**Computational *de novo* Design of Antibodies Binding to a Peptide with High Affinity**

Venkata Giridhar Poosarla1,2ǂ, Tong Li1ǂ, Boon Chong Goh3, Klaus Schulten3, Thomas K. Wood1,2\*, and Costas D. Maranas1\*

1Department of Chemical Engineering and 2Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, 16802, USA. 3Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA

\*To whom correspondence should be addressed: costas@psu.edu, tuw14@psu.edu

ǂ These authors have contributed equally to this work

**Supplementary Text 1.**

**MATERIALS AND METHODS**

**Computational generation of peptide-binding antibodies.** The designs were initiated by sampling positions of the dodecapeptide antigen (DVFYPYPYASGS) (PDB 4H0H, chain D) in a general antibody-binding site, which is represented by a rectangle grid box located close to the origin([Li et al. 2014](#_ENREF_21)). The X, Y and Z coordinates of the box were within the ranges of (−10 Å, 5 Å), (−5 Å, 10 Å) and (3.75 Å, 16.25 Å), respectively. The box was divided into a set of grid points by assigning grid spacings of 1.25, 1.25 and 1.25 Å for the X-, Y- and Z- axes, respectively. All dodecapeptide residues were identified as the epitope. During the positioning, the coordinate center of the epitope was placed into its corresponding grid box and rotated around the X-, Y-, Z-axes by 60°, 60° and 360° in steps of 30°, 30° and 20°, respectively, to generate an ensemble of initial antigen positions. 301,158 (13×13×11×3×3×18) initial positions of the antigen were generated in total. A clash-checking step was applied to filter out the antigen poses that clashed with a representative antibody framework (without six CDRs). Two structures are assumed clashed if at least a pair of antigen-antibody heavy atoms is within 1.5 Å distance. After clash checking, 241, 936 non-clash antigen conformations remained.

A library of "germline" antibody models was constructed to target each antigen pose by assembling six germline modular antibody parts (MAPs)([Pantazes and Maranas 2013](#_ENREF_28)). The interaction energies, including van der Waals and electrostatic contributions, between the dodecapeptide and the MAPs were calculated and stored. A “softening” atom van der Waals radius with a radius half of that employed in the CHARMM force field([Vanommeslaeghe et al. 2010](#_ENREF_39)) was used to estimate the hydrophobic interaction to avoid large penalty for atomic clashes. The modular antibody parts were previously constructed in the spirit of template-based modeling, with each part being a prototype structure of the random variable (V), diversity (D), and joining (J) regions in the MAPs database([Pantazes and Maranas 2013](#_ENREF_28)). The database contains 929 parts constructed from an analysis of 1168 human, humanized, chimeric, and mouse antibody structures and encompasses all currently observed structural diversity of antibodies. V, CDR3 and J structures can assemble both H and L chains of an antibody. There exist two types of light chain, KAPPA and LAMBDA, which are treated separately. Each MAPs structure was numbered using IMGT's unique numbering([Ehrenmann et al. 2011](#_ENREF_8); [Lefranc 2011a](#_ENREF_18); [Lefranc 2011b](#_ENREF_19); [Lefranc et al. 2005](#_ENREF_20)) and consistently placed in the 3D space so that its CDRs attachment points were approximately on the same X-Y plane and centered on the origin with CDRs perpendicularly directed in the positive Z direction. Once the interaction energies are calculated, the problem of selecting the best scoring combination of non-clashing antibody parts at each position could be mathematically represented using a MILP representation([Li et al. 2014](#_ENREF_21)).

The top 2,000 assembled germline antibody models with the lowest interaction energies were collected. The antigens were redocked to all the selected 2,000 antibodies using ZDOCK([Pierce et al. 2011](#_ENREF_30)) with default ZDOCK parameters. To improve the docking resolution, an extra block list that specified residues for blocking (parts of the framework residues) from the antibody-binding site was used as the constraint during the docking. For each antibody, the top 50 docked conformations of dodecapeptides were collected and compared with the best-positioned pose ranked by interaction energies by calculating the RMSDs. We used two independent criteria to select the best antibody candidates for further stability evaluation: (i) the top lowest interaction energies (20 scFvs chosen) and (ii) RMSD values between docked and the best-positioned poses lower than 4 Å (11 scFvs chosen). Hence, thirty-one antibodies were chosen for further refinement and stability evaluation using molecular dynamics.

**Molecular dynamics of designed antibodies with the dodecapeptide.** Each antibody-antigen complex was simulated for 100 ns to assess its binding stability and dynamics. Simulations were performed using NAMD 2.10([Phillips et al. 2005](#_ENREF_29)) with the CHARMM36 force field([Best et al. 2012](#_ENREF_2)) for the protein and the TIP3P model for water molecules. Each model was solvated in a sufficiently large water box such that the minimal distance between solute and box boundary was 15 Å along all three axes. The net charge of the solvated systems was adjusted to zero; the salinity of the solvent was set to 150 mM NaCl. The ionized systems were then minimized for 20,000 steps and subsequently thermalized to 310 K within 4 ps. All equilibrium simulations were performed in the NPT ensemble. Periodic boundary conditions were assumed, and the particle-mesh-Ewald summation method([Darden et al. 1993](#_ENREF_6)) was employed for the evaluation of Coulomb forces. The van der Waals energy was calculated using a cutoff of 12 Å. Temperature and pressure were maintained at 310 K and 1 atm, respectively, using a Langevin thermostat with a damping constant of 1 ps-1 and Nosé-Hoover Langevin piston methods. The integration time step was 2 fs. The non-bonded interactions were evaluated for every 2 fs, and electrostatics for every 4 fs. The procedure was followed by 10 ns of equilibrium simulation with the positions of protein backbone atoms restrained via a harmonic potential of force constant set to 1 kcal mol-1Å-2. Finally, the harmonic potential was removed and the systems were equilibrated for 100 ns.

**Computational affinity maturation.** The MD-refined antibody models (twenty-seven designs) were redesigned with Iterative Protein Redesign and Optimization (IPRO)([Pantazes et al. 2015](#_ENREF_26)) in order to find sequences that maximally improve the binding affinity and possess minimal computational immunogenicity. The standard IPRO design protocol was modified for use in OptMAVEn, which consists of five main steps([Li et al. 2014](#_ENREF_21)) in each iteration: (1) sequence design; (2) backbone perturbation; (3) optimal rotamer selection; (4) antigen redocking; and (5) energy evaluation. In step 1, a set of 1–3 continuous residues in either VH or VL is randomly selected for mutation. To increase the relevance of the identified designs, the permitted amino acid mutations at each framework position were pre-selected according to the amino acid frequency of each kind of amino acid at that position in alignments of existing antibodies (Supplementary Data 1). The residues in CDRs were allowed to mutate into any of the 20 standard amino acids. In step 2, the perturbed region, including 5 more residues on both sides of the mutation positions and surrounding residues within 4.5 Å, is subjected to backbone perturbation. In step 3, a rotamer library and mixed integer linear programing optimization algorithm are used to repack the amino acid side chains in and around the perturbed region. In step 4, the antigen is re-refined in the antibody-binding site by random perturbation and interaction energy evaluation based on which the movement of the antigen is kept or rejected using the Metropolis criterion ([Metropolis et al. 1953](#_ENREF_24)). In step 5, the complex and interaction energies are evaluated and a Metropolis criterion is used to determine whether or not to retain the results of this iteration. A high-resolution score function that evaluates van der Waals, electrostatics, bonds, angles, dihedral angles, improper dihedral angles, and generalized Born energies with molecular volume integration implicit solvation energy functions from CHARMM([Best et al. 2012](#_ENREF_2)) was used. We carried out 2,000 iterations for the respective calculations. Finally, the five best affinity-maturated computational designs, whose "germline" precursors maintain stably bound to the dodecapeptide during 100 ns-long MD simulations, were submitted to experimental validation.

**Reagents, bacterial strains, and cultivation conditions.** The dodecapeptide, DVFYPYPYASGS conjugated with BSA([Goel et al. 2004](#_ENREF_11)) was synthesized from LifeTein, LLC., Somerset, NJ. Cysteine was added to the C-terminus of the antigen, and the thiol group of the cysteine was used to conjugate the antigen with BSA (DVFYPYPYASGS-C-BSA). BSA was chosen because of its good water solubility, and it is commonly used in immunoassays([Gomez et al. 1998](#_ENREF_12)). Ratio of the dodecapeptide antigen to carrier protein (BSA) was 1:1. The relative protein concentration was determined using the Bradford assay with BSA as standard protein([Sun et al. 2014](#_ENREF_35)). Furthermore, BSA conjugation adds mass to the antigen that helps to study scFv-antigen binding interactions using the Octet QK instrument (ForteBio, Menlo Park, CA) because this method detects the proteins or peptides of the molecular weight range greater than 10 kDa([Mader and Kunert 2012](#_ENREF_23)).

*E. coli* HST08 (Clontech Laboratories Inc., Mountain View, CA) that provides high transformation efficiency (> 5 × 108 cfu/µg) was used to construct and propagate recombinant scFv expression plasmids. Oligonucleotide sequences used for amplifying the desired scFv sequences are given in Supplementary Table I. Expression plasmid pET27b(+) and expression host *E. coli* Rosetta (DE3) pLacI (the rare tRNA genes (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) are present on the same plasmid that carries the *lac* repressor gene) for expression of recombinant scFvs (Novagen/EMD Millipore, Billerica, MA). Kanamycin (50 µg/mL) was used to maintain the recombinant pET27b-scFv plasmids, and chloramphenicol (34 µg/mL) was used to maintain the *E. coli* Rosetta (DE3) plasmid pLacI. The *E. coli* was cultured in Luria-Bertani (LB) broth at 37 oC with shaking at 250 rpm, unless otherwise stated.

**scFvs cloning, expression, and purification**. The genes encoding the scFvs were synthesized as gBlocks obtained from Integrated DNA Technologies (IDT, Coralville, IA) and were codon optimized with respect to *E. coli* using the Codon Optimization Tool from IDT. The sequences of all *de novo* designed scFvs used in this work are available in the FASTA format in the Supplementary Data 2. The DNA sequences encoding the 5 best *de novo* scFv designs and scFv-2D10([Goel et al. 2004](#_ENREF_11); [Krishnan et al. 2007](#_ENREF_16); [Tapryal et al. 2013](#_ENREF_37); [Tapryal et al. 2010](#_ENREF_38)), were amplified with the respective primers (Supplementary Table I), and cloned into pET27b(+)([Ding et al. 2010](#_ENREF_7)) between the *Nco*I and *Xho*I sites. We used the In-FusionTM cloning technique (Clontech) to clone each of the amplified scFv fragments. The positive clones were identified by colony PCR and confirmed by DNA sequencing using T7 promoter (forward) and T7 terminator (reverse) primers (Supplementary Table I). The VH and VL domains of each scFv were attached by the linker (Gly4Ser)3([Campana et al. 2009](#_ENREF_4)) which provides flexibility, protects from proteases, keeps the scFv predominantly monomeric, and prevents interference with the function of the protein([Chichili et al. 2013](#_ENREF_5)). Each gene for the scFvs was preceded by DNA encoding the PelB signal peptide for periplasmic expression([Miller et al. 2005](#_ENREF_25)). Also, DNA encoding a His6 tag fused to the C-terminus of the light chain of each scFv with a short linker (GSG) was used to facilitate the protein to actively fold without being affected by the poly-histidine tag([Bachmann et al. 2005](#_ENREF_1)), and the His6 tag was followed by a stop codon. The advantage of the His6 tag joined to the C-terminus of the scFv instead of a C-terminal His6 tag encoded by the vector backbone was to avoid an additional 20 amino acids at the C-terminus. The His6 tag facilitated both purification and loading on the Ni-NTA biosensors (Octet).

*E. coli* Rosetta (DE3) pLacI was used to produce each His6-tagged scFv by following the protocol for isolation of proteins from inclusion bodies followed by *in vitro* refolding([Tapryal et al. 2010](#_ENREF_38)). *E. coli* Rosetta (DE3) pLacI cells harboring recombinant pET27b-scFv plasmids (scFv-1, 2, 3, 4, 5, and 2D10) were grown as 1000 mL cultures in terrific broth([Zarschler et al. 2013](#_ENREF_42)) until the turbidity at 600 nm reached 0.8, then isopropyl-β-D-thiogalactopyranoside (IPTG, 1 mM) was added to induce the genes encoding the scFvs. Growth was continued for 6 h at 37 oC, then the cells were chilled on ice for 10 min, harvested by centrifugation, and were suspended in 20 mL of buffer A (20 mM sodium phosphate, 500 mM NaCl, 40 mM imidazole, pH 7.4) supplemented with 200 µg/mL lysozyme (Sigma-Aldrich, St. Louis, MO) and 100µL of protease inhibitor cocktail (Sigma-Aldrich) and incubated on ice for 30 min. Sonication (Qsonica, LLC, Model Q700, Newtown, CT) was performed six times for 30 s at amplitude 45%. Following disruption of cells through sonication, the lysate was centrifuged at 18,000*g* for 20 min. The pellet was washed twice with 25 mL of buffer A with 0.5% Triton X-100 followed by centrifugation at 18,000*g* for 20 min to separate the contaminating cell debris from the inclusion body pellet. The inclusion body pellet was dissolved in 20 mL of solubilization buffer, buffer B (20 mM sodium phosphate, 500 mM NaCl, 40 mM imidazole, pH 7.4, containing 8 M urea), left at room temperature for 1 h with gentle shaking and centrifuged at 20,000*g* for 30 min to remove unsolubilized debris. The supernatant was loaded onto the 5 mL HisTrapTM FF column (GE Healthcare, Piscataway, NJ) equilibrated with the buffer B. Following 10 column volumes wash with buffer B, the bound protein was eluted with buffer C (20 mM sodium phosphate, 500 mM NaCl, 250 mM imidazole, pH 7.4, containing 8 M urea). The eluted proteins were analyzed by SDS-PAGE and the purified scFv protein fractions were pooled and concentrations were estimated by Bradford method using BSA as standard.

**scFv refolding.** *In vitro* refolding for the purified scFvs was performed as published previously([Tapryal et al. 2010](#_ENREF_38)) with slight modifications. To avoid aggregation, the purified proteins were diluted to 10 µM with buffer B. β-mercaptoethanol was added to the denatured protein solution to a final concentration of 10 mM, incubated for 1 h at room temperature with slow mixing and subsequently subjected to dialysis in Slide-A-Lyzer Dialysis Cassettes with a molecular weight cut-off of 10 kDa (Thermo Fisher Scientific, Waltham, MA) for protein refolding against 200 mL of buffer D (50 mM Tris-HCl, 100 mM NaCl, pH 8.0 containing 8 M urea and 1 mM EDTA). Step-wise dialysis (8 M, 6 M, 4 M, 2 M, 1 M, 0.5 M, 0.25 M, 0.125 M, and 0 M, each for 2 h) was performed to decrease the urea concentration gradually; and these solutions were prepared in buffer D with only the urea concentrations changing. For the change of urea concentration from 2 M to 0 M, 375 µM oxidized glutathione and 400 mM L-arginine hydrochloride were added to the refolding buffer. Later these additions were subsequently dialysed in four more steps, each step reducing the concentration to half. Lastly the refolded proteins were dialyzed three times (2 h each) against the final buffer, buffer E (50 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA, pH 8.0), with a final dialysis step overnight to completely remove the traces of chemicals/reagents used for protein refolding. The sequential buffer changes performed for efficient scFv refolding are outlined in detail in Supplementary Table II.

**Analysis of refolded scFvs by circular dichroism (CD).** The secondary structures of refolded scFvs were analyzed using Far-UV CD([Blanco-Toribio et al. 2014](#_ENREF_3); [Glaven et al. 2012](#_ENREF_10); [Song et al. 2014](#_ENREF_33)). CD spectra for all the proteins were collected on Jasco J-1500 Spectrophotometer (JASCO Ltd, UK) between wavelengths 240 nm and 190 nm at 25 oC in 10 mM Tris, 10 mM NaCl, 0.2 mM EDTA([Greenfield 2006](#_ENREF_13)) at pH 8.0. Spectra were averaged for three scans after subtracting the baseline scan (buffer in which refolded scFvs was excluded). The spectra were deconvoluted into α-helix, β-sheet, and unordered regions, using the CDPro software([Sreerama and Woody 2000](#_ENREF_34)).

***K*D determination via biolayer interferometry.** *K*D is the affinity constant, or equilibrium dissociation constant, which measures how tightly the ligand (scFv) binds to the analyte (dodecapeptide antigen). *K*D represents the ratio of the association rate constant (*k*a) to the dissociation rate constant (*k*d)([Hulme and Trevethick 2010](#_ENREF_14)). *K*D values for the study of antigen-antibody interactions were determined by bio-layer interferometry using an Octet QK instrument([Fischer et al. 2015](#_ENREF_9); [Lee et al. 2014](#_ENREF_17); [Prischi et al. 2010](#_ENREF_31); [Tang et al. 2014](#_ENREF_36)). His6 tagged *de novo* designed scFvs (15 µg/mL) in binding buffer (10 mM sodium phosphate, 50 mM NaCl, pH 7.4, 0.01% BSA, and 0.005% Tween 20)([Lee et al. 2014](#_ENREF_17)) were immobilized onto Dip and ReadTM Ni-NTA (NTA) Biosensors (ForteBio) which were pre-soaked off-line in 200 µl of binding buffer for 30 min, and incubated with varying concentrations of dodecapeptide-conjugated with BSA to obtain the best fit for the assessment of binding affinities. To rule out the non-specific binding, we tested a non-related scFv designed for the yeast transcription factor, GCN4 (YHLENEVARLKK-C-BSA) that activates transcription of genes involved in amino acid biosynthesis([Zahnd et al. 2004](#_ENREF_41)) with the antigen of the current study, as well as tested all the scFvs of the current system titrated against the non-related antigen, GCN4 (YHLENEVARLKK-C-BSA). All binding data were collected at 25 oC. The experiments included five steps: (i) determination of the baseline (300 s), (ii) scFv loading onto sensors (600 s), (iii) second baseline (60 s), (iv) association of antigen (*k*a, 900 s), and (v) dissociation of antigen (*k*d, 900 s). Baseline and dissociation steps were conducted in binding buffer. The *k*a and *k*d values of each scFv with dodecapeptide were used to determine *K*D using the ForteBio Data Analysis software 4.0.7([Weatherill et al. 2012](#_ENREF_40)) with a 1:1 global fitting model.

**Isothermal titration calorimetry.** Experiments were performed using a MicroCal AutoiTC200 (MicroCal, Inc., Northampton, MA), and both cell and syringe samples were loaded into a 96-well plate. Free dodecapeptide antigen, DVFYPYPYASGS (antigen not conjugated to BSA) was used in a syringe. The best binders from biolayer interferometry, scFv-1 (3 µM) and scFv-2D10 (8 µM) were dialyzed extensively in ITC binding buffer (10 mM sodium phosphate, 50 mM NaCl, 0.7% DMSO, pH 7.4) in sample cells and were titrated with 80 µM of dodecapeptide antigen in the same ITC binding buffer. All the titrations were conducted at 25 oC with the standard 5 µcal/s reference power. Each binding analysis was accompanied by a series of control experiments by titrating ITC binding buffer alone, buffer with dodecapeptide antigen, or scFvs with ITC binding buffer. The control runs (data not shown) suggested that the heat of dilution was minimal and constant, indicating the signals obtained during titrations were due to protein-peptide interactions only. The data were integrated and analyzed using the Origin AutoITC software (OriginLab, Northampton, MA). Before testing our samples, the instrument was calibrated using the EDTA test kit for the MicroCal AutoiTC200 supplied by the manufacturer.

**Supplementary Text 2.**

**RESULTS**

**Computational workflow.** OptMAVEn([Li et al. 2014](#_ENREF_21)) was used to *de novo* design a complete antibody variable domain targeting any specific epitope by expanding the concepts pioneered in OptCDR([Pantazes and Maranas 2010](#_ENREF_27)). However, OptMAVEn does not directly account for antibody dynamics. Protein design procedures as OptMAVEn often generate designs with poor stability characteristics given that functional sequences are so close to the limit of thermodynamic and colloidal stability([Li et al. 2013](#_ENREF_22); [Rouet et al. 2014](#_ENREF_32)). It was found that a pre-screening step using MD simulation prior to experimental testing is indispensable for discarding designs that exhibit poor stability and other design flaws that are inaccessible to static evaluations([Kiss et al. 2013](#_ENREF_15)). Therefore, we introduced all-atom MD simulation into our design protocol. Figure 1 pictorially depicts the updated OptMAVEn based design workflow: (1) Sampling of dodecapeptide (DVFYPYPYASGS) (Fig. 1a) conformations in the antibody-binding site (Fig. 1b); (2) Assembly of germline antibodies (variable regions of heavy and light chains) by best V, D, J modular antibody parts (MAPs) (Fig. 1c); (3) Structural refinement and stability evaluation using molecular dynamics (Fig. 1d); and (4) Iterative sequence design and structural relaxation to select affinity mature antibody sequences (Fig. 1e).

**Computational *de novo* design of "germline" antibodies binding the dodecapeptide.** Step 1 and 2 aim to generate "germline" antibody models with favorable interactions with the antigen. In step 1, we sampled 241,936 dodecapeptide conformations within the antibody-binding site. The conformations were selected not to involve any clash with the backbone of a representative antibody framework. To achieve such selection, the dodecapeptide was placed onto a grid of 1,859 points representing the binding site by a 15×15×12.5 Å rectangular cuboid and rotating it along the X, Y, Z coordinate. For each antigen conformation, the interaction energy (IE) including van der Waals and electrostatic terms between the antigen and all MAPs in the database were calculated and the top 2,000 best combinations of MAPs parts were selected based on mixed-integer linear programming (MILP)([Li et al. 2014](#_ENREF_21)). In step 2, we used the IE or root-mean-square deviation (RMSDs) between the docked and best-positioned antigen conformations filters to select thirty-one best *de novo* designed "germline" antibody models (Table I) assembled by six MAPs from the top 2,000 candidates for further evaluation using MD simulation. Twenty out of the thirty-one designs were selected based on (i) the lowest interaction energy between the antibody and the antigen criterion while (ii) eleven out of the thirty-one designs were selected based on lowest RMSD of the antigen to their docked conformation. The second criterion was introduced so as to capture designs that may score sub-optimally on interaction energy but are worth exploring further as the docked conformation matched well the chosen position of the antigens (step 1). In designs from both criteria, it can be seen that the V and CDR3s part in both VH and VL (Table I) exhibit considerable diversity, in accordance with their critical roles in recognizing different antigen poses.

**Conformation refinement and stability evaluation using molecular dynamics.** In step 3, all-atom MD simulation was usedfor computational stability evaluation and refinement of binding site residues with the dodecapeptide. Each one of the thirty-one antigen-antibody designed complexes generated under OptMAVEn steps 1 and 2 were simulated for 100 ns. Based on the dynamic characteristics of the antigen observed in the MD simulations (Fig. 2a, Supplementary Table III), we categorized our designs into four groups: stably bound (MD\_SB), relocating/reorientation (MD\_RE), partially bound (MD\_PB) and unbound (MD\_UB). Eight (group MD\_SB) of the thirty-one designs (Fig. 2b) remained stably bound to the antigen throughout the 100 ns-long MD simulations. Four out of the eight stable complexes remained bound to the antigen with an average root-mean-square deviation (RMSD) of less than 5 Å from the original coordinates. In all eight designs of group MD\_SB, MD simulations show some backbone flexibility of the CDRs (Complementarity Determining Regions) loops (Fig. 2b). In addition, the side-chains of several residues in direct contact with the dodecapeptide undergo reorganization in order to form more favorable contacts (Fig. 2b). The ten designs in group MD\_RE exhibited even more significant conformational changes with antigens relocating to new binding pockets (Fig. 2c) or re-positioning in the original pocket (Fig. 2d). From the remaining thirteen germline antibodies that did not show stable binding to the antigen, nine designs (group MD\_PB) dynamically evolved to only partially bound antigens (Fig. 2e) while four (group MD\_UB) became completely unbound from the complex during the MD simulation (Fig. 2f) and thus rejected. Finally, a total of twenty-seven designs from groups MD\_SB, MD\_RE and MD\_PB (Fig. 2g) with either MD-refined or initial geometries were retained and submitted for *in silico* maturation for improved affinity.

**Supplementary Text 3.**

**Explaining the role of molecular dynamics simulations in OptMAVEn, a computational method for antibody design.** OptMAVEn had been developed and employed in the past without use of MD simulations. In the present study, MD simulations were integrated into OptMAVEn to improve maturation that seeks to develop antibodies with optimal binding to a given antigen. The present study was carried out for the purpose of developing new methodology. To establish solid test criteria, the antigen chosen was one for which a native, i.e., highly evolved, antibody existed already, namely scFv-2D10.

Developing antibodies using solely experimental methods is a time-consuming process. Therefore, computational methods have been developed to design antibodies. One of the methods is OptMAVEn, which performs antibody design in two phases: antibody generation (phase A) and antibody maturation (phase B). Phase A involves generating antibodies from a library of antibody fragments, docking the generated antibodies to the target antigen, and score the resulting antigen-antibody complexes by their interaction energies. Phase B mimics the natural evolution of an antibody *in vivo* by mutating the amino acid residues of the antibody to enhance the interaction of the antigen-antibody complexes generated from phase A. The goal is to increase the binding constant between antibody and the given antigen by introducing computationally mutations into the antibody. The resulting antigen-antibody complexes are then scored computationally according to their antigen binding energies. Typically, around 5-10 complexes with the lowest interaction energies are selected for experimental validation.

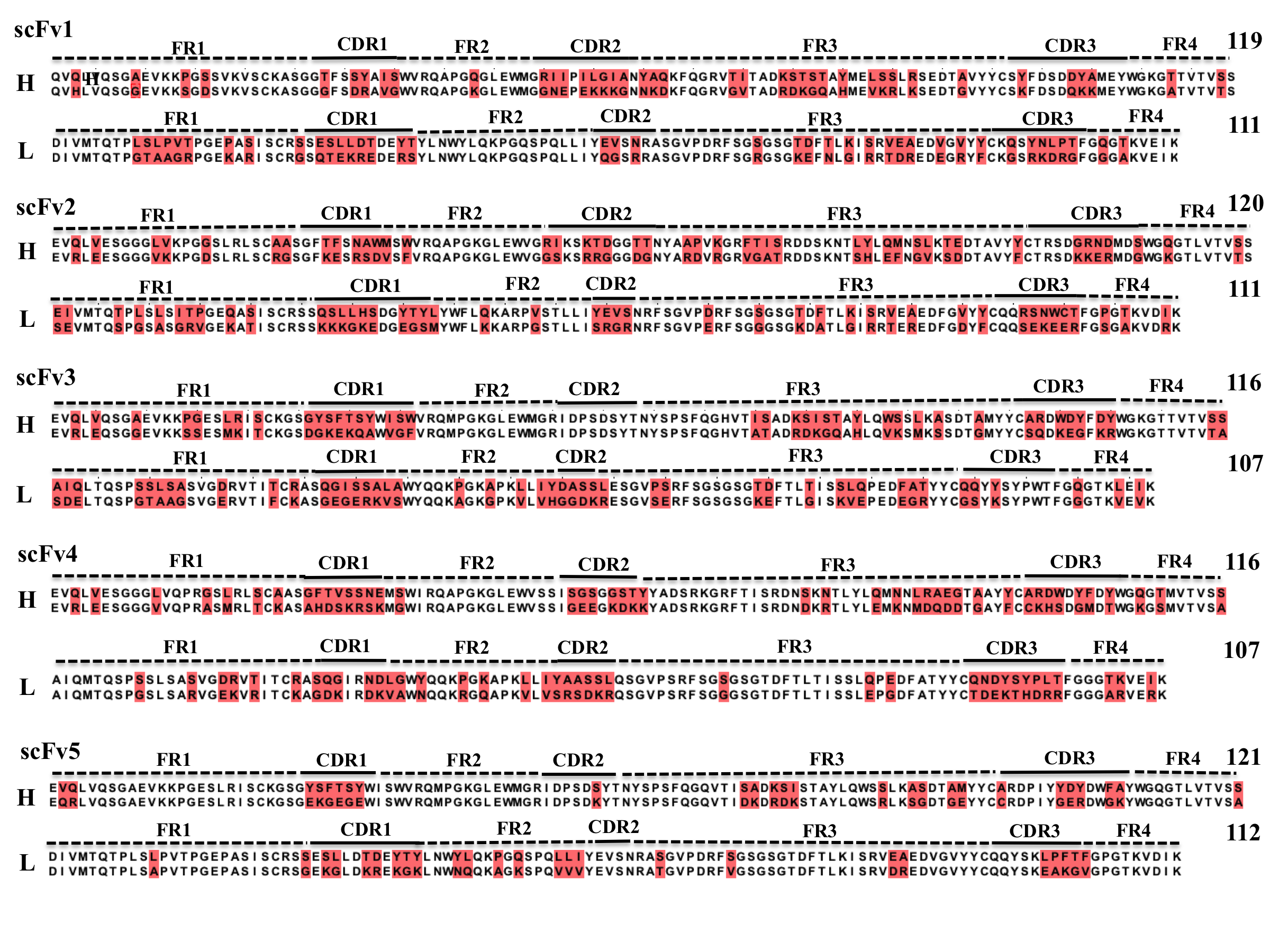
At this point starts the costly experimental phase of the project. The computationally determined antibodies are manufactured and exposed to the antigen. Dissociation constants, Kd, of the antibodies are measured using biolayer interferometry; smaller Kd corresponds to a higher binding affinity.

The success rate of computationally-designed antibodies, unfortunately, is very low. The calamity can partly be attributed to the lack of dynamic information during the design process as only static structures are considered. To generate antibodies of high quality, molecular dynamics (MD) simulation is incorporated in the present study into the design workflow of OptMAVEn to account for the dynamic nature of the antigen-antibody interaction. For this purpose the antigen-antibody complexes generated from phase A of OptMAVEn are subjected to MD simulations. The antigen-antibody complexes are each simulated independently starting in a way typical for MD simulations: the complexes are energy minimized and then thermally annealed ending at physiological temperature. The resulting solvated system is then subjected to a 100 ns of equilibrium simulation. The outcomes of the MD simulations are the structures reached at the ends of the equilibration runs. The outcomes are categorized into consistent binding, altered binding, partial binding and unbinding cases depending on if the antigen remained, at the end of the 100 ns simulations, completely bound to the same binding pocket, relocated to a new binding pocket, partially bound, or completely unbound. The designs with the antibody unbound from the antigen were discarded from the design workflow. The simulated structures of the designs with bound or partially-bound antibody are used as input structures for phase B of OptMAVEn. The top antibody designs, ranked by their antigen-antibody interaction energies, are then selected for experimental validation.

In the present study, 31 antibody designs were generated in phase A. Each antigen-antibody system was subjected 100 ns of equilibrium simulation. Out of the 31 antibody designs, 8 designs showed consistent binding, 10 designs exhibited altered binding, 9 designs remained partially bound, and 4 designs completely unbound from the antigen. The 4 bad designs were removed from the design workflow; the 27 simulated structures of the antigen-antibody complexes were used in phase B where mutations were computationally introduced to enhance binding. Finally, 5 top-ranked antibodies were chosen for experimental validation, where the binding affinities between the antibody and antigen were measured. The experiments revealed that 3 out of the 5 antibodies reported binding affinities that are comparable to 2D10, the native antibody.

The MD simulations performed in the present study not only assessed the stability of the antigen-antibody complexes, but also refined the binding interface of the antibodies that is instrumental to the accuracy of the antibody maturation calculations. In short, OptMAVEn combined with MD simulation has been demonstrated to have a great capability in generating high-affinity antibodies via *de novo* design and *in silico* optimization.

**Supplementary Figure 1.** **Alignments between mature and "germline" antibody sequences for the five *de novo* designs.** FRs, CDRs regions and the lengths of sequences are indicated on top of each alignment. Red shading shows introduced amino acid mutations. The lengths of each sequence are also labeled.

****

**Supplementary Figure 2. ITC titrations of scFvs (scFv-1 and scFv-2D10) with the dodecapeptide antigen.** Top panels are the raw thermograms while the bottom panels are the integrated data. The line in the integrated data panels represents the results of a single-site binding model fit. Eighty µM of dodecapeptide antigen (free peptide) was titrated into 3 µM of scFv-1 and 8 µM of scFv-2D10 at 25 oC.

C:\Users\Giridhar\Dropbox\Dr. Thomas K. Wood Lab PSU\Alternative Systems\Proteins MS\ITC\Refit 09222016\ITC Final 09222016.tif

**Supplementary Table I.** Primers used in this study.

|  |  |
| --- | --- |
| Primers | DNA sequence (5’-3’) |
| scFv-1-F | CCAGCCGGCGATGGCCCAGGTGCATCTGGTCCAAAGT |
| scFv-1-R | CCCGTTTGATCTCGAGTCAATGATGGTGGTGATGATG |
| scFv-2-F | CCAGCCGGCGATGGCCGAAGTGCGCCTCGAGGAATCC |
| scFv-2-R | CCCGTTTGATCTCGAGTCAATGATGGTGGTGGTGGTG |
| scFv-3-F | CCAGCCGGCGATGGCCGAAGTTCGCCTGGAACAGAGC |
| scFv-3-R | CCCGTTTGATCTCGAGTCAATGATGATGGTGGTGATG |
| scFv-4-F | CCAGCCGGCGATGGCCGAAGTTCGCCTGGAAGAAAGC |
| scFv-4-R | CCCGTTTGATCTCGAGTCAATGATGGTGATGGTGGTG |
| scFv-5-F | CCAGCCGGCGATGGCCGAACAGCGCTTGGTACAGTCG |
| scFv-5-R | CCCGTTTGATCTCGAGTCAGTGATGATGGTGGTGGTG |
| scFv-2D10-F | CCAGCCGGCGATGGCCATGGAAATCCAGTTACAGCAG |
| scFv-2D10-R | CCCGTTTGATCTCGAGTCAATGATGATGGTGGTGATG |
| T7-F | TAATACGACTCACTATAGGG |
| T7term-R | GCTAGTTATTGCTCAGCGG |

**Supplementary Table II.** Temporal change in dialysis buffers for the quick and efficient refolding of scFvs purified from inclusion bodies.All solutions were 1000 mL. T: 50 mM Tris; N**:** 100 mM NaCl; E: 1 mM EDTA; pHs adjusted to pH 8.0 with concentrated HCl.

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Time | Buffer Change (From) | Buffer Change (To) |
|
| 1 | O/N | TNE + 8 M urea + 0.25 M imidazole + 10 mM β mercaptoethanol | TNE + 8 M urea |
| 2 | 2 h | TNE + 8 M urea | TNE + 6 M urea |
| 3 | 2 h | TNE + 6 M urea | TNE + 4 M urea |
| 4 | 2 h | TNE + 4 M urea | TNE + 2 M urea |
| 5 | 2 h | TNE + 2 M urea | TNE + 1 M urea + 400 mM Arginine + 375 µM Glutathione oxidized |
| 6 | 2 h | TNE + 1 M urea + 400 mM arginine + 375 µM glutathione oxidized | TNE + 0.5 M Urea + 400 mM Arginine + 375 µM Glutathione oxidized |
| 7 | O/N | TNE + 0.5 M urea + 400 mM arginine + 375 µM glutathione oxidized | TNE + 0.25 M urea + 400 mM arginine + 375 µM glutathione oxidized |
| 8 | 2 h | TNE + 0.25 M urea + 400 mM arginine + 375 µM glutathione oxidized | TNE + 0.125 M urea + 400 mM arginine + 375 µM glutathione oxidized |
| 9 | 2 h | TNE + 0.125 M urea + 400 mM arginine + 375 µM glutathione oxidized | TNE + 0 M urea + 400 mM arginine + 375 µM glutathione oxidized |
| 10 | 2 h | TNE + 0 M urea + 400 mM arginine + 375 µM glutathione oxidized | TNE + 0 M urea + 200 mM arginine + 188 µM glutathione oxidized |
| 11 | 2 h | TNE + 0 M urea + 200 mM arginine + 188 µM glutathione oxidized | TNE + 0 M urea + 100 mM arginine + 94 µM glutathione oxidized |
| 12 | O/N | TNE + 0 M urea + 100 mM arginine + 94 µM glutathione oxidized | TNE + 0 M urea + 50 mM arginine + 47 µM glutathione oxidized |
| 13 | 2 h | TNE + 0 M urea + 50 mM arginine + 47 µM glutathione oxidized | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized |
| 14 | 2 h | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized |
| 15 | 2 h | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized |
| 16 | 2 h | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized |
| 17 | O/N | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized | T + N (50 mM NaCl) + E + 0 M urea + 0 mM arginine + 0 µM glutathione oxidized |

**Supplementary Table III.** The results of molecular dynamics.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group | Design | frame used | selected | RMSD (20ns) | RMSD (50ns) | RMSD (100ns) |
| antigens stably bound | 120\_15439\_2 | 100 | y | 7.65 ± 0.48 | 6.86 ± 0.55 | 5.83 ± 1.40 |
| 120\_6290\_1 | 100 | y | 3.41 ± 0.36 | 3.40 ± 0.57 | 4.67 ± 0.22 |
| 140\_10149\_1 | 100 | y | 3.65 ± 0.45 | 5.74 ± 0.57 | 5.47 ± 0.40 |
| 140\_15899\_5 | 100 | y | 1.92 ± 0.26 | 5.28 ± 0.88 | 3.66 ± 0.32 |
| 160\_8161\_1 | 100 | y | 3.11 ± 0.61 | 5.41 ± 0.59 | 4.89 ± 0.78 |
| 160\_15235\_2 | 100 | y | 4.66 ± 0.42 | 3.25 ± 0.31 | 4.05 ± 0.22 |
| 200\_15222\_2 | 100 | y | 6.93 ± 0.35 | 6.04 ± 0.41 | 6.77 ± 0.49 |
| 200\_15222\_5 | 100 | y | 5.71 ± 0.42 | 6.30 ± 0.35 | 6.96 ± 1.10 |
| antigens bound to new binding pocket or orientation | 120\_10148\_1 | 100 | y | 6.95 ± 0.42 | 6.89 ± 0.64 | 7.06 ± 0.73 |
| 120\_13234\_5 | 100 | y | 5.64 ± 0.17 | 5.82 ± 0.35 | 5.24 ± 0.26 |
| 120\_13389\_2 | 100 | y | 8.99 ± 1.32 | 6.76 ± 1.70 | 9.34 ± 0.74 |
| 140\_16152\_1 | 100 | y | 9.64 ± 0.71 | 19.25 ± 3.62 | 11.56 ± 0.30 |
| 140\_9977\_4 | 100 | y | 11.21 ± 0.52 | 5.40 ± 0.40 | 3.94 ± 0.78 |
| 160\_10162\_1 | 100 | y | 7.04 ± 1.67 | 10.18 ± 0.55 | 9.79 ± 0.39 |
| 160\_10175\_1 | 100 | y | 6.43 ± 0.55 | 8.19 ± 0.38 | 7.88 ± 0.54 |
| 160\_14896\_1 | 100 | y | 10.37 ± 1.10 | 6.82 ± 0.40 | 7.24 ± 0.43 |
| 200\_15729\_2 | 100 | y | 3.48 ± 0.52 | 6.14 ± 0.38 | 8.02 ± 0.49 |
| 240\_4115\_2 | 100 | y | 4.18 ± 0.25 | 6.08 ± 0.47 | 8.65 ± 1.46 |
| antigens partially bound | 120\_10148\_2 | 0 | y | 9.86 ± 0.83 | 6.24 ± 1.03 | 6.79 ± 0.85 |
| 140\_13234\_5 | 50 | y | 4.60 ± 0.48 | 4.95 ± 0.51 | 5.37 ± 0.39 |
| 140\_4402\_1 | 0 | y | 7.19 ± 0.77 | 7.92 ± 0.69 | 10.49 ± 1.09 |
| 140\_9976\_2 | 50 | y | 4.25 ± 0.66 | 5.60 ± 0.47 | 7.74 ± 0.70 |
| 160\_10005\_1 | 0 | y | 6.54 ± 1.01 | 5.86 ± 0.74 | 7.93 ± 1.05 |
| 160\_10005\_2 | 0 | y | 13.11 ± 0.50 | 13.19 ± 0.67 | 12.35 ± 1.04 |
| 160\_10162\_2 | 0 | y | 8.05 ± 0.50 | 7.97 ± 0.65 | 13.20 ± 0.90 |
| 160\_15235\_1 | 20 | y | 4.46 ± 0.66 | 4.93 ± 0.57 | 9.22 ± 1.11 |
| 160\_8161\_2 | 0 | y | 7.35 ± 0.61 | 6.80 ± 0.92 | 12.69 ± 0.82 |
| antigens unbound | 120\_13389\_1 | 0 | n | 11.62 ± 1.56 | 29.65 ± 0.57 | 44.44 ± 5.26 |
| 120\_15439\_1 | 0 | n | 12.72 ± 0.83 | 15.66 ± 0.83 | 26.95 ± 1.00 |
| 140\_10149\_2 | 0 | n | 15.16 ± 1.82 | 16.52 ± 0.55 | 16.81 ± 1.04 |
| 140\_9976\_1 | 0 | n | 20.47 ± 2.90 | 14.40 ± 1.43 | 13.60 ± 1.78 |

**Supplementary Table IV.** Summary of energies of the five *de novo* designed scFvs against peptide 2D10 after in silico affinity maturation.

|  |  |  |  |
| --- | --- | --- | --- |
| Antibody | MD description | Complex  Energya | IEb |
| 120\_6290\_1 (scFv-1) | antigens stably bind | -11939 | -474 |
| 120\_15439\_2 (scFv-2) | -11629 | -428 |
| 140\_10149\_1 (scFv-3) | -7168 | -360 |
| 140\_15899\_5 (scFv-4) | -11110 | -550 |
| 160\_15235\_2 | -9089 | -342 |
| 160\_8161\_1 | -8548 | -281 |
| 200\_15222\_2 | -7954 | -282 |
| 200\_15222\_5 (scFv-5) | -8704 | -225 |
| 120\_10148\_1 | antigens bind to a new binding pocket or adopt a new orientation | -11057 | -394 |
| 120\_13234\_5 | -11047 | -523 |
| 120\_13389\_2 | -11278 | -428 |
| 140\_9977\_4 | -9197 | -462 |
| 140\_16152\_1 | -7441 | -353 |
| 160\_10175\_1 | -8437 | -370 |
| 160\_14896\_1 | -9595 | -213 |
| 160\_10162\_1 | -7827 | -242 |
| 200\_15729\_2 | -10077 | -399 |
| 240\_4115\_2 | -8007 | -329 |
| 120\_10148\_2 | antigen partially bind | -8680 | -323 |
| 140\_4402\_1 | -8825 | -322 |
| 140\_9976\_2 | -8335 | -330 |
| 140\_13234\_5 | -11432 | -340 |
| 160\_8161\_2 | -6427 | -191 |
| 160\_10005\_1 | -5576 | -135 |
| 160\_10005\_2 | -9177 | -353 |
| 160\_10162\_2 | -7827 | -242 |
| 160\_15235\_1 | -7781 | -290 |

a The entire complex energy. Unit in kcal/mol.

b The interaction energy between the antibody and antigen using CHARMM force field. Unit in kcal/mol.

**Supplementary Table V.** Comparison of sequences of five *de novo* designed scFvs with those of existing antibody sequences using BLAST search.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Antibody | Chain | Description of antibodies from the first hit by BLAST search | Species | Accession number | Identitya |
| scFv-1 | Heavy | Chain H, Crystal Structure Of Human Germline Antibody 1-69B3 | Homo sapiens | 3QOT\_H | 60% |
| Light | Immunoglobulin G heavy chain variable region | Homo sapiens | AIT38683.1 | 63% |
| scFv-2 | Heavy | Immunoglobulin A heavy chain variable region | Homo sapiens | AGP01213.1 | 59% |
| Light | Immunoglobulin kappa chain variable region | Homo sapiens | AIZ06502.1 | 59% |
| scFv-3 | Heavy | Immunoglobulin heavy chain variable region | Homo sapiens | BAI52483.1 | 65% |
| Light | Immunoglobulin light chain variable region | Homo sapiens | ABG38375.1 | 60% |
| scFv-4 | Heavy | Anti-oligomeric synuclein single-chain Fv antibody D5E | Synthetic construct | AFR23376.1 | 60% |
| Light | Immunoglobulin light chain variable region | Homo sapiens | AKU38972.1 | 64% |
| scFv-5 | Heavy | Immunoglobulin variable region | Homo sapiens | CAA81438.1 | 74% |
| Light | Immunoglobulin kappa chain | Macaca mulatta | AAR84040.1 | 71% |

a Sequence similarity. Only the top hit with the maximum identify was reported.

**Supplementary Table VI.** Analysis of CD spectra of 6 scFvs with CDSSTR, CONTIN and GOR4. CDSSTR and CONTIN values are actual values deconvoluted from raw data, and GOR4 is an information theory-based method for the prediction of secondary structures in proteins. r: regular; d: distorted; U: Unordered. For GOR4 analysis, the values in parentheses for α-helix, β-sheet and Unordered indicate the number of amino acids considered for analysis.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Antibody | CDSSTR | | | | | CONTIN | | | | | GOR4 | | |
| **α helix** | | **β sheet** | | **U** | **α helix** | | **β sheet** | | **U** | **α**  **helix** | **β**  **sheet** | **U** |
| **r** | **d** | **r** | **D** | **r** | **d** | **r** | **d** |
| scFv-1 | 0 | 2.6 | 22.4 | 13.3 | 37.1 | 0.8 | 5.2 | 24.3 | 14.1 | 33.5 | 7.87 (20) | 28.74 (73) | 63.39 (161) |
| scFv-2 | -0.2 | 2.6 | 23.9 | 13.6 | 35.0 | 1 | 5.3 | 24.6 | 12.9 | 34.6 | 3.53 (9) | 27.06 (69) | 69.41 (177) |
| scFv-3 | -0.1 | 3.7 | 24.0 | 13.3 | 34.9 | 0.2 | 6 | 22.7 | 13.0 | 35.2 | 10.12 (25) | 33.20 (82) | 56.68 (140) |
| scFv-4 | 0.1 | 4.1 | 21.7 | 13.6 | 34.6 | 1 | 5.5 | 22.7 | 13.6 | 33.0 | 11.74 (29) | 29.55 (73) | 58.7 (145) |
| scFv-5 | 0.8 | 4.0 | 21.4 | 12.7 | 35.5 | 3.2 | 5.8 | 22.5 | 11.6 | 34.4 | 8.17 (21) | 27.24 (70) | 64.59 (166) |
| scFv-2D10 | 0.7 | 2.6 | 25.9 | 13.6 | 34.8 | 0.3 | 5.1 | 25.4 | 13.0 | 33.9 | 8.63 (22) | 30.2 (77) | 61.18 (156) |

**Supplementary Table VII.** Thermodynamic parameters for the two scFvs binding to dodecapeptide antigen obtained by ITCa at 25 oC.

|  |  |  |  |
| --- | --- | --- | --- |
|  | scFv-1 | scFv-2D10 | Positive controlb |
| *N* | 0.93 ± 0.04 | 0.91 ± 0.03 | 0.96 (0.957) |
| *ΔH* (cal/mol) | -561 ± 35 | -779 ± 36 | -4021±76  (-4327) |
| *ΔS* (cal/mol/deg) | 30.1 | 29.4 | 10.2 (14.5) |
| *K*d (nM) | 100 ± 20 | 99 ± 20 | 6000 ± 33000 (7000) |

a. Errors are from model fitting

b. Positive control (MicroCal AutoiTC200), EDTA, 0.4 mM and CaCl2, 5mM in 10 mM MES buffer at pH 5.6. Values in parentheses are the original standard values indicated by the manufacturer.

**Supplementary Data 1.** Permitted amino acid kinds during the computational affinity maturation for H and L chains of the scFvs.

IMGT residue number Permitted amino acid kinds

1: ['Q' 'S' 'E']

2: ['Q' 'V']

3: ['Q' 'H' 'R' 'N']

4: ['L']

5: ['E' 'L' 'Q''S' 'R' 'V']

6: ['Q' 'E']

7: ['S']

8: ['G']

9: ['A' 'P' 'S' 'T' 'G']

11: ['A' 'E' 'D' 'G' 'Q' 'V']

12: ['I' 'M' 'L' 'V']

13: ['K' 'R' 'T' 'V']

14: ['Q' 'P' 'K' 'T']

15: ['P' 'S']

16: ['S' 'T' 'G']

17:['A' 'E' 'D' 'K' 'Q' 'S' 'R']

18: ['I' 'S' 'T']

19: ['I' 'P' 'M' 'L' 'V']

20: ['K' 'R' 'S' 'T']

21: ['I' 'L' 'V']

22: ['P' 'S' 'T']

23: ['C']

24: ['A' 'E' 'K' 'Q' 'S' 'R' 'T' 'W' 'V']

25: ['A' 'G' 'F' 'S' 'T' 'V']

26: ['A' 'S']

27: ['E' 'D' 'G']

28: ['D' 'G' 'F' 'I' 'L' 'P' 'S' 'Y']

29: ['E' 'D''F' 'I' 'K' 'M' 'N' 'P' 'S' 'R' 'T']

30: ['E' 'F' 'I' 'M' 'L' 'S' 'T' 'W' 'Y']

35: ['G' 'I' 'L' 'N' 'S' 'R' 'T']

36: ['A' 'E' 'D' 'G' 'N' 'S' 'T' 'Y']

37: ['C' 'E' 'F' 'H' 'M' 'S' 'T' 'Y']

38: ['A' 'G' 'F' 'I' 'H' 'L' 'P' 'S' 'T' 'V' 'Y']

39: ['I' 'M' 'L' 'W' 'V']

40: ['H' 'S' 'T' 'G' 'N']

41: ['W' 'F']

42: ['I' 'L' 'W' 'V']

43: ['R']

44: ['Q' 'L']

45: ['A' 'I' 'S' 'P']

46: ['P']

47: ['D' 'G']

48: ['Q' 'K' 'R']

49: ['Q' 'A' 'S' 'R' 'G']

50: ['P' 'L' 'F']

51: ['Q' 'E' 'D']

52: ['W']

53: ['I''M' 'L' 'V']

54: ['A' 'S' 'T' 'G']

55: ['G' 'I' 'L' 'S' 'W' 'V' 'Y']

56: ['I' 'M' 'L' 'V']

57: ['D''F' 'I' 'H' 'K' 'N' 'S' 'T' 'W' 'Y']

58: ['A' 'G' 'H' 'P' 'T' 'W' 'Y']

59: ['A' 'D' 'G' 'I' 'K''L' 'N' 'S' 'R' 'T' 'V' 'Y']

62: ['D' 'G' 'F' 'I' 'N' 'S' 'T' 'W' 'Y']

63: ['A' 'S' 'D' 'G']

64: ['A' 'E' 'D' 'K' 'N' 'Q' 'S' 'T']

65: ['I' 'K' 'Q' 'P' 'R' 'T' 'V' 'Y']

66: ['A' 'E' 'D' 'K' 'N' 'Q' 'S' 'Y']

67: ['Y' 'N' 'F']

68: ['A' 'P' 'S' 'K']

69: ['A' 'D' 'Q' 'P' 'S' 'R' 'W']

70: ['D' 'K''N' 'Q' 'P' 'S' 'R' 'Y']

71: ['V' 'M' 'L' 'F']

72: ['E' 'H' 'K' 'N' 'Q' 'R']

74: ['A' 'D' 'G' 'H''Q' 'S' 'T']

75: ['R' 'W']

76: ['I' 'F' 'L' 'V']

77: ['G' 'I' 'N' 'Q' 'S' 'T']

78: ['A' 'F' 'I' 'M''L' 'V']

79: ['S' 'D' 'T']

80: ['A' 'K' 'R' 'T' 'V']

81: ['Q' 'H' 'D']

82: ['A' 'E' 'D' 'G' 'I' 'L' 'N' 'R' 'T' 'V']

83: ['Y' 'A' 'S' 'D' 'F']

84: ['A' 'E' 'I' 'K' 'M' 'L' 'Q' 'S' 'R' 'T' 'W']

85: ['E' 'D' 'G' 'F' 'N' 'S' 'R' 'T']

86: ['A' 'Q' 'S' 'T' 'V']

87: ['A' 'S' 'V' 'L' 'F']

88: ['Y' 'H' 'S' 'V' 'F']

89: ['M' 'L']

90: ['E' 'D' 'K' 'Q' 'R' 'T']

91: ['I' 'V' 'M' 'L' 'F']

92: ['K' 'R' 'S' 'T' 'N']

93: ['A' 'G' 'F' 'N' 'S' 'R']

94: ['M' 'L' 'V']

95: ['S' 'K' 'R' 'T' 'D']

96: ['A' 'Q' 'P' 'S' 'V' 'Y']

97: ['A' 'E' 'D' 'G' 'V']

98: ['D']

99: ['T']

100: ['A' 'G']

101: ['E' 'I' 'M' 'L' 'T' 'V']

102: ['Y']

103: ['Y' 'F']

104: ['C']

105: ['A' 'T' 'V']

106: ['S' 'R' 'K' 'V']

107: ['A' 'E' 'D' 'G' 'H' 'L' 'Q' 'P' 'R' 'T' 'V'] \

108: ['A' 'E' 'G' 'K' 'M' 'L' 'Q' 'P' 'R' 'V']

109: ['F' 'L' 'N' 'Q' 'P' 'S' 'R' 'Y']

110: ['C' 'E' 'G' 'K' 'M' 'S' 'Y']

111: ['E' 'D' 'G' 'Q' 'P' 'R' 'T' 'W' 'V']

112: ['A' 'E' 'D' 'G' 'F' 'K' 'N' 'S' 'R' 'W' 'V' 'Y']

113: ['A' 'E' 'G' 'F' 'S' 'R' 'W' 'Y']

114: ['A' 'D' 'G' 'H' 'P' 'S' 'T' 'W' 'Y']

115: ['F' 'M' 'L' 'S' 'W' 'Y']

116: ['A' 'E' 'D' 'G' 'H' 'Q' 'R']

117: ['H' 'K' 'L' 'N' 'P' 'V' 'Y']

118: ['W']

119: ['S' 'G']

120: ['Q' 'P' 'K' 'R' 'S']

121: ['G']

122: ['A' 'I' 'S' 'T']

123: ['M' 'L' 'Q' 'P' 'T' 'V']

124: ['I' 'V']

125: ['A' 'I' 'S' 'T' 'V']

126: ['I' 'V']

127: ['S' 'T']

128: ['A' 'P' 'S']

Permitted amino acid kinds for L chain

1: ['E' 'D' 'Q' 'S' 'A']

2: ['I' 'P' 'S' 'Y']

3: ['Q' 'A' 'E' 'V']

4: ['M' 'L']

5: ['T']

6: ['Q']

7: ['P' 'S']

8: ['P']

9: ['A' 'S' 'T' 'G']

10: ['I' 'S' 'T']

11: ['A' 'L' 'V']

12: ['A' 'S']

13: ['A' 'L' 'G' 'V']

14: ['A' 'S' 'R' 'T']

15: ['P' 'L' 'V']

16: ['G']

17: ['Q' 'E' 'D']

18: ['A' 'K' 'R' 'T']

19: ['A' 'I' 'V']

20: ['I' 'S' 'R' 'T']

21: ['I' 'L' 'F']

22: ['S' 'N' 'T' 'F']

23: ['C']

24: ['G' 'K' 'Q' 'S' 'R' 'T']

25: ['A' 'S' 'T' 'G']

26: ['A' 'S' 'G' 'N']

27: ['D' 'G' 'I' 'H' 'N' 'Q' 'S' 'R' 'V']

28: ['G' 'F' 'I' 'S' 'R' 'T' 'Y']

29: ['F' 'G' 'H' 'K' 'M' 'S']

36: ['F' 'N' 'R' 'S' 'T']

37: ['D' 'G' 'I' 'K' 'N' 'S' 'R' 'Y']

38: ['A' 'D' 'H' 'P' 'S' 'R' 'T' 'V' 'Y']

39: ['M' 'L' 'V']

40: ['A' 'C' 'H' 'N' 'Q' 'S' 'T' 'Y']

41: ['W']

42: ['Y' 'N' 'F']

43: ['Q']

44: ['Q' 'H' 'K' 'V']

45: ['F' 'K' 'L' 'R' 'T' 'V']

46: ['A' 'P' 'R']

47: ['G']

48: ['Q' 'A' 'K' 'R' 'T']

49: ['A' 'P' 'S' 'G''V']

50: ['P']

51: ['K' 'R' 'E' 'V']

52: ['L' 'V']

53: ['L' 'V']

54: ['I' 'V']

55: ['Y' 'H' 'S' 'C' 'F']

56: ['A' 'E' 'D' 'G' 'S' 'R' 'W' 'Y']

57: ['A' 'D' 'G' 'N' 'S' 'T' 'V']

65: ['Y' 'S' 'E' 'D' 'N']

66: ['E' 'D' 'I' 'H' 'K' 'N' 'Q' 'R' 'T' 'V']

67: ['R' 'L']

68: ['A' 'H' 'Q' 'E' 'P']

69: ['A''S' 'R' 'T' 'G']

70: ['G' 'V']

71: ['I' 'V']

72: ['P' 'S']

74: ['A' 'S' 'E' 'D' 'T']

75: ['R']

76: ['I' 'F']

77: ['S' 'V']

78: ['A' 'G']

79: ['S' 'R' 'G']

80: ['Q' 'K' 'R' 'G' 'N']

83: ['S' 'W' 'F']

84: ['A' 'H' 'G']

85: ['A' 'K' 'N' 'Q' 'P' 'S' 'T']

86: ['E' 'D' 'N' 'Q' 'S' 'T']

87: ['A' 'Y' 'F']

88: ['S' 'N' 'T' 'F']

89: ['L']

90: ['A' 'I' 'S' 'T' 'G']

91: ['I']

92: ['S' 'R' 'T' 'N']

93: ['G' 'K' 'N' 'S' 'R' 'T']

94: ['V' 'M' 'L' 'T']

95: ['Q' 'E' 'D']

96: ['A' 'P' 'S' 'R' 'T' 'V']

97: ['I' 'A''E' 'D' 'G']

98: ['D']

99: ['I' 'V' 'E' 'F']

100: ['A' 'G']

101: ['E' 'D' 'G' 'L' 'N' 'R' 'T' 'V']

102: ['Y']

103: ['Y' 'F']

104: ['C']

105: ['A' 'Q' 'H']

106: ['Q' 'A' 'T' 'H' 'V']

107: ['F' 'H' 'L' 'W' 'V' 'Y']

108: ['D' 'G' 'S' 'Y']

109: ['D' 'G' 'I' 'N' 'R' 'S' 'T']

110: ['A' 'L' 'R' 'S' 'T' 'W']

113: ['L' 'P' 'S' 'T']

114: ['D' 'F' 'K' 'N' 'P' 'R' 'S' 'T' 'W']

115: ['A' 'G' 'H' 'Q' 'R']

116: ['A' 'E' 'D' 'G' 'Q' 'W' 'V']

117: ['A' 'F' 'I' 'S' 'T' 'W' 'V']

118: ['I' 'V' 'F']

119: ['G' 'V']

120: ['Q' 'P' 'S' 'L' 'G']

121: ['G']

122: ['A' 'S' 'T']

123: ['Q' 'K' 'R' 'E']

124: ['L' 'V']

125: ['Q' 'V' 'E' 'D' 'T']

126: ['I' 'R' 'L' 'V']

127: ['K' 'L']

**Supplementary Data 2. Sequences of *de novo* designed scFvs used in binding studies.** The sequences listed below represent the full-length ORF as cloned in the pET-27b(+) expression vector. The C-terminal linker (GSG) and His6 tag are retained on all proteins. The theoretical molecular weights and pIs for each scFv used in this study were also indicated. a. *de novo* designed scFv protein sequences. b. DNA sequences in FASTA format.

VH-Linker-VL-His(linker)-Histag-**stop**

**a.** *de novo* designed scFv protein sequences**.**

**>scFv-1**

QVHLVQSGGEVKKSGDSVKVSCKASGGGFSDRAVGWVRQAPGKGLEWMGGNEPEKKKGNNKDKFQGRVGVTADRDKGQAHMEVKRLKSEDTGVYYCSKFDSDQKKMEYWGKGATVTVTSGGGGSGGGGSGGGGSDIVMTQTPGTAAGRPGEKARISCRGSQTEKREDERSYLNWYLQKPGQSPQLLIYQGSRRASGVPDRFSGRGSGKEFNLGIRRTDREDEGRYFCKGSRKDRGFGGGAKVEIKGSGHHHHHH**stop**

Theoretical molecular weight: 26543.5 Daltons

Theoretical pI: 9.64

**>scFv-2**

EVRLEESGGGVKKPGDSLRLSCRGSGFKESRSDVSFVRQAPGKGLEWVGGSKSRRGGGDGNYARDVRGRVGATRDDSKNTSHLEFNGVKSDDTAVYFCTRSDKKERMDGWGKGTLVTVTSGGGGSGGGGSGGGGSSEVMTQSPGSASGRVGEKATISCRSSKKKGKEDGEGSMYWFLKKARPGSTLLISRGRNRFSGVPERFSGGGSGKDATLGIRRTEREDFGDYFCQQSEKEERFGSGAKVDRKGSGHHHHHH**stop**

Theoretical molecular weight: 27033.6 Daltons

Theoretical pI: 9.78

**>scFv-3**

EVRLEQSGGEVKKSSESMKITCKGSDGKEKQAWVGFVRQMPGKGLEWMGRIDPSDSYTNYSPSFQGHVTATADRDKGQAHLQVKSMKSSDTGMYYCSQDKEGFKRWGKGTTVTVTAGGGGSGGGGSGGGGSSDELTQSPGTAAGSVGERVTIFCKASGEGERKVSWYQQKAGKGPKVLVHGGDKRESGVSERFSGSGSGKEFTLGISKVEPEDEGRYYCGSYKSYPWTFGGGTKVEVKGSGHHHHHH**stop**

Theoretical molecular weight: 26205.8 Daltons

Theoretical pI: 9.01

**>scFv-4**

EVRLEESGGGVVQPRASMRLTCKASAHDSKRSKMGWIRQAPGKGLEWVSSIGEEGKDKKYADSRKGRFTISRDNDKRTLYLEMKNMDQDDTGAYFCCKHSDGMDTWGKGSMVTVSAGGGGSGGGGSGGGGSAIQMTQSPGSLSARVGEKVRITCKAGDKIRDKVAWNQQKRGQAPKVLVSRSDKRQSGVPSRFSGGGSGTDFTLTISSLEPGDFATYYCTDEKTHDRRFGGGARVERKGSGHHHHHH**stop**

Theoretical molecular weight: 26543.5 Daltons

Theoretical pI: 9.64

**>scFv-5**

EQRLVQSGAEVKKPGESLRISCKGSGEKGEGEWISWVRQMPGKGLEWMGRIDPSDKYTNYSPSFQGQVTIDKDRDKSTAYLQWSRLKSGDTGEYYCCRDPIYGERDWGKYWGQGTLVTVSAGGGGSGGGGSGGGGSDIVMTQTPLSAPVTPGEPASISCRSGEKGLDKREKGKLNWNQQKAGKSPQVVVYEVSNRATGVPDRFVGSGSGTDFTLKISRVDREDVGVYYCQQYSKEAKGVGPGTKVDIKGSGHHHHHH**stop**

Theoretical molecular weight: 27252.7 Daltons

Theoretical pI: 8.83

**>scFv-2D10**

MEIQLQQSGPELVKPGASVKISCKASGYSFTDYIMLWVKQSHGKSLEWIGNINPYYGSTSYNLKFKGKATLTVDKSSSTAYMQLNSLTSEDSAVYYCARKNYYGSSLDYWGQGTTLTVSSAKTTGGGGSGGGGSGGGGSDVVMTQTPFSLPVSLGDQASISCRSSQSLVHSNGNTYLHWYLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYFCSQSTHVPYTFGGGTKLEIKGSGHHHHHH**stop**

Theoretical molecular weight: 27893.0 Daltons

Theoretical pI: 9.00

b. scFv DNA sequences in FASTA format.

**>scFv-1**

CAGGTGCATCTGGTCCAAAGTGGCGGCGAAGTCAAAAAGTCAGGTGATAGCGTCAAAGTGAGCTGTAAAGCGAGTGGCGGCGGCTTCAGCGATCGTGCAGTTGGGTGGGTTCGGCAGGCACCCGGTAAGGGGCTGGAATGGATGGGGGGTAATGAGCCGGAGAAAAAAAAGGGGAACAACAAAGATAAATTCCAGGGACGCGTGGGGGTTACGGCCGACCGTGATAAGGGCCAGGCTCATATGGAAGTGAAACGTTTAAAAAGCGAAGATACGGGGGTGTACTATTGTAGTAAGTTCGATAGCGATCAGAAAAAAATGGAATACTGGGGCAAAGGCGCCACAGTAACAGTCACGAGTGGTGGTGGAGGCTCGGGTGGCGGTGGCAGCGGCGGCGGCGGCAGCGACATTGTCATGACCCAAACTCCGGGGACTGCGGCCGGGCGTCCGGGGGAAAAAGCTCGTATCTCGTGTCGCGGGTCCCAGACCGAAAAGCGTGAGGACGAACGGAGCTACTTAAATTGGTATTTACAAAAACCAGGTCAATCGCCGCAGCTGTTAATTTATCAAGGCAGTCGTCGTGCATCTGGTGTGCCCGACCGTTTCTCAGGACGTGGCAGTGGCAAGGAATTTAACCTGGGAATCCGTCGTACCGATCGTGAGGACGAGGGTCGCTATTTTTGCAAAGGCAGTCGCAAGGATCGTGGCTTTGGGGGTGGGGCGAAGGTCGAAATCAAAGGTAGCGGCCATCATCACCACCATCAT**TGA**

**>scFv-2**

GAAGTGCGCCTCGAGGAATCCGGCGGCGGTGTAAAAAAACCGGGTGATTCACTGCGCCTTAGTTGTCGGGGCTCGGGATTTAAAGAGTCTCGTTCGGATGTCAGTTTTGTGCGCCAAGCGCCTGGGAAGGGGCTCGAATGGGTTGGCGGTTCCAAATCACGCCGTGGCGGGGGTGATGGGAATTATGCGCGCGACGTACGTGGTCGGGTGGGTGCGACGCGCGACGATTCAAAAAACACGTCCCACCTCGAGTTCAACGGCGTTAAAAGTGATGATACCGCTGTATACTTCTGTACACGTTCAGATAAGAAAGAGCGCATGGATGGATGGGGTAAAGGCACGTTGGTTACGGTTACCTCGGGAGGAGGTGGTTCAGGTGGCGGCGGTTCAGGGGGGGGTGGCTCTAGTGAGGTTATGACCCAGTCCCCGGGTTCGGCCTCCGGCCGCGTCGGCGAAAAAGCGACCATTTCTTGTCGTTCTTCTAAGAAAAAAGGTAAAGAGGACGGTGAGGGTAGCATGTATTGGTTTCTGAAAAAGGCCCGTCCAGGATCGACGCTGCTGATTTCTCGGGGGCGCAATCGGTTTAGTGGTGTGCCGGAACGTTTCTCCGGGGGCGGTAGCGGTAAAGATGCCACGCTTGGCATCCGTCGTACAGAACGGGAAGACTTTGGTGATTATTTTTGTCAACAGTCAGAAAAAGAGGAACGCTTCGGCAGTGGCGCGAAAGTTGACCGTAAGGGCTCGGGCCACCACCACCACCATCAT**TGA**

**>scFv-3**

GAAGTTCGCCTGGAACAGAGCGGTGGCGAGGTTAAAAAATCATCCGAAAGCATGAAAATTACCTGTAAAGGGTCAGACGGTAAGGAAAAACAGGCCTGGGTTGGTTTCGTGCGGCAGATGCCAGGGAAGGGATTGGAGTGGATGGGTCGTATCGACCCAAGCGACTCCTATACGAACTATTCCCCATCGTTTCAGGGCCATGTGACCGCGACCGCTGATCGTGATAAGGGGCAAGCCCACCTGCAGGTGAAGTCAATGAAAAGCTCTGATACCGGCATGTACTATTGCTCGCAAGACAAGGAGGGTTTCAAACGTTGGGGTAAAGGAACAACAGTGACCGTTACCGCGGGCGGGGGCGGTAGTGGTGGCGGTGGCAGTGGTGGTGGAGGATCCTCTGATGAATTAACGCAGTCACCGGGAACTGCTGCCGGTAGCGTGGGTGAACGTGTTACCATTTTTTGCAAAGCCTCGGGTGAAGGAGAACGCAAAGTTAGCTGGTATCAACAGAAAGCCGGTAAAGGTCCCAAAGTGTTGGTGCATGGGGGTGATAAACGGGAATCTGGGGTATCAGAACGCTTTTCTGGCTCGGGCAGTGGCAAAGAATTTACCCTGGGTATTAGTAAAGTCGAACCGGAAGATGAAGGCCGCTATTATTGCGGAAGCTACAAATCATATCCGTGGACTTTCGGTGGCGGCACCAAGGTTGAAGTTAAAGGCAGCGGACATCACCACCATCATCAT**TGA**

**>scFv-4**

GAAGTTCGCCTGGAAGAAAGCGGAGGTGGGGTTGTACAACCGCGTGCCAGTATGCGTCTGACCTGTAAAGCCAGCGCGCACGATAGCAAACGCAGTAAAATGGGATGGATTCGGCAGGCTCCGGGGAAAGGTTTAGAATGGGTTTCATCCATCGGCGAAGAGGGCAAAGATAAAAAATATGCCGATAGCCGCAAAGGCCGTTTCACTATCAGCCGGGATAACGATAAACGTACTCTTTACCTGGAAATGAAAAATATGGACCAGGATGATACAGGCGCGTACTTCTGCTGCAAGCATTCAGATGGAATGGATACCTGGGGGAAAGGCAGCATGGTTACGGTTTCGGCGGGTGGCGGTGGGAGCGGTGGGGGAGGTAGTGGCGGTGGCGGGAGCGCGATCCAGATGACGCAGAGTCCAGGCAGCTTGTCTGCGCGGGTAGGTGAGAAAGTGCGCATCACCTGCAAAGCTGGTGACAAAATTCGTGACAAAGTAGCATGGAACCAACAAAAGCGTGGTCAAGCACCGAAAGTGCTGGTCAGCCGCAGTGACAAACGTCAGTCAGGCGTTCCGAGCCGGTTTTCCGGAGGCGGTAGCGGTACTGATTTTACATTGACTATTTCTTCACTGGAGCCGGGTGATTTTGCGACCTACTACTGCACCGACGAGAAAACCCACGATCGCCGTTTCGGCGGTGGCGCTCGTGTGGAACGGAAAGGTTCTGGCCACCACCATCACCATCAT**TGA**

**>scFv-5**

GAACAGCGCTTGGTACAGTCGGGGGCGGAAGTAAAGAAACCGGGTGAATCACTTCGCATTAGCTGCAAAGGCTCAGGCGAAAAGGGCGAGGGGGAGTGGATTAGCTGGGTTCGTCAAATGCCCGGTAAGGGTCTCGAATGGATGGGTCGCATTGACCCGAGCGATAAATACACGAACTACAGTCCTTCGTTCCAAGGTCAGGTGACCATCGATAAGGACCGTGATAAATCTACCGCATATTTACAATGGTCTCGCCTGAAGTCAGGCGACACTGGCGAGTACTACTGCTGCCGTGACCCAATTTATGGCGAGCGCGACTGGGGTAAATATTGGGGTCAGGGTACTCTGGTGACAGTGAGTGCAGGTGGAGGGGGCAGTGGAGGAGGAGGTAGCGGTGGCGGTGGTTCCGATATCGTAATGACACAAACCCCCTTATCCGCGCCCGTAACCCCTGGTGAACCGGCGAGCATCTCTTGCCGCTCCGGTGAAAAGGGTTTGGACAAGCGTGAAAAGGGAAAACTGAATTGGAATCAGCAGAAAGCGGGTAAAAGTCCGCAAGTCGTCGTGTACGAAGTCTCAAATCGTGCGACCGGCGTGCCTGATCGCTTCGTGGGTAGTGGCTCGGGAACGGACTTCACTTTAAAGATTAGTCGTGTAGATCGTGAAGATGTGGGCGTCTATTACTGTCAGCAGTATTCTAAGGAAGCTAAGGGCGTAGGTCCAGGCACGAAAGTTGATATTAAAGGTAGCGGCCACCACCACCATCATCAC**TGA**

**>scFv-2D10**

ATGGAAATCCAGTTACAGCAGTCTGGTCCGGAATTAGTGAAGCCGGGCGCGTCTGTTAAGATTAGCTGCAAAGCTTCCGGATATAGCTTTACCGATTACATCATGTTGTGGGTCAAACAATCTCATGGTAAAAGTTTGGAATGGATTGGCAACATTAATCCGTATTATGGCAGCACTAGCTACAATTTGAAATTCAAAGGAAAAGCCACCTTGACTGTTGACAAGAGCAGCTCAACCGCCTACATGCAACTTAACTCACTGACCTCTGAAGATTCCGCGGTCTACTACTGTGCGCGTAAGAACTACTATGGTAGCTCGTTGGATTACTGGGGCCAGGGCACTACTCTTACGGTATCCAGCGCCAAAACGACCGGAGGCGGGGGGAGTGGTGGTGGAGGCAGCGGTGGCGGCGGCTCCGATGTGGTCATGACGCAGACTCCGTTCTCACTGCCAGTTTCGCTTGGTGACCAGGCATCGATCTCTTGTCGCAGTAGTCAGAGTCTGGTTCATTCCAACGGAAACACGTACCTGCATTGGTATTTACAGAAACCTGGGCAGAGCCCCAAATTGCTGATTTACAAAGTTAGTAACCGCTTCTCTGGTGTTCCCGACCGCTTTTCGGGTTCCGGCAGCGGCACAGACTTCACATTGAAAATTTCGCGTGTTGAAGCCGAAGATCTGGGCGTTTATTTTTGCAGTCAATCCACCCATGTGCCGTATACGTTTGGTGGTGGAACCAAACTGGAGATTAAAGGTAGCGGTCATCACCACCATCATCAT**TGA**

**REFERENCES**

Bachmann A, Kiefhaber T, Boudko S, Engel J, Bachinger HP. 2005. Collagen triple-helix formation in all-trans chains proceeds by a nucleation/growth mechanism with a purely entropic barrier. Proc. Natl. Acad. Sci. USA 102(39):13897-13902.

Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. J. Chem. Theory Comput. 8(9):3257-3273.

Blanco-Toribio A, Lacadena J, Nunez-Prado N, Alvarez-Cienfuegos A, Villate M, Compte M, Sanz L, Blanco FJ, Alvarez-Vallina L. 2014. Efficient production of single-chain fragment variable-based N-terminal trimerbodies in Pichia pastoris. Microb. Cell Fact. 13:116.

Campana V, Zentilin L, Mirabile I, Kranjc A, Casanova P, Giacca M, Prusiner SB, Legname G, Zurzolo C. 2009. Development of antibody fragments for immunotherapy of prion diseases. Biochem. J. 418:507-515.

Chichili VPR, Kumar V, Sivaraman J. 2013. Linkers in the structural biology of protein-protein interactions. Protein Sci. 22(2):153-167.

Darden T, York D, Pedersen L. 1993. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J. Chem. Phys. 98(12):10089-10092.

Ding L, Azam M, Lin YH, Sheridan J, Wei SH, Gupta G, Singh RK, Pauling MH, Chu WH, Tran A and others. 2010. Generation of High-Affinity Fully Human Anti-Interleukin-8 Antibodies from its cDNA by Two-Hybrid Screening and Affinity Maturation in Yeast. Protein Sci. 19(10):1957-1966.

Ehrenmann F, Giudicelli V, Duroux P, Lefranc MP. 2011. IMGT/Collier de Perles: IMGT standardized representation of domains (IG, TR, and IgSF variable and constant domains, MH and MhSF groove domains). Cold Spring Harb. Protoc. 2011(6):726-36.

Fischer N, Elson G, Magistrelli G, Dheilly E, Fouque N, Laurendon A, Gueneau F, Ravn U, Depoisier JF, Moine V and others. 2015. Exploiting light chains for the scalable generation and platform purification of native human bispecific IgG. Nat. Commun. 6:6113.

Glaven RH, Anderson GP, Zabetakis D, Liu JL, Long NC, Goldman ER. 2012. Linking Single Domain Antibodies that Recognize Different Epitopes on the Same Target. Biosensors 2(1):43-56.

Goel M, Krishnan L, Kaur S, Kaur KJ, Salunke DM. 2004. Plasticity within the antigen-combining site may manifest as molecular mimicry in the humoral immune response. J. Immunol. 173(12):7358-7367.

Gomez CE, Lopez-Campistrous AE, Duarte CA. 1998. An immunoassay with bovine serum albumin coupled peptides for the improved detection of anti V3 antibodies in HIV-1 positive human sera. J. Virol. Methods 71(1):7-16.

Greenfield NJ. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1(6):2876-2890.

Hulme EC, Trevethick MA. 2010. Ligand binding assays at equilibrium: validation and interpretation. Brit. J. Pharmacol. 161(6):1219-1237.

Kiss G, Pande VS, Houk KN. 2013. Molecular dynamics simulations for the ranking, evaluation, and refinement of computationally designed proteins. Methods Enzymol. 523:145-70.

Krishnan L, Lomash S, Raj BP, Kaur KJ, Salunke DM. 2007. Paratope plasticity in diverse modes facilitates molecular mimicry in antibody response. J. Immunol. 178(12):7923-31.

Lee PS, Ohshima N, Stanfield RL, Yu WL, Iba Y, Okuno Y, Kurosawa Y, Wilson IA. 2014. Receptor mimicry by antibody F045-092 facilitates universal binding to the H3 subtype of influenza virus. Nat. Commun. 5:3614.

Lefranc MP. 2011a. IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. Cold Spring Harb. Protoc. 2011(6):643-51.

Lefranc MP. 2011b. IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. Cold Spring Harb. Protoc. 2011(6):633-42.

Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M and others. 2005. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. 29(3):185-203.

Li T, Pantazes RJ, Maranas CD. 2014. OptMAVEn--a new framework for the de novo design of antibody variable region models targeting specific antigen epitopes. PLoS One 9(8):e105954.

Li T, Verma D, Tracka MB, Casas-Finet J, Livesay DR, Jacobs DJ. 2013. Thermodynamic Stability and Flexibility Characteristics of Antibody Fragment Complexes. Protein Pept. Lett.

Mader A, Kunert R. 2012. Evaluation of the potency of the anti-idiotypic antibody Ab2/3H6 mimicking gp41 as an HIV-1 vaccine in a rabbit prime/boost study. PLoS One 7(6):e39063.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics 21(6):1087-1092.

Miller KD, Weaver-Feldhaus J, Gray SA, Siegel RW, Feldhaus MJ. 2005. Production, purification, and characterization of human scFv antibodies expressed in Saccharomyces cerevisiae, Pichia pastoris, and Escherichia coli. Protein Expr. Purif. 42(2):255-267.

Pantazes RJ, Grisewood MJ, Li T, Gifford NP, Maranas CD. 2015. The Iterative Protein Redesign and Optimization (IPRO) suite of programs. J. Comput. Chem. 36(4):251-63.

Pantazes RJ, Maranas CD. 2010. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. Protein Eng. Des. Sel. 23(11):849-58.

Pantazes RJ, Maranas CD. 2013. MAPs: a database of modular antibody parts for predicting tertiary structures and designing affinity matured antibodies. BMC Bioinformatics 14(1):168.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26(16):1781-802.

Pierce BG, Hourai Y, Weng Z. 2011. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One 6(9):e24657.

Prischi F, Konarev PV, Iannuzzi C, Pastore C, Adinolfi S, Martin SR, Svergun DI, Pastore A. 2010. Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. Nat. Commun. 1:95.

Rouet R, Lowe D, Christ D. 2014. Stability engineering of the human antibody repertoire. FEBS Lett. 588(2):269-77.

Song HN, Jang JH, Kim YW, Kim DH, Park SG, Lee MK, Paek SH, Woo EJ. 2014. Refolded scFv Antibody Fragment against Myoglobin Shows Rapid Reaction Kinetics. Int. J. Mol. Sci. 15(12):23658-23671.

Sreerama N, Woody RW. 2000. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal. Biochem. 287(2):252-260.

Sun H, Wu GM, Chen YY, Tian Y, Yue YH, Zhang GL. 2014. Expression, production, and renaturation of a functional single-chain variable antibody fragment (scFv) against human ICAM-1. Brazilian Journal of Medical and Biological Research 47(7):540-547.

Tang XC, Agnihothram SS, Jiao YJ, Stanhope J, Graham RL, Peterson EC, Avnir Y, Tallarico AS, Sheehan J, Zhu Q and others. 2014. Identification of human neutralizing antibodies against MERS-CoV and their role in virus adaptive evolution. Proc. Natl. Acad. Sci. USA 111(19):6863-6863.

Tapryal S, Gaur V, Kaur KJ, Salunke DM. 2013. Structural evaluation of a mimicry-recognizing paratope: plasticity in antigen-antibody interactions manifests in molecular mimicry. J. Immunol. 191(1):456-63.

Tapryal S, Krishnan L, Batra JK, Kaur KJ, Salunke DM. 2010. Cloning, expression and efficient refolding of carbohydrate-peptide mimicry recognizing single chain antibody 2D10. Protein Expr. Purif. 72(2):162-8.

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I and others. 2010. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. 31(4):671-90.

Weatherill EE, Cain KL, Heywood SP, Compson JE, Heads JT, Adams R, Humphreys DP. 2012. Towards a universal disulphide stabilised single chain Fv format: importance of interchain disulphide bond location and vL-vH orientation. Protein Eng. Des. Sel. 25(7):321-9.

Zahnd C, Spinelli S, Luginbuhl B, Amstutz P, Cambillau C, Pluckthun A. 2004. Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. J Biol Chem 279(18):18870-7.

Zarschler K, Witecy S, Kapplusch F, Foerster C, Stephan H. 2013. High-yield production of functional soluble single-domain antibodies in the cytoplasm of Escherichia coli. Microb. Cell Fact. 12.