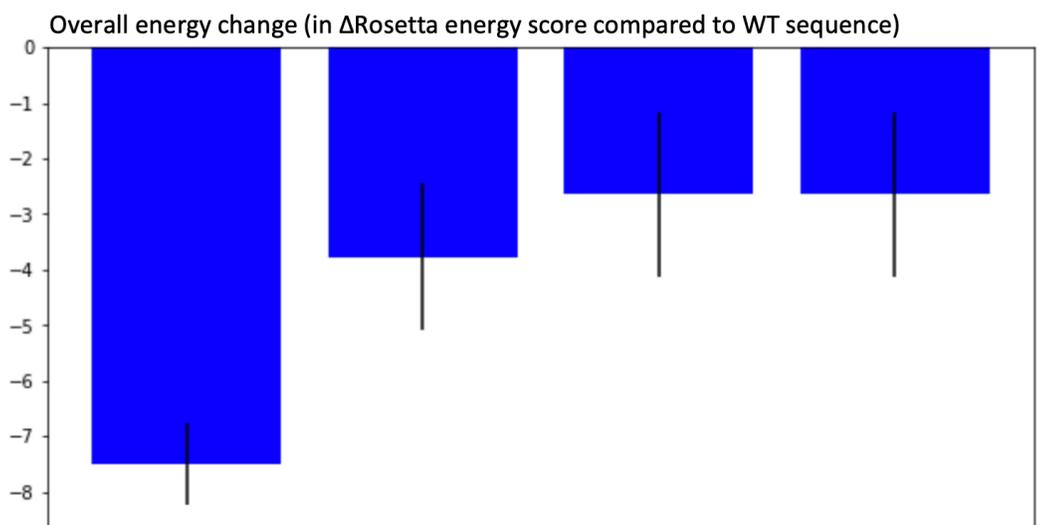


**Figure 2 Analysis of C002 combinatorial design results against WT SARS-CoV-2**

**RBD :** Top 20 sequences from 11 combinatorial sequence designs that were either restricted on HC or LC alone were analyzed by overall energy score improvement, interaction energy, and shape complementarity (SC). Data was shown as average value in the 20 sequences with standard deviations. For interaction energy and shape complementarity, analysis was performed only on the antibody chain where mutations were modeled in. The interaction energy between heavy chain and RBD was calculated to be -10.8 and the interaction energy between light chain and RBD was calculated to be -6.6. The SC between heavy chain and RBD was calculated to be 0.62, and the SC between light chain and RBD was calculated to be 0.74. The detailed description of each design is as follows. Heavy chain designs: all – All amino acid sites and choices from list 1 on figure 1, inter\_energy – Amino acid sites and choices observed in both list 1 and 2 on figure 1, SC – Amino acid sites and choices observed in both list 1 and 3 on figure 1, SC\_int – Amino acid sites and choices observed in all of list 1, 2, and 3, top\_5\_all – Top 5 amino acid sites according to the ranks on list 1, top\_5\_inter\_energy – Top 5 amino acid sites according to the ranks on list 1 and amino acid choices observed in both list 1 and 2, and top\_SC\_int – Top 5 amino acid sites according to the ranks on list 1 and amino acid choices observed in all of list 1, 2, and 3. Light chain designs: L\_int – Amino acid sites and choices observed in list 2, L\_SC – Amino acid sites and choices observed in list 3, L\_int\_skip3292 – L\_int design without sites 32 and 92, since they are a potential interaction hotspot.

1. Mutation choices with improved overall energy score	2. Mutation choices with improved interaction energy	3. Mutation choices with improved shape complementarity
<pre> [-aas P/96 His_Val,Ile,Thr,Asp,Ala,Ser,Met,Phe,Glu,WT', '-aas O/528 Trp,Glu,WT', '-aas O/56 Tyr,WT', '-aas O/35 WT', '-aas O/50 WT', '-aas O/95 WT', '-aas O/34 WT', '-aas O/33 WT', '-aas O/51 WT', '-aas P/93 WT', '-aas P/94 WT', '-aas P/92 WT', '-aas O/100E WT', '-aas O/52 WT', '-aas P/91 WT', '-aas O/100G WT', '-aas O/58 Ile,WT'] </pre>	<pre> [-aas O/100E His,Met,Trp,Tyr,Pro,Gly,Ser,Glu,Arg,Phe,Thr,Ile,Gln,Leu,Ala', '-aas O/100G His,Met,Trp,Asp,Tyr,Pro,Asn,Glu,Arg,Ile,Gln,Lys,Leu,Val', '-aas O/33 Ala', '-aas O/34 Asp,Tyr,Pro,Gly,Ser,Asn,Glu,Arg,Thr,Ile,Gln', '-aas O/35 Asp,Tyr,Pro,Gly,Glu,Arg,Phe,Thr,Ile,Leu,Val,Ala', '-aas O/30 His,Met,Trp,Asp,Tyr,Gly,Ser,Glu,Arg,Phe,Ile,Gln,Lys,Leu', '-aas O/51 His,Met,Asp,Tyr,Pro,Gly,Ser,Asn,Glu,Arg,Phe,Thr,Gln,Lys,Leu,Val,Ala', '-aas O/52 His,Met,Trp,Asp,Tyr,Pro,Gly,Glu,Arg,Phe,Thr,Ile,Leu,Val,Ala', '-aas O/28 Trp,Phe', '-aas O/56 His,Met,Trp,Arg,Phe,Ile,Lys,Leu,Val,Ala', '-aas O/58 His,Trp,Asp,Ser,Gln,Phe,Thr,Ile,Lys,Leu,Val,Ala', '-aas O/95', '-aas P/91 His,Met,Trp,Asp,Tyr,Pro,Gly,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas P/92', '-aas P/93 Trp,Gly,Arg,Gln,Lys', '-aas P/94 His,Met,Trp,Tyr,Phe', '-aas P/96 His,Met,Trp,Asp,Tyr,Pro,Gly,Ser,Asn,Glu,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala'] </pre>	<pre> [-aas O/100E Met,Trp,Asp,Ser,Asn,Phe,Thr,Gln,Lys,Leu', '-aas O/100G His,Asp,Glu,Arg', '-aas O/33 His,Met,Trp,Asp,Tyr,Pro,Asn,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas O/34 His,Trp,Asn,Phe,Thr,Ile,Lys,Val', '-aas O/35 Trp,Asp,Tyr,Ser,Asn,Glu,Arg,Phe,Lys,Val,Ala', '-aas O/50 Asp,Asn,Arg,Phe', '-aas O/51 His,Met,Trp,Tyr,Asn,Glu,Arg,Phe,Gln,Lys,Leu,Val', '-aas O/52 His,Trp,Tyr,Gly,Asn,Glu,Phe,Thr,Ile,Lys,Leu,Val', '-aas O/52A Met,Trp,Asp,Ser,Asn,Arg,Phe,Thr,Ile,Gln,Leu,Val,Ala', '-aas O/56 Trp,Pro,Gly,Arg,Gln,Lys,Leu,Ala', '-aas O/58 His,Met,Trp,Pro,Gly,Ser,Asn,Glu,Arg,Thr,Ile,Lys,Leu,Val,Ala', '-aas O/95 Tyr,Phe', '-aas P/91 His,Met,Trp,Asp,Pro,Gly,Asn,Glu,Arg,Phe,Thr,Gln,Lys,Leu,Ala', '-aas P/92 His,Trp,Asp,Arg,Thr,Lys', '-aas P/93 Met,Gly,Glu,Thr,Gln,Lys,Ala', '-aas P/94 His,Trp,Tyr,Gly,Ser,Asn,Phe,Gln,Lys,Leu', '-aas P/96 Pro,Gly,Ser,Thr,Ala'] </pre>

**Figure 3 Single-site saturation analysis results of C002 CDR sites against B.1.351 SARS-CoV2 RBD:** Mutations with predicted improvements in overall Rosetta energy score, interaction energy, and shape complementarity than WT sequence choices were reported for each position. “O” refers to the heavy chain, “P” refers to the light chain, and “–aas” refers to amino acid site positions.

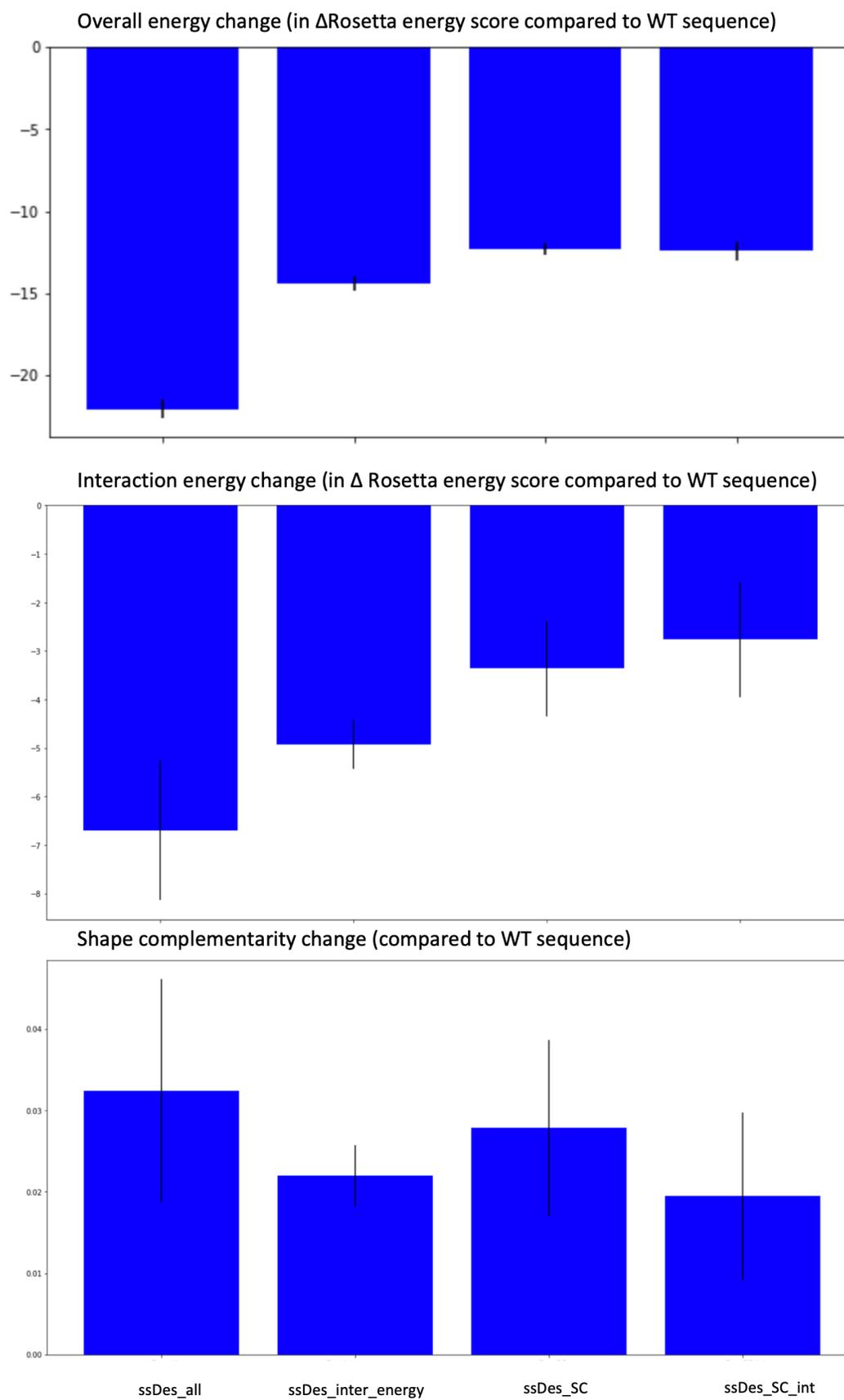


**Figure 4 Analysis of C002 combinatorial design results against B.1.351 SARS-CoV-**

**2 RBD Four combinatorial designs were performed as follows:** ssDes\_all – used all residue choices that showed up to improve the total energy (Figure 3 list 1). ssDes\_all\_inter\_energy – used residues that showed up to both improve the total energy and the interaction energy (residues that appear in both Figure 3 List 1 and List 2). ssDes\_SC – used residues that showed up to both improve the total energy and the shape complementarity score (residues that appear in both Figure 3 List 1 and List 3). ssDes\_SC\_int – used the residues that showed up to improve the total energy, interaction energy, and shape complementarity (residues that appear in all three lists. Top reported sequences were analyzed for overall energy, interaction energy, and shape complementarity change over WT sequence. Data is reported as average and standard deviations.

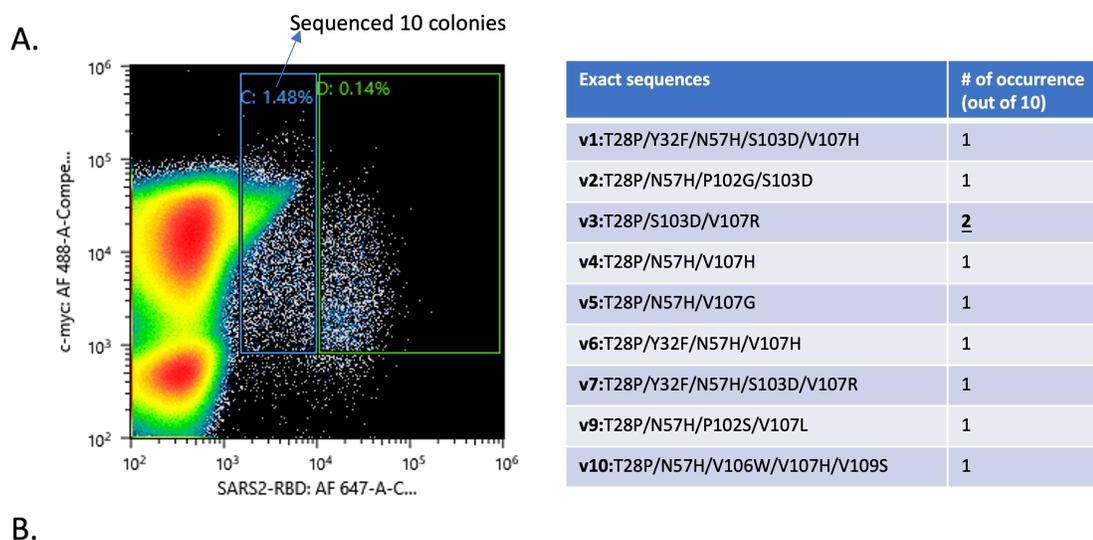
1. Mutation choices with improved overall energy score	2. Mutation choices with improved interaction energy	3. Mutation choices with improved shape complementarity
<pre> [-aas B/31 His,Phe,Leu,Tyr,Glu,Gln,WT', '-aas B/28 Asp,Asn,Gln,Glu,WT', '-aas B/10 Ala,Gln,Ser,Glu,WT', '-aas B/57 Leu,His,Glu,WT', '-aas B/56 Tyr,Phe,WT', '-aas B/53 Trp,Phe,WT', '-aas B/101 Ile,WT', '-aas B/51 Tyr,Phe,WT', '-aas B/108 Asn,Asn,Gly,WT', '-aas B/56 Thr,WT', '-aas B/57 His,WT', '-aas B/26 WT', '-aas B/27 WT', '-aas B/32 WT', '-aas B/46 WT', '-aas B/49 WT', '-aas B/53 WT', '-aas B/54 WT', '-aas B/55 WT', '-aas B/58 WT', '-aas B/59 WT', '-aas B/60 WT', '-aas B/61 WT', '-aas B/100 WT', '-aas B/102 WT', '-aas B/103 WT', '-aas B/105 WT', '-aas B/109 WT', '-aas B/111 WT']                     </pre>	<pre> ['-aas B/100', '-aas B/101 Met,Tyr,Pro,Asn,Glu,Arg,Ile,Lys,Leu,Val', '-aas B/102 His,Met,Trp,Asp,Tyr,Ser,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/103 Trp,Arg,Phe,Leu', '-aas B/105 Phe', '-aas B/108 His,Met,Trp,Asp,Tyr,Pro,Gly,Ser,Asn,Glu,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/109 Trp,Tyr,Ser,Arg,Thr,Gln', '-aas B/110 Trp,Pro,Glu,Gln,Lys,Leu', '-aas B/111 His,Met,Trp,Asp,Pro,Gly,Ser,Asn,Glu,Arg,Thr,Ile,Gln,Lys,Val,Ala', '-aas B/26', '-aas B/27 Ile', '-aas B/28 His,Met,Trp,Tyr,Ser,Glu,Arg,Phe,Ile,Lys,Leu,Val,Ala', '-aas B/31 His,Met,Trp,Asp,Tyr,Pro,Ser,Phe,Thr,Ile,Gln,Leu,Ala', '-aas B/32 His,Asp,Pro,Gly,Ser,Asn,Thr,Ile,Leu,Val,Ala', '-aas B/33 Trp,Arg,Phe', '-aas B/36 Met,Asp,Tyr,Pro,Gly,Asn,Glu,Arg,Phe,Thr,Gln,Lys,Leu,Val,Ala', '-aas B/37 Met,Asp,Pro,Glu', '-aas B/46 Lys', '-aas B/49 Glu,Arg', '-aas B/51 His,Met,Trp,Tyr,Gly,Ser,Glu,Arg,Phe,Thr,Ile,Gln,Leu,Val', '-aas B/52 Pro,Ser,Asn,Phe,Ile', '-aas B/54 His,Met,Trp,Tyr,Pro,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val', '-aas B/55 His,Met,Trp,Asp,Tyr,Pro,Asn,Arg,Phe,Ile,Val,Ala', '-aas B/56 Met,Trp,Tyr,Glu,Arg,Phe,Ile,Lys,Val,Ala', '-aas B/57 His,Met,Trp,Asp,Tyr,Pro,Gly,Glu,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/58 Phe', '-aas B/59 His,Met,Trp,Asp,Tyr,Ser,Asn,Glu,Arg,Thr,Ile,Gln,Leu,Val,Ala', '-aas B/60 His,Met,Trp,Tyr,Pro,Gly,Ser,Glu,Arg,Phe,Thr,Ile,Leu,Val,Ala', '-aas B/61 Asn,Glu,Gln,Lys,Leu']                     </pre>	<pre> ['-aas B/100 Met,Trp,Gln', '-aas B/101 His,Met,Tyr,Pro,Gly,Asn,Arg,Phe,Gln,Leu,Val,Ala', '-aas B/102 His,Met,Trp,Asp,Tyr,Pro,Ser,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/103 Lys,Leu', '-aas B/105 Phe', '-aas B/108 His,Asp,Pro,Gly,Ser,Asn,Glu,Thr,Ile,Lys,Val,Ala', '-aas B/109 Pro,Asn,Leu', '-aas B/110 Met,Trp,Pro,Gly,Thr,Gln,Val,Ala', '-aas B/111 His,Met,Trp,Asp,Pro,Gly,Ser,Asn,Glu,Arg,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/26 His,Met,Asp,Tyr,Pro,Ser,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/27 His,Met,Asp,Pro,Gly,Ser,Asn,Glu,Leu,Ala', '-aas B/28 His,Met,Trp,Asp,Tyr,Asn,Glu,Phe,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/31 His,Tyr,Gly,Glu,Phe,Ile,Gln,Lys,Leu,Val', '-aas B/32 His,Asp,Pro,Gly,Ser,Asn,Thr,Ile,Leu,Val,Ala', '-aas B/33 His,Trp,Asp,Pro,Gly,Ser,Asn,Arg,Thr,Ile,Gln,Leu,Val,Ala', '-aas B/34 His,Met,Asp,Pro,Gly,Asn,Glu,Thr,Ile,Lys,Val,Ala', '-aas B/37 Asp,Pro,Gly,Ser,Glu,Thr,Gln,Val,Ala', '-aas B/46 His,Met,Asp,Gly,Ser,Asn,Glu,Thr,Ile,Gln,Lys,Val,Ala', '-aas B/49 His,Trp,Tyr,Pro,Gly,Ser,Asn,Glu,Phe,Ile,Gln,Val', '-aas B/51 Trp,Asp,Ser,Arg,Phe,Thr', '-aas B/53 His,Met,Trp,Tyr,Pro,Gly,Ser,Glu,Arg,Phe,Thr,Gln,Lys,Leu,Val,Ala', '-aas B/54 His,Met,Trp,Asp,Tyr,Pro,Ser,Asn,Glu,Arg,Phe,Thr,Ile,Lys,Leu,Val,Ala', '-aas B/55 His,Met,Trp,Asp,Tyr,Pro,Gly,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/56 Met,Asp,Pro,Gly,Ser,Asn,Glu,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/57 Met,Trp,Asp,Pro,Gly,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Lys,Leu,Val,Ala', '-aas B/58 Met,Trp,Asp,Tyr,Pro,Gly,Ser,Asn,Glu,Arg,Phe,Thr,Ile,Gln,Leu,Val,Ala', '-aas B/59 Met,Lys,Leu,Val', '-aas B/60 Met,Trp,Pro,Gly,Ser,Glu,Arg,Thr,Ile,Lys,Val,Ala', '-aas B/61 His,Met,Pro,Ser,Asn,Thr,Ile,Val,Ala']                     </pre>

**Figure 5 Single-site saturation analysis results of C118 CDR sites against WT SARS-CoV2 RBD:** Mutations with predicted improvements in overall Rosetta energy score, interaction energy, and shape complementarity than WT sequence choices were reported for each position. “B” refers to the heavy chain, “H” refers to the light chain, and “-”aas” refers to amino acid site positions.



**Figure 6 Analysis of C118 combinatorial design results against WT SARS-CoV-2**

**RBD:** Four combinatorial designs were performed as follows. ssDes\_all – used all residue choices that showed up to improve the total energy (Figure 5 list 1). ssDes\_all\_inter\_energy – used residues that showed up to both improve the total energy and the interaction energy (residues that appear in both Figure 5 List 1 and List 2). ssDes\_SC – used residues that showed up to both improve the total energy and the shape complementarity score (residues that appear in both Figure 5 List 1 and List 3). ssDes\_SC\_int – used the residues that showed up to improve the total energy, interaction energy, and shape complementarity (residues that appear in all three lists). Top reported sequences were analyzed for overall energy, interaction energy, and shape complementarity change over WT sequence. Data is reported as average and standard deviations.



\*Converted from pM using a MW of 150kDa

**Figure 7 Designed C002 variants were isolated by FACS to confirm design effectiveness:** (A) FACS sort of a MACS-enriched sub-library of C002 at 2 nM WT RBD concentration. 10 sequences were identified from the sorted cell population. (B) Potency improvement in one variant (v4) was confirmed in three independent technical repeats (data generated by Priyanthi Gnanapragasam and Jennifer Keeffe. ). In comparison, IC<sub>50</sub> values of approved nAbs and published C002 WT value were shown [5,13,14].

### *Bibliography*

1. P. Zhou, *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. R. Yan, *et al.*, Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science* (2020) <https://doi.org/10.1126/science.abb2762> (March 15, 2020).
3. D. Wrapp, *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
4. Q. Wang, *et al.*, Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* (2020) <https://doi.org/10.1016/j.cell.2020.03.045> (May 8, 2020).
5. D. F. Robbiani, *et al.*, Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).
6. C. O. Barnes, *et al.*, SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* (2020) <https://doi.org/10.1038/s41586-020-2852-1>.
7. D. M. Weinreich, *et al.*, REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *New England Journal of Medicine* **384**, 238–251 (2021).
8. “Fact Sheet For Health Care Providers Emergency Use Authorization (Eua) Of Bamlanivimab.” *FDA*, 2020, <https://www.fda.gov/media/143603/download> (Accessed Apr 30, 2021)
9. “Fact Sheet For Health Care Providers Emergency Use Authorization (Eua) Of Casirivimab and Imdevimab.” *FDA*, 2020, <https://www.fda.gov/media/143892/download> (Accessed Apr 30, 2021)
10. D. Zhou, *et al.*, Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348-2361.e6 (2021).
11. C. A. Jette, *et al.*, Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *bioRxiv*, 2021.04.23.441195 (2021).
12. R. F. Alford, *et al.*, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
13. J. Hansen, *et al.*, Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **369**, 1010–1014 (2020).
14. B. E. Jones, *et al.*, The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in non-human primates. *Science Translational Medicine* (2021) <https://doi.org/10.1126/scitranslmed.abf1906> (April 30, 2021).
15. *Triad*, Protabit LLC., <https://triad.protabit.com>
16. “InterfaceResidues” *PyMOL Wiki*, <https://pymolwiki.org/index.php/InterfaceResidues> (Accessed May 24, 2018)