CHEMBIOCHEM

Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2014

Cell Surface Display Yields Evolvable, Clickable Antibody Fragments

James A. Van Deventer,[a, b] Kai P. Yuet,[a] Tae Hyeon Yoo,[a, c] and David A. Tirrell*[a]

cbic_201402184_sm_miscellaneous_information.pdf

Materials and Methods

Materials. The plasmids pB18D and pB30D, which encode the cell surface-displayed 26-10 anti-digoxigenin single chain variable fragment (scFv) in pBAD18 and pBAD30 vector backbones, were generous gifts from Professor George Georgiou.[1] The plasmid pAK400, a vector used for periplasmic expression scFvs, was a generous gift from Professor Andreas Plückthun.[2] The plasmids pQE-80L and pREP4 were obtained from Qiagen (Valencia, CA). All DNA oligomers were purchased from Integrated DNA technologies (IDT, Coralville, Iowa), and all restriction enzymes were obtained from New England Biolabs (Ipswich, MA). DNA polymerases for cloning and for error prone PCR (Mutazyme II) were obtained from Stratagene/Agilent (Santa Clara, CA), and deoxynucleoside triphosphates were obtained from either Stratagene or Roche (Indianapolis, IN). The *E. coli* cell strains XL-1 Blue and DH10B were obtained from Qiagen and Invitrogen (Carlsbad, CA), respectively. The methionine auxotrophic *E. coli* strain TYJV2 was made in-house using the red recombinase gene knockout method of Datsenko and Wanner to eliminate the gene *metE* from *E. coli* strain DH10B.[3,4] Chemical reagents were purchased from Sigma-Aldrich (Madison, WI) unless otherwise noted. Canonical amino acids and L-norleucine (Nrl, **4**) were obtained from Sigma-Aldrich. L-Homopropargylglycine (Hpg, **2**) was purchased from Chiralix (Nijmegen, Netherlands). L-Azidohomoalanine (Aha, **3**) was synthesized by the method of Link et al.[5] with minor modifications. BODIPY FL Digoxigenin **5** used in flow cytometry experiments was a generous gift from Professor Patrick Daugherty (originally purchased from Invitrogen), and propidium iodide for viability staining was from Invitrogen. Streptavidin-phycoerythrin (SA-PE) was purchased from EBioSciences. Isopropyl-β-D-thiogalactoside (IPTG) was obtained from Gold Biosciences. Dibenzocyclooctyne Alexa Fluor 488 **6** was purchased from Invitrogen. Synthesis of biotin cyclooctyne **7**[6,7] has been described previously. Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) and 3-azidopropanol were prepared in-house by Janek Szychowski. THPTA was synthesized as described previously,[8] and 3-azidopropanol was prepared by azide displacement of bromine in 3-bromopropanol. The amine-reactive version of digoxigenin, 3-amino-3-deoxydigoxigenin hemisuccinamide, succinimidyl ester, was purchased from Invitrogen. Bovine serum albumin was from Equitech Bio (Kerrville, TX) or Pierce (Rockford, IL). Columns used in protein purification and reagents used in Biacore assays were obtained from GE Healthcare Life Sciences (Picataway, NJ). Nickel-nitrilotriacetic acid (Ni-NTA) resin was purchased from Qiagen. Dialysis membranes were from Spectrum Labs (Rancho Dominguez, CA). Nitrocellulose membrane was from GE Healthcare Life Sciences. Alexa Fluor 647-labeled anti-penta-His antibodies were purchased from Qiagen. Sequencing grade porcine trypsin was purchased from Promega (Madison, WI). Desalting col-

umns for purifying mass spectrometry samples were purchased from Millipore (Billerica, MA). Zeba spin desalting columns for small-scale buffer exchanges were from Pierce, and PD-10 desalting columns for intermediate volume buffer exchanges were from GE Healthcare Life Sciences.

Cloning and library construction. The plasmid pQE-80L-antidig-HisGS-Base was constructed in several steps. An *Eco*RI site in pB30D was eliminated via site-directed mutagenesis using the primer EcoRIElimFwd (sequence given in Table S10) and its reverse complement EcoRIElimRev. This resulted in the generation of plasmid pJAV2, which was confirmed by DNA sequencing. The Lpp-OmpA-scFv fusion was then amplified in two steps from pB18D. In the first step, an internal *Hind*III site was eliminated by performing two PCRs. The 5'-fragment was amplified using primers Lpp-OmpA-antidigFwd and HindIIIElimRev, and the 3'-fragment was amplified using HindIIIElimFwd and Lpp-OmpA-antidigRev1. After gel purification, a second amplification with the 5'- and 3'-fragments and primers Lpp-OmpA-antidigFwd and Lpp-OmpA-antidigRev2 was undertaken. The purified PCR product and pJAV2 were doubly digested with *Xma*I and *Hind*III and ligated. After transformation of electrocompetent XL-1 Blue cells and selection on plates containing chloramphenicol, clones containing the correctly sized insert were verified by DNA sequencing at Laragen (Culver City, CA). The resulting plasmid was designated pJAV2-antidig-RGSHis. This plasmid was used as a template for additional PCR amplification of the Lpp-OmpA-scFv fusion with Lpp-OmpA-antidigFwd and LppHisRescue. Double digestion of pQE-80L and the PCR product with *Hind*III and *Xma*I, purification, and ligation resulted in pQE-80L-antidig-HisGS, which was validated by DNA sequencing. Introduction of restriction sites into the Lpp-OmpA-scFv fusion gene at locations suitable for library construction was accomplished through assembly PCR. Four PCRs were performed in which pQE-80L-antidig-HisGS was amplified with the following pairs of primers: 80LLibFwd and PstIAddRev; PstIAddFwd and PstIElimRev; PstIElimFwd and BglIIAddRev; BglIIAddFwd and 80LLibRev. The resulting four PCR products were gel purified, mixed, and amplified using the primers 80LLibFwd and 80LLibRev. The PCR product was gel purified and doubly digested with *Hind*III and *Xma*I along with pQE-80L. The digestion products were gel purified, ligated, and transformed into electrocompetent *E. coli* XL-1 Blue cells, which were plated on ampicillin plates. Colonies were tested for the proper insert size and sequence verified for the desired restriction site modifications (*Pst*I and *Bgl*II), resulting in plasmid pQE-80L-antidig-HisGS-Base. The introduction of these restriction sites shortly after the Lpp portion of the construct and after the scFv sequence did not appear to affect the function of the display construct when compared to

pQE-80L-antidig-HisGS or pB18D. Use of pQE-80L-antidig-HisGS-Base and related display plasmids was always performed in cells containing pREP4 in order to maximize the amount of constitutively expressed *laqI*^{q} gene product, which is encoded in pQE-80L derivatives and in pREP4.

Libraries were constructed using pQE-80L-antidig-HisGS-Base (Library 1) or the thrice-sorted Library 1 after expression in medium containing Hpg (Library 2) as templates for error-prone PCR. The polymerase Mutazyme II from the Stratagene GeneMorph II kit was used to introduce errors throughout the Lpp-OmpA-scFv gene at targeted error rates of roughly two to five mutations per gene (Lpp and histidine tags excluded from mutagenesis). After an initial error-prone PCR step, additional amplification of DNA was performed using a higher fidelity polymerase. The insert and pQE-80L-antidig-HisGS-Base were doubly digested with *Pst*I and *Bgl*II. After gel purification and ligation, the resulting DNA was used to transform electrocompetent *E. coli* TYJV2 cells (~200 μL) containing pREP4. These cells were rescued for 1 hour in super optimized broth with catabolic repression (SOC; 20–25 mL), followed by inoculation into 2×YT medium (0.25–1.0 L) containing ampicillin (200 mg/L) and kanamycin (34 mg/L) (2×YT KA). Before growing cells, small amounts (~0.5–50 μL) of the broth were distributed on agar plates containing ampicillin and kanamycin in order to estimate the total number of transformants in the library. The libraries were grown at 37 °C until surpassing an OD_{600} of 1.0, and plates were incubated at 37 °C until colonies were large enough to count. Plasmid DNA was isolated from culture (~200 mL) using a Qiagen Maxiprep kit. Aliquots of cells containing the library were stored at a 1:1 ratio with cell stock buffer (65% glycerol (vol/vol), 25 mM Tris, 100 mM $MgSO_4$, pH 8.0) and frozen at −80 °C. In the construction of Library 1, two separate transformations were performed (1 and 1a), aliquoted, and maxiprepped. Colony counts indicated that transformation 1 yielded about 4.5×10^5 transformants, while transformation 1a yielded roughly 7.7×10^4 transformants. Random picking of ten clones revealed nine with the proper insert. Sequencing of these transformants using the primer 80LLibFwd revealed an approximate error rate of 2.7 mutations per kilobase, or 3.0 per gene. Construction of Lib2 resulted in approximately 5.6×10^6 transformants. Restriction fragment analysis of twelve clones showed that all twelve transformants selected had the proper inserts. Sequencing of ten clones using 80LLibRev revealed an approximate error rate of 4.2 mutations per kilobase, or 4.7 per gene, in Lib2 relative to Base (this includes mutations acquired during the first round of screening).

Miniprepped plasmid DNA isolated from sorted libraries was transformed into fresh TYJV2 cells containing pREP4; the transformed cells were grown on 2×YT plates supplemented with ampicillin and kanamycin. Individual colonies were randomly picked from plates, sub-

jected to medium shifts and flow cytometry experiments to determine binding function on-cell (see below), miniprepped, and sequenced. The scFv portions of DNA from the variants to be studied in soluble form were PCR amplified using primers AntidigpAK400Fwd and AntidigpAK400Rev with the exception of clone Nrl4x3, which contained a mutation within the priming region of AntidigpAK400Fwd and was therefore amplified in the forward direction with the primer AntidigpAK400FwdNrl4x3. After PCR amplification, all PCR products and the vector pAK400 were digested with *Sfi*I. Ligation of scFv genes and vector were performed using the New England Biolabs Quick Ligation Kit and transformed into electrocompetent DH10B cells. These cells were grown on agar plates at 30 °C containing chloramphenicol (35 mg/L) and glucose (1 % weight/vol). Resistant colonies were grown at 30 °C in liquid culture containing chloramphenicol and glucose, miniprepped, tested for the desired inserts, and verified by sequencing.

Expression of cell surface display constructs. Expression of all cell surface-displayed sequence variants and library populations to be studied via flow cytometry was performed using a standard medium shift procedure suitable for cell surface display.[9] On the day of expression, cells harboring the sequence variants or populations of interest were grown at 37 °C in M9 minimal medium (M9 salts containing glucose (0.4% w/v), thiamine hydrochloride (35 mg/L), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), and 20 amino acids (40 mg/L)) supplemented with ampicillin (200 mg/L), and kanamycin (35 mg/L) (M9 KA Glucose 20AA). Upon reaching an OD_{600} of roughly 0.5–1.0, cells were pelleted (7 minutes at 6000 × *g* in a fixed angle rotor or 10 minutes at 3000–4000 × *g* in a variable angle rotor), resuspended in minimal medium lacking Met, and allowed to grow at 37 °C for ten minutes. Cells were again pelleted and resuspended in minimal medium lacking Met and supplemented with Met or with the ncAA of interest. Cells were induced with IPTG (1 mM) and allowed to grow at 25 °C. After 6 hours, cells were pelleted and prepared for flow cytometry according to procedures described below.

Flow cytometry, Library 1. A single colony bearing plasmids pQE-80L-antidig-HisGS-Base (Base) and pREP4 was used to inoculate an overnight culture of minimal medium (2 mL) containing the twenty canonical amino acids, kanamycin, ampicillin, and glucose (M9 KA Glucose 20AA) and then grown overnight at 37 °C. The culture was then diluted 1:100 into fresh M9 KA Glucose 20AA (~10 mL). In the initial round of screening, equal volumes of transformations 1 and 1a were mixed, diluted 1:30 into M9 KA Glucose 20AA (60 mL), and grown at 37 °C. After

reaching OD_{600} = 0.7–0.8, cells were shifted into new medium as described above. Cells bearing pQE-80L-Antidig-HisGS-Base were aliquoted (3–4 mL/aliquot) and induced with IPTG (1 mM) in M9 KA Glucose containing 20 amino acids or 19 amino acids (−Met) supplemented with Hpg (2.5 mM). Cells bearing Library 1 were aliquoted into several small samples and induced in medium containing twenty canonical amino acids or 19 amino acids (−Met) with Hpg (2.5 mM) supplemented with Met (0, 10, 20, or 30 μM). After 6 hours, cells were pelleted, washed once in phosphate-buffered saline (PBS, pH 7.4: 8.00 g/L NaCl, 0.20 g/L KCl, 1.15 g/L $Na_2HPO_4 \bullet H_2O$, 0.20 g/L KH_2PO_4), and resuspended in PBS to an OD_{600} of 1.0. Cells were then treated with **5** (200 nM) in aliquots (0.25–1.0 mL) for at least 45 minutes with gentle agitation at room temperature, diluted 1:20 into PBS, and filtered using filters (25 mm diameter) containing Acrodisc Supor membranes (5 μm; Pall Life Sciences, Ann Arbor, Michigan).

A MoFlo flow cytometer (Beckman Coulter, Miami, FL) was used for all scanning and library sorting. Control experiments confirmed that expression of Base in media supplemented with Met yielded cells exhibiting high fluorescence after exposure to **5**, while cells expressed in Hpg (2.5 mM) exhibited low fluorescence levels after exposure to **5**, confirming that the medium shift was successful. Library 1 expressed in Hpg (2.5 mM) and Met (20 μM) was judged to have a sufficient number of positive events to make it suitable for sorting. A gate was set to collect the brightest ~0.1 % of the events, and sorting was allowed to proceed for a total of 35 million events in the sort mode Single 1. Approximately 36,000 events satisfied the gating criteria and were deposited directly into a tube of SOC medium on ice. Upon completion of the sort, cells were rescued for 1 hour at 37 °C in SOC (2.5 mL), diluted with 2×YT medium (10.5 mL) containing ampicillin and kanamycin, and allowed to grow overnight at 37 °C. The next day, aliquots of the sorted population were mixed 1:1 with cell stock buffer and frozen at −80 °C. The remaining rescued cells were miniprepped in order to isolate plasmid DNA from the sort. Subsequent rounds of expression, scanning, and sorting were performed as described above, with a frozen aliquot from the previous round of sorting used as the input for the next round of expression and cell sorting. In rounds 2 and 3, the population to be sorted was expressed in M9 KA Glucose containing nineteen amino acids plus Hpg (2.5 mM; no Met). Sorting was performed with comparable stringencies and event totals as in round 1.

Miniprepped DNA from the third round of sorting was used to transform electrocompetent TYJV2 cells containing pREP4. Cells were rescued for 1 hour with SOC and plated on agar containing ampicillin and kanamycin. Ten colonies were chosen randomly for sequencing and characterization. Following inoculation into M9 KA Glucose 20AA and overnight growth at 37 °C, cells were diluted 1:100 into fresh M9 KA Glucose 20AA and 1:200 into fresh 2×YT

supplemented with kanamycin and ampicillin (2×YT KA). Cells in minimal media were then subjected to medium shifts, induced with IPTG (1 mM) in Hpg (2.5 mM) for 6 hours, and harvested. Cells were treated with **5** (200 nM) and subjected to flow cytometry to assess binding in the Hpg amino acid context. Cells grown in 2×YT KA were miniprepped for plasmid DNA and submitted for sequencing with the primers 80LLibFwd and 80LLibRev.

Flow cytometry, Library 2. Rounds one and two of sorting of Library 2 in various amino acid contexts were performed essentially as described for sorting Library 1. Aliquots of the library were thawed and diluted 1:30 into fresh M9 KA Glucose 20AA, grown, shifted into fresh medium, and subjected to 6-hour induction of protein expression in medium containing either Met (0.27 mM), Hpg (2.5 mM), Aha (0.28 mM), or Nrl (0.30 mM). In round 1, cells were sorted in Purify 1 mode after treatment with **5** (200 nM), dilution, and filtering. A total set of events numbering more than ten times the library size was screened in each amino acid context, with 0.2%–0.5% of the most fluorescent events retained and rescued as described above. In round 2, the populations were treated with a reduced concentration (100 nM) of **5** prior to screening, and 0.1%–0.2% of the most fluorescent events were retained and rescued.

Beginning with the third round of sorting, significant changes were made to the expression and sorting protocol. Frozen aliquots of library fractions were thawed and diluted 1:100 into fresh M9 KA Glucose 20AA and allowed to grow to saturation overnight at 37 °C, then diluted 1:50 the next morning into fresh M9 KA Glucose 20AA and grown, medium shifted, and induced for 6 hours with IPTG (1 mM). Cells were washed once in PBS, and library fractions of cells induced in the presence of ncAAs were resuspended at an OD_{600} of 0.2. The library fraction Hpg2x expressed in Hpg was sorted on the same day it was induced. The library fraction was treated with **5** (100 nM), with control samples (i.e., library fractions expressed in Met and Base in Met and Hpg) resuspended and treated with **5** (100 nM) at an OD_{600} of 1.0. All other populations to be sorted (i.e., Aha2x, Aha3x, Hpg3x, Nrl2x, and Nrl3x) were incubated with gentle agitation overnight at 4 °C in PBS along with control samples (at the OD_{600}s listed above).

Washing, viability staining, and kinetic competitions were used with all samples to be sorted in rounds three and four of sorting in an attempt to enrich sorted populations for clones with improved kinetic properties and to exclude dead cells. Samples to be sorted were treated with **5** (100 nM) for at least 45 minutes before being washed twice in PBS, diluted, filtered, and scanned. Cells were also treated with propidium iodide (PI; 5 μM) at least five minutes before

being scanned on the flow cytometer. Prior to sorting library fractions, the kinetic competition method of Daugherty et al. was employed.[10] Two kinetic competitions were run in parallel: cells to be monitored as a function of time were washed twice after treatment with **5**, diluted to an OD_{600} of 0.05, filtered, and exposed to PI prior to competition. The population to be sorted was washed twice after treatment with **5** and diluted to an OD_{600} of 0.5 without exposure to PI prior to competition. The cells diluted to OD_{600} = 0.05 were scanned on the flow cytometer, and then competition was initiated in both samples by adding unlabeled digoxin (2 μM). Total competition duration was determined empirically by monitoring the competition progress with scans of the more dilute cells every 5 minutes. When the mean fluorescence of the scanned population had been reduced by roughly 40%–60%, the competition of the cells at OD_{600} = 0.5 was stopped by washing the cells twice in ice-cold PBS. Cells were then diluted to an OD_{600} of 0.05, filtered, treated with PI, and sorted on the MoFlo using sort mode Single 1. Approximately 0.1%–1% of events were retained in each sort, which was gated to isolate cells exhibiting high levels of fluorescence after kinetic competition while excluding dead cells, doublets, and other aberrant events. Events totaling at least ten times the size of the sorted population (≤ 30,000 distinct clones, based on the number of events retained in previous sorts) were scanned in the course of a sort. Cells were retained and rescued as above.

On-cell estimates of dissociation constants. Individual clones from sorted library populations were obtained by transforming electrocompetent TYJV2 cells containing pREP4 with plasmid DNA samples from sorted populations Hpg3x, Aha3x, Hpg4x, Aha4x, and Nrl4x. Transformed cells were then allowed to grow on agar plates supplemented with ampicillin and kanamycin. Individual colonies were picked and used to inoculate liquid cultures in M9 KA Glucose 20AA and grown overnight at 37 °C. The next morning, samples were diluted 1:50 into fresh M9 KA Glucose 20AA and grown at 37 °C in preparation for medium shifts. During the medium shifts performed on cells bearing individual mutants, a portion of each culture was diluted into fresh 2×YT KA and allowed to grow to saturation. Plasmid DNA from the 2×YT cultures was extracted and subsequently sequenced to determine the mutations present in each clone. Another portion of cells was shifted into fresh medium, and protein expression was induced in the ncAA context in which the clones were isolated. After expression, all cells were washed once and resuspended in PBS at an OD_{600} of 1.0 and kept at 4 °C overnight with gentle agitation. The next day, samples were treated with **5** (100 nM) for at least 45 minutes at an OD_{600} of 0.2 and stored on ice. While on ice, all samples to be used in kinetic dissociation experiments[11] were treated with PI. Prior to scanning, samples that had been induced in the

presence of Met, Hpg, or Aha were washed twice and then resuspended in PBS, again containing PI. For samples induced in the presence of Nrl, cells were pelleted, but not washed, prior to resuspension in PBS containing PI due to rapid loss of fluorescence in these samples. An initial scan of each clonal population was obtained prior to initiation of competition. Samples were then treated with unlabeled digoxin competitor (1 μM), mixed, and scanned every 60 seconds over a period of 10 minutes.

Data from timed scans were analyzed using FlowJo (Tree Star, Ashland, OR) and Igor (Wavemetrics, Lake Oswego, OR) according to a previously described approach.[11] All sample scans were gated to exclude PI-positive cells from the analysis. The mean fluorescence of all PI-negative cells was then calculated. To eliminate background fluorescence from all samples, the mean fluorescence of a sample of cells containing pQE-80L-Antidig-HisGS-Base induced in the presence of Hpg (obtained on the same day as all data for kinetic dissociation experiments and gated to eliminate PI-positive cells) was subtracted from the fluorescence of all other samples. The background-corrected data for an entire competition were then normalized based on the brightness of the sample of cells involved in the competition at t = 0 minutes. The resulting background-corrected, normalized data were fit to a first-order exponential equation of the form

$$F = e^{-k_{off} t}, \qquad \text{(S1)}$$

where F is the relative fluorescence of the sample compared to the sample at time t = 0 and k_{off} is the dissociation rate constant. Results reported for all clones discussed here were performed once and served as a means for identifying individual clones to study in more detail. On each day these experiments were performed, the rate constant of either Base induced in Met or the protein product of pQE-80L-Antidig-HisGS containing Met was estimated as a control for day-to-day consistency of technique. Data are reported in groups of samples that were all assayed at the same time. All data were analyzed using the program Igor (Wavemetrics, Lake Oswego, OR), and the errors reported are the 95% confidence intervals of the fits to Equation S1.

Flow cytometry, simultaneous investigation of binding and chemical modification. TYJV2 cells containing a plasmid encoding cell surface-displayed Aha4x5 and the plasmid pREP4, and TYJV2 cells lacking plasmids were grown in M9 KA Glucose 20AA (cells containing plasmids) or M9 Glucose 20AA (cells without plasmids) as described above and induced with IPTG (1 mM) after using the standard cell surface display medium shift procedure, with

induction performed in medium containing either Met (0.27 mM) or Aha (1 mM). Samples left uninduced (TYJV2 cells with or without plasmids) were also subjected to medium shifts and expression conditions, without the addition of IPTG. After the 6-hour expressions, cells were washed twice in ice-cold PBS and resuspended to an OD_{600} of 1.0. Samples of cells were treated with biotin cyclooctyne **7** (100 μM) and incubated at 37 °C for 16 hours.[7] Additional portions of cultures were also incubated for 16 hours at 37 °C without exposure to **7**. All samples were pelleted and resuspended, with cells treated with **7** washed twice in PBS prior to resuspension. Fluorescent probes were then added to probe the binding and modification states of cellular populations. All samples were treated with **5** (100 nM), a 1:100 dilution of streptavidin-phycoerythrin (SA-PE; 0.2 mg/mL), or both dyes simultaneously for a period of 45 minutes or longer. Cells displaying the Aha form of Aha4x5 were also treated with digoxin (10 μM) prior to exposure to **5** and SA-PE. After dye exposure, all samples were washed twice in PBS, diluted by at least tenfold, and filtered.

Samples were scanned using a MoFlo XDP flow cytometer (Beckman Coulter) upgraded from MoFlo to MoFlo XDP status by Propel Labs (Fort Collins, CO). Manual compensation was performed to enable simultaneous detection of BODIPY and phycoerythrin fluorescence without cross talk between fluorescence channels. Antigen binding and chemical modification were examined, first in separate experiments and then in a double-labeling experiment (Figure S6). A model sorting experiment was set up by combining four populations of cells bearing the plasmid encoding for Aha4x5: uninduced cells grown in Aha, induced cells expressed in Met, induced cells expressed in Aha blocked with unlabeled digoxin, and induced cells expressed in Aha. Each population was treated with **7** and exposed to **5** and SA-PE. Elliptical regions were established in the two-dimensional dot plot of BODIPY versus phycoerythrin fluorescence to capture populations of cells with various combinations of fluorescence properties. A four-way sort performed in Purify mode enabled isolation of cells with all possible combinations of high and low BODIPY and phycoerythrin fluorescence. The fluorescence properties of the isolated populations were examined by running each population on the flow cytometer immediately after completion of the sort.

Expression and purification of soluble scFvs. TYJV2 colonies bearing individual scFv sequence variants in the pAK400 backbone were inoculated into M9 minimal medium (2 mL) containing twenty amino acids and glucose (0.4% weight/vol) supplemented with chloramphenicol (M9 Cam Glucose 20AA). Cells were allowed to grow at 30 °C for 4 or more hours,

diluted 1:20 into fresh M9 Cam Glucose 20AA (20–40 mL) and grown at 30 °C overnight. Saturated cultures were then diluted 1:20 into M9 Cam Glucose 20AA (0.25–1.0 L) and grown at 30 °C until reaching an OD_{600} of approximately 0.9–1.0. Cells were pelleted (15 minutes at 5000 × *g*, 4 °C), washed three times in ice-cold NaCl (0.9% weight/vol), and resuspended in fresh M9 medium containing 19 amino acids (minus Met), glycerol (0.4% weight/vol), and chloramphenicol (M9 Cam Glycerol 19AA). Aliquots were supplemented with Met (0.27 mM), Hpg (1 mM), Aha (1 mM), or Nrl (2.3 mM) and grown at 25 °C for 30 minutes. Cultures were then induced with IPTG (1 mM) and allowed to grow for 4 hours at 25 °C.

Upon completion of expression, cells were pelleted (15 minutes at 5000 × *g*, 4 °C) and resuspended vigorously in ice-cold Tris/sucrose buffer (0.75 M sucrose, 0.1 M Tris, pH 8.0; 32 mL per liter culture volume). All subsequent scFv purification steps were performed at 4 °C or on ice unless otherwise noted. An adaptation of a previously described osmotic shock procedure[12] was used to isolate the periplasmic fractions of cells, which contain the expressed scFvs. After resuspension, the Tris/sucrose buffer was supplemented with ice-cold lysozyme (10 mg/mL) in Tris/sucrose buffer (3.2 mL per liter culture volume). Pellets were rotated at 140 RPM on ice during dropwise addition of ethylenediaminetetraacetic acid (EDTA; 1 mM; 64 mL per liter culture volume). After ten minutes (still with rotation), magnesium chloride (0.5 M; 3 mL per liter culture volume) was added dropwise to the cell suspension. After another ten-minute incubation period, the samples were pelleted (20 minutes at 10,000 × *g*, 4 °C). The supernatants were dialyzed overnight against a solution of Ni-NTA start buffer (0.3 M NaCl, 0.05 M NaH_2PO_4, 0.01 M imidazole, pH 8.0) at 4 °C using 12–14 kilodalton molecular weight cutoff (MWCO) dialysis tubing. Dialyzed samples were filtered using Acrodisc Supor filters (25 mm diameter with 5 μm pores) and incubated for 2–4 hours with Ni-NTA agarose prewashed with Ni-NTA start buffer (4.0 mL per liter culture volume resuspended slurry) at 4 °C. The Ni-NTA slurries were added to columns and the flow-through was collected. After two washes with Ni-NTA wash buffer (0.3 M NaCl, 0.05 M NaH_2PO_4, 0.02 M imidazole, pH 8.0; 20 mL per liter culture volume), samples were eluted with Ni-NTA elution buffer (0.3 M NaCl, 0.05 M NaH_2PO_4, 0.25 M imidazole, pH 8.0; 4 × 4.0 mL per liter culture volume). The presence of protein in the eluent was confirmed by spotting small portions of each elution fraction onto filter paper, staining the paper with coomassie blue stain (2.5 g/L coomassie blue and 2.5 g/L cupric sulfate in a solution of 5:4:1 water:ethanol:acetic acid) and destaining using 5:4:1 water:ethanol:acetic acid. Fractions found to contain large amounts of protein were pooled and dialyzed against ion exchange start buffer (125 mM NaCl, 20 mM Tris, pH 8.0), saving aside small amounts of eluent for sodium docedecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

Ion exchange and size exclusion chromatography were performed on dialyzed eluents from Ni-NTA purification using an AktaPrime Plus fast performance liquid chromatography (FPLC, GE Healthcare) system at 4 °C. A HiTrapQ XL ion exchange column (1.0 mL bed volume) was used according to the manufacturer's instructions with ion exchange start buffer and ion exchange high salt buffer (1 M NaCl, 20 mM Tris pH 8.0). The majority of scFv protein samples eluted during initial injection onto the column, while impurities tended to bind to the column until the salt concentration on the column was increased. The column flow-through containing scFv was collected, pooled, and concentrated (to a volume of 1.0–1.5 mL) using Amicon Ultra-15 10 kilodalton MWCO concentration devices primed with HBS +EDTA (150 mM NaCl, 10 mM HEPES, 3 mM EDTA, pH 7.4). Size exclusion chromatography was performed using a HiPrep 16/60 Sephacryl S-100 HR column using HBS +EDTA at a flow rate of 1 mL/minute. This column allowed resolution of dimer and monomer scFv peaks. Fractions of monomeric protein were collected, pooled, and concentrated using Amicon Ultra-15 concentrators.

Protein characterization. All purification processes were assessed by SDS-PAGE analysis. After Ni-NTA, ion exchange, and size exclusion chromatography, all samples were found to be more than 90 % pure as judged by quantification of band intensities after colloidal blue staining (Invitrogen). Gels were imaged on a Typhoon Trio (GE Healthcare, Piscataway, NJ) imager, and the resulting images were processed using ImageQuant software (GE Healthcare). Concentrations of protein samples were determined using bicinchoninic acid (BCA) assay kits from Pierce. Proteins were stored in HBS +EDTA buffer at 4 °C long term. Size exclusion chromatography on monomeric fractions of select proteins after storage for several months revealed no evidence of dimerization or aggregation.

Matrix-assisted laser desorption ionization (MALDI) mass spectrometry was performed on trypsinized samples in order to assess amino acid replacement levels. All samples to be examined were buffer exchanged into denaturing buffer (8 M urea, 0.1 M Tris, pH 8.0) using Amicon Ultra-0.5 mL concentration devices. Samples were reduced by treatment with tris(2-carboxyethyl)phosphine (TCEP) (3.75 mM) at room temperature for 15 minutes. In the case of azide-containing samples, some samples were reduced with dithiothreitol (DTT; 2.5 mM) at 55−60 °C for 15 minutes. Reduced samples were alkylated by the addition of iodoacetamide (12.5 mM) and incubation at room temperature in the dark for 15 minutes. Following alkylation, ammonium bicarbonate in water (50 mM, pH 7.8; 10 volumes) was added to each solution along with trypsin (Promega, 0.1 μg/μL stock concentration; 2.5–5 μL/sample); cleavage was

allowed to proceed overnight at 37 °C. Trifluoroacetic acid (TFA) was added to each cleaved sample (0.1 % final concentration), and C_{18} Zip Tips (Millipore) were used to desalt peptide samples according to the manufacturer's protocol with one slight modification. In between wetting and equilibrating the columns, a 50/50 acetonitrile/0.1% TFA solution was used to wash the columns. All samples were then mixed with α-cyano-4-hydroxycinnamic acid, dried on a MALDI target, and assayed on a Voyager DE Pro (Applied Biosystems, Carlsbad, CA) at the Caltech Division of Chemistry and Chemical Engineering Mass Spectrometry Facility. All observed peptide masses fell within instrument tolerances of mass accuracies. Some Aha residues were converted to diaminobutyric acid residues during acquisition of MALDI mass spectra.[13]

Incorporation levels of ncAAs were determined using processed MALDI mass spectra. All spectra were baselined and deisotoped using Data Explorer software (Applied Biosystems). The fraction of peptides bearing the ncAA substitution was calculated based on the counts of peaks corresponding to substituted and unsubstituted peptides. In most cases, the deisotoping procedure resulted in the presence of single peaks at the substituted and unsubstituted mass positions. In cases where the deisotoping still resulted in multiple peaks, the counts of all peaks corresponding to unsubstituted and substituted peptides were summed prior to calculating the fraction of substituted peaks.

MALDI mass spectrometry was also used on trypsinized and whole-protein samples in order to determine modification sites after click chemistry (see below). Trypsinized samples of modified proteins were prepared as described above. Whole-protein samples were also buffer exchanged into a denaturing buffer (8 M urea, 0.1 M Tris, pH 8.0). To each solution was added ammonium bicarbonate in water (50 mM, pH 7.8; 10 volumes), followed by addition of TFA (0.1 % final concentration). Protein samples were then desalted using C_4 Zip Tips and further prepared for MALDI as described above.

Protein modification. All strain-promoted click chemistry on soluble scFvs was performed using Alexa Fluor 488 dibenzocyclooctyne (DIBO) (**6**, Invitrogen). Compound **6** (10 μM) was added to solutions of Aha- or Met-containing scFvs (1.75 μM) in HBS +EDTA buffer. Solutions were vortexed briefly, and reaction was allowed to proceed for 1 hour at room temperature. Reactions were quenched by adding 3-azidopropanol (10–20 mM) and vortexing.

All dye-labeled proteins were run on denaturing SDS-PAGE gels in order to assess dye labeling. After electrophoretic separation, gels were destained in a solution of 50% methanol,

40% water, and 10% acetic acidic and rinsed in water. The Typhoon Trio was used to detect fluorescent bands on the gels. Gels were then stained overnight in colloidal blue staining solution, washed with water, and imaged again on the Typhoon Trio. Gel images were processed using ImageQuant software to assess the relative amounts of fluorescent dyes attached to protein samples of various types. Quantification of the extent of reaction between azide-containing proteins and compound **6** was performed by using Alexa Fluor 488-labeled streptavidin as an in-gel fluorescence standard. Absorbance measurements on a Cary 50 ultraviolet-visible spectrophotometer (Agilent) were used to obtain protein concentrations and the extent of labeling of the concentrated fluorescent streptavidin standard. Known concentrations of standard and known protein quantities of labeled samples were run on SDS-PAGE gels and imaged as above. Establishment of a standard curve enabled estimation of the quantity of fluorescent dye present in the bands of each experimental sample, and knowledge of the protein concentration enabled the number of dye molecules per protein to be calculated. Results reported here are the averages of three triplicate experiments. In the case of the Aha form of Aha4x5, protein expressed and purified in two separate batches was used to estimate the extent of dye labeling. All other proteins assessed for extent of modification were from single batches of purified proteins.

Western blotting. Digoxigenin-labeled bovine serum albumin (BSA) was prepared by conjugating amine-reactive 3-amino-3-deoxydigoxigenin hemisuccinamide, succinimidyl ester (Invitrogen) to BSA using reaction conditions described by the manufacturer. Briefly, BSA (10.2 mg) was dissolved in sodium bicarbonate (0.1 M, pH 9.0; 1 mL) and stirred in a scintillation vial. A small amount (≤ 1 mg) of 3-amino-3-deoxydigoxigenin hemisuccinamide, succinimidyl ester dissolved in *N,N*-dimethylformamide (150 μL) was added dropwise to the BSA solution. The mixture was stirred at room temperature for 5 minutes, diluted using PBS (final volume of 2.5 mL) and run on a preequilibrated PD-10 desalting column in order to isolate the BSA from the unreacted labeling agent and to perform a buffer exchange into PBS.

Detection of digoxigenin-labeled BSA (BSA-Dig) in protein samples was achieved using standard Western blotting procedures. Samples of BSA and BSA-Dig mixed with *E. coli* lysates (~5.5 μg/lane) were run on SDS-PAGE gels and transferred to nitrocellulose membranes. Membranes were blocked for at least 1 hour in PBS containing TWEEN-20 (0.1%) and either BSA (3% weight/vol) or powdered milk solution (5% weight/vol) (Nestle, Glendale, CA). Membranes were washed in PBS containing TWEEN-20 (0.1%; PBS-TWEEN) before expo-

sure to labeled scFvs. Antibody fragment samples (100–250 μL, 1.75 μM protein) treated with **6** were diluted into PBS-TWEEN (approximately 10 mL) containing either BSA (3% weight/vol) or powdered milk (5% weight/vol) and exposed to membranes for approximately 1 hour. Membranes were washed several times with PBS-TWEEN and imaged using the Typhoon Trio. In some cases, membranes were placed back into PBS-TWEEN containing milk (5% weight/vol) and exposed to Alexa Fluor 647-labeled anti-penta-His antibodies added to solution at a 1:10,000 dilution for 1 hour, washed, and imaged. These experiments enabled secondary detection of digoxigenin using the hexahistidine tag of the scFvs and served to probe the binding function of scFvs regardless of whether the antibody fragments in question were labeled with fluorescent dye.

Kinetic characterization of soluble scFvs. A Biacore T100 instrument was used to determine the binding kinetics of scFvs using surface plasmon resonance. All assays were performed on CM5 chips to which BSA and BSA-Dig were immobilized using standard amine coupling procedures for the T100 processing unit. Flow cells to which BSA was immobilized served as a means to double reference all data obtained on flow cells containing immobilized BSA-Dig. Two separate CM5 chips containing immobilized BSA and BSA-Dig were prepared and used to assay the binding kinetics of scFvs. One chip contained 92.9 response units of BSA immobilized to one flow cell and 12.7 response units of BSA-Dig immobilized to a second flow cell. The other chip contained one flow cell to which 39.7 response units of BSA was bound and three flow cells to which BSA-Dig was bound at response unit levels of 10.6, 5.4, and 45.9. Standard multicycle kinetics assays were performed at 25 °C on all scFvs of interest, including dye-labeled samples, using HBS EP+ (HBS +EDTA with 0.005 % surfactant P20) as a running buffer at a 90 μL/min flow rate. Surface regeneration was achieved by exposing the chip to glycine-HCl (10 mM glycine-HCl, pH 2.0) for 60 seconds (90 μL/min flow rate). scFv samples were injected onto the chip in concentrations ranging from 0.3125 to 40 nM in twofold increments (i.e., 0.3125, 0.625, 1.25, 2.5, 5, 10, 20, and 40 nM concentrations). Blank samples (no scFv) were also run, and each sample/blank was run twice during the course of an assay. Injection and dissociation times depended on the amino acids present in the scFv sample in question. scFvs containing Hpg were injected for a period of 180 seconds, while scFvs containing all other amino acids were injected for 120 seconds. The dissociation of scFvs containing Aha from the chip surface was monitored for 650 seconds, while all other scFv dissociations were monitored for 450 seconds. All scFv samples (including those samples treated with **6**) were diluted to the final concentrations need for kinetic assays using HBS EP+, without di-

alysis or other buffer exchange into the Biacore running buffer. Before assaying experimental samples, and in between different scFv samples, the chip was reconditioned using at least five startup cycles.

Kinetic constants for all scFvs were determined using Biacore T100 Evaluation software. All binding curves were subjected to software-provided double referencing procedures before performing data fits using the standard 1:1 binding kinetic transport model provided with the software.[14] Data fits usually included all data up to and including 20 nM injections, excluding samples in which air bubbles or other deviations marred the sensorgrams. Most fitting could be performed without invoking significant mass transport limitations (mass transport constant $k_t \geq 10^{10}$). However, all data taken on Hpg-containing fragments showed evidence of substantial mass transport limitation ($k_t \sim 10^6$–10^8), meaning that the values for k_{on} and k_{off} reported here for these scFvs may be slightly different from values determined in the absence of significant mass transport limitations.[15] All data reported here are the averages of independent experiments performed on the two chips prepared as described above, with a total of four determinations of kinetic parameters through data fitting on each chip surface containing immobilized BSA-Dig.

High-throughput sequencing. Frozen aliquots of Library 1 sorted three times after expression in media containing Hpg, Library 2 Hpg4x, Library 2 Aha4x, and Library 2 Nrl4x were thawed, diluted 1:200 into 2×YT KA (200 mL), and grown at 37 °C until the OD_{600} of each culture exceeded 1.0. The cultures were pelleted (15 minutes at 5000 × *g*, 4 °C), decanted, and frozen at −20 °C for at least 12 hours. Plasmid DNA from each sorted population was isolated using a Maxiprep kit. All DNA was digested using restriction enzymes *Bgl*II and *Pst*I, and fragments containing the Lpp-OmpA-scFv fusion gene (excluding Lpp and His tag portions of the gene) were separated from other DNA fragments via gel electrophoresis. DNA of each library population (approximately 1–3 μg) was submitted to the Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology for fragmentation and high-throughput sequencing using the Illumina Genome Analyzer IIx platform (Illumina, San Diego, CA). DNA samples were subjected to fragmentation and cluster generation according to the manufacturer's recommendations, and 38-base sequencing runs were performed on each sample (one lane/sample). Each run yielded over 3×10^7 sequencing reads on an estimated 3.5–4.2×10^5 clusters. All data were aligned to the DNA sequence of pQE-80L-Antidig-HisGS-Base, and the total number of calls of A, C, G, T, and N, at every position of each population was tallied using a script developed at the genomics laboratory. Every position within the con-

struct was read at least 8,500 times in all populations. Substantial variation in the total numbers of reads at each position was observed (ranging from approximately 8,500 to more than 1 million), most likely due to nonuniform fragmentation of the restriction fragment. However, even the minimum coverage of 8,500 yields good sampling of the populations to be characterized, which are the result of repeated isolations of no more than 20,000–30,000 sequence variants via flow cytometry. All sequencing data were searched for positions at which 5% or more of the total reads indicated mutations from the nucleotide present in the base construct in a given population using scripts developed in Matlab (only single nucleotide mutations were considered in this analysis). Data are presented here on all positions at which at least one population displayed 5% deviation from pQE-80L-Antidig-HisGS-Base with the additional criterion that the base call quality (Q, as defined by the Illumina analysis software) of at least one nucleotide (A, C, G, or T) at the position of interest exceeded 30.

References

[1] P. S. Daugherty, M. J. Olsen, B. L. Iverson, G. Georgiou, *Protein Eng.* **1999**, *12*, 613–621.

[2] A. Krebber, S. Bornhauser, J. Burmester, A. Honegger, J. Willuda, H. R. Bosshard, A. Plückthun, *J. Immunol. Methods* **1997**, *201*, 35–55.

[3] K. A. Datsenko, B. L. Wanner, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6640–6645.

[4] T. H. Yoo, D. A. Tirrell, *Angew. Chem.* 2007, *119*, 5436–5439; *Angew. Chem. Int. Ed.* **2007**, *46*, 5340–5343.

[5] A. J. Link, M. K. S. Vink, D. A. Tirrell, *Nat. Protoc.* **2007**, *2*, 1879–1883.

[6] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

[7] A. J. Link, M. K. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10180–10185.

[8] V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angew. Chem.* **2009**, *121*, 10063–10067; *Angew. Chem. Int. Ed.* **2009**, *48*, 9879–9883.

[9] A. J. Link, M. K. S. Vink, D. A. Tirrell, *J. Am. Chem. Soc.* **2004**, *126*, 10598–10602.

[10] P. S. Daugherty, G. Chen, B. L. Iverson, G. Georgiou, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2029–2034.

[11] P. S. Daugherty, G. Chen, M. J. Olsen, B. L. Iverson, G. Georgiou, *Protein Eng.* **1998**, *11*, 825–832.

[12] A. Hayhurst, S. Happe, R. Mabry, Z. Koch, B. L. Iverson, G. Georgiou, *J. Immunol. Methods* **2003**, *276*, 185–196.

[13] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24.

[14] R. Karlsson, *J. Mol. Recognit.* **1999**, *12*, 285–292.

[15] A. Önell, K. Andersson, *J. Mol. Recognit.* **2005**, *18*, 307–317.

[16] P. D. Jeffrey, R. K. Strong, L. C. Sieker, C. Y. Chang, R. L. Campbell, G. A. Petsko, E. Haber, M. N. Margolies, S. Sheriff, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10310–10314.

Table S1. Amino acid mutations in clones isolated from Library 1 after sorting the library three times for clones tolerant of Hpg in place of Met.

Clone	M(H20)	S(H24)	M(H80)	L(H82C)	I(L2)	A(L12)
1	-	P	L	-	V	-
3	-	-	L	-	-	-
6	I	-	L	-	-	T
7	-	-	L	-	-	-
10	-	-	L	-	-	-
2	I	-	L	M	-	-
4	I	-	L	M	-	-
5	I	-	L	M	-	-
8	I	-	L	M	-	-
9	I	-	L	M	-	-

Table S2. Summary of conditions used in flow cytometric sorting of Library 2.

	Expression in Medium Containing		
Population	Hpg (2.5 mM)	Aha (0.28 mM)	Nrl (0.30 mM)
Naïve Library 2	High Fluorescence	High Fluorescence	High Fluorescence
1x sort	High Fluorescence Lower Antigen Conc.	High Fluorescence Lower Antigen Conc.	High Fluorescence Lower Antigen Conc.
2x sort	Fluorescence, Viability 15 minute competition	Fluorescence, Viability 40 minute competition	Fluorescence, Viability 8 minute competition
3x sort	Fluorescence, Viability 40 minute competition	Fluorescence, Viability 100 minute competition	Fluorescence, Viability 15 minute competition

Table S3. Amino acid mutations in clones isolated from Library 2.

Colony	Notes	E (H1)	E (H10)	L (H11)	M (H20)	S (H24)	Y (H27)	F (H29)	T (H30)	M (H34)	H (H41)	D (H46)	Y (H53)	Q (H61)	M (H80)	L (H82C)	Y (H102)	A (H107)	S (H113)	G (Linker1)	G (Linker7)	G (Linker9)
Hpg4x1		-	-	-	I	Y	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x3		-	-	-	I	F	N	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x4		-	-	-	I	-	N	S	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x5	†	-	-	-	I	-	-	S	-	-	-	-	-	-	L	M	-	-	-	S	-	-
Hpg4x6		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x7		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x8		-	-	-	I	Y	-	S	-	-	-	-	-	-	L	M	N	-	-	-	-	-
Hpg4x9		-	D	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x10		-	-	-	L	-	-	-	-	-	-	-	-	-	L	M	-	T	-	-	-	-
Hpg4x11		-	-	-	I	F	N	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg4x12	‡	-	-	-	I	-	-	-	-	-	-	-	-	R	L	M	-	-	-	-	-	-
Hpg3x1		-	-	M	I	Y	-	-	-	-	N	-	-	-	L	M	-	-	-	-	-	R
Hpg3x2		-	-	-	I	-	-	-	-	-	-	E	-	-	L	M	-	-	-	-	-	-
Hpg3x3		-	-	S	I	F	-	-	S	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg3x4		-	-	-	I	-	-	S	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg3x5		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Hpg3x6		-	-	-	I	F	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha3x1		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha3x2		-	-	-	I	-	-	S	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha3x3		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha3x4		-	-	-	I	-	-	S	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha3x5		-	-	-	L	-	-	S	-	-	-	-	-	-	L	-	-	-	-	-	-	-
Aha3x6		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha4x1		-	-	-	I	Y	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha4x2	†	-	-	-	I	Y	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Aha4x4		-	-	-	I	Y	-	-	-	I	-	-	-	-	L	M	-	-	-	-	-	-
Aha4x5		-	A	-	I	-	-	S	-	-	-	-	-	-	L	-	-	-	-	-	-	V
Aha4x6		-	-	-	I	-	-	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Nrl4x1		-	-	-	I	F	N	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Nrl4x2		-	-	-	-	Y	-	-	-	-	-	-	-	-	L	-	-	-	-	-	-	-
Nrl4x3		G	-	-	I	F	-	S	-	-	-	-	N	-	L	M	-	-	-	-	-	-
Nrl4x4		-	-	-	I	-	-	S	-	-	-	E	-	-	L	M	-	-	-	-	-	-
Nrl4x5		-	-	-	-	-	-	S	-	-	-	-	-	-	L	-	-	-	-	-	-	-
Nrl4x6		-	-	-	-	Y	-	S	-	-	-	-	-	-	L	-	-	-	-	-	-	-
Nrl4x7		-	-	-	I	Y	-	-	-	-	-	-	-	-	L	-	-	-	-	-	-	-
Nrl4x8		-	-	-	I	F	N	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Nrl4x9		-	-	-	I	-	-	S	-	V	-	-	-	-	L	M	-	-	-	-	R	-
Nrl4x10		-	-	-	I	F	N	-	-	-	-	-	-	-	L	M	-	-	-	-	-	-
Nrl4x11		-	-	-	L	Y	-	S	-	-	-	-	-	-	L	-	-	-	F	-	-	-

†Truncated fragment.
‡Missing amino acids 400−404.¶
¶Amino acids located after the end of the fragment are numbered according to their position within the open reading frame of the Lpp-OmpA-scFv fusion protein, starting at number one (initiator methionine; leader sequence included in numbering).

Table S3, continued. Amino acid mutations in clones isolated from Library 2.

Colony	Notes	I (L2)	A (L12)	S (L22)	S (L25)	Q (L27)	P (L40)	N (L53)	S (L64)	E (L68)	F (L71)	T (L74)	D (L76)	V (L78)	I (L85)	K394¶	R395¶	E404¶
Hpg4x1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x3		-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-
Hpg4x4		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x5	†	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Stop	-	-
Hpg4x6		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x7		-	-	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x8		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x9		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x10		-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-
Hpg4x11		-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-
Hpg4x12	‡	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	H	-
Hpg3x1		-	-	-	-	-	-	-	-	-	-	S	N	-	-	-	-	-
Hpg3x2		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x4		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x5		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x6		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x2		-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-
Aha3x3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x4		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x5		-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x6		-	-	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-
Aha4x1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x2	†	-	T	-	-	K	-	-	-	-	-	-	-	-	-	-	-	Stop
Aha4x4		-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x5		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x6		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x1		-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-
Nrl4x2		-	-	-	-	-	S	-	-	-	-	-	-	-	K	-	-	-
Nrl4x3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x4		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x5		V	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-
Nrl4x6		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x7		V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x8		-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-
Nrl4x9		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x10		-	-	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-
Nrl4x11		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

†Truncated fragment.
‡Missing amino acids 400−404.¶
¶Amino acids located after the end of the fragment are numbered according to their position within the open reading frame of the Lpp-OmpA-scFv fusion protein, starting at number one (initiator methionine; leader sequence included in numbering).

Table S4. Off-rate estimates performed using cell surface-displayed scFvs.*

Hpg4x Clones			
Sample	AA Context	k_{off} (10^{-3} s^{-1})	Remarks
Mut2	Hpg	1.097 ± 0.081	†
Base	Met	0.672 ± 0.047	
Mut2	Met	0.528 ± 0.029	
Hpg4x1	Hpg	0.795 ± 0.059	
Hpg4x2	Hpg	1.153 ± 0.175	†
Hpg4x3	Hpg	0.695 ± 0.021	
Hpg4x4	Hpg	0.912 ± 0.030	
Hpg4x5	Hpg	0.626 ± 0.016	‡
Hpg4x6	Hpg	0.966 ± 0.070	
Hpg4x7	Hpg	1.368 ± 0.059	
Hpg4x8	Hpg	0.883 ± 0.040	
Hpg4x9	Hpg	0.914 ± 0.051	
Hpg4x10	Hpg	1.274 ± 0.090	
Hpg4x11	Hpg	0.871 ± 0.039	
Hpg4x12	Hpg	1.241 ± 0.087	
Hpg3x Clones			
Sample	AA Context	k_{off} (10^{-3} s^{-1})	Remarks
80L HisGS	Met	0.490 ± 0.046	
Hpg3x1	Hpg	0.563 ± 0.012	
Hpg3x2	Hpg	1.372 ± 0.061	
Hpg3x3	Hpg	0.668 ± 0.033	
Hpg3x4	Hpg	0.906 ± 0.055	
Hpg3x5	Hpg	0.912 ± 0.026	†
Hpg3x6	Hpg	1.021 ± 0.030	
Aha3x and Aha4x Clones			
Sample	AA Context	k_{off} (10^{-3} s^{-1})	Remarks
Base	Met	0.661 ± 0.071	
Mut2	Met	0.506 ± 0.026	
Mut2	Aha	0.200 ± 0.007	
Aha3x1	Aha	0.054 ± 0.017	
Aha3x2	Aha	0.092 ± 0.019	
Aha3x3	Aha	0.161 ± 0.089	
Aha3x4	Aha	0.131 ± 0.016	
Aha3x5	Aha	0.193 ± 0.024	
Aha3x6	Aha	0.143 ± 0.013	
Aha4x1	Aha	0.061 ± 0.008	
Aha4x2	Aha	0.086 ± 0.014	‡
Aha4x4	Aha	0.122 ± 0.021	
Aha4x5	Aha	0.076 ± 0.018	
Aha4x6	Aha	0.154 ± 0.024	
Nrl4x Clones			
Sample	AA Context	k_{off} (10^{-3} s^{-1})	Remarks
Base	Met	0.736 ± 0.061	
Mut2	Met	0.533 ± 0.030	
Base	Nrl	1.364 ± 0.047	
Mut2	Nrl	1.224 ± 0.023	
Nrl4x1	Nrl	0.509 ± 0.019	
Nrl4x2	Nrl	0.518 ± 0.026	
Nrl4x3	Nrl	0.366 ± 0.056	
Nrl4x4	Nrl	0.745 ± 0.025	
Nrl4x5	Nrl	0.787 ± 0.045	
Nrl4x6	Nrl	1.025 ± 0.243	
Nrl4x7	Nrl	0.435 ± 0.030	
Nrl4x8	Nrl	0.514 ± 0.024	
Nrl4x9	Nrl	0.494 ± 0.047	
Nrl4x10	Nrl	0.548 ± 0.022	
Nrl4x11	Nrl	0.262 ± 0.020	

*Cells displaying scFvs were subjected to off rate characterizations using an adaptation of the method of Daugherty et al.[11] (see Materials and Methods). Cells displaying the Nrl forms of scFvs were subjected to a less stringent wash procedure prior to estimating off rates. Errors reported are the 95% confidence intervals in the fits as determined by the program Igor (Wavemetrics, Lake Oswego, OR).

†A large portion of **5** appeared to have dissociated from cells prior to kinetic competition.

‡Truncated antibody fragment.

Table S5. Frequently mutated positions (>5%) of scFvs identified in high-throughput sequencing of sorted populations.

Nucleotide Position*	Base	Frequent Mutation	Amino Acid Mutation†	Mutational Frequency			
				Library 1 Hpg Sort 3	Lib2 Hpg4x	Lib2 Aha4x	Lib2 Nrl4x
678	c	t		0.00	0.04	0.13	0.00
693	g	a,c	M(H20)I	0.73	0.86	0.94	0.57
703	t	c	S(H24)P	0.06	0.03	0.01	0.02
704	c	a	S(H24)Y	0.03	0.18	0.32	0.34
		t	S(H24)F	0.01	0.11	0.24	0.21
712	t	a	S(H27)N	0.00	0.03	0.01	0.12
		c	S(H27)H	0.00	0.01	0.00	0.01
719	t	c	F(H29)S	0.02	0.32	0.29	0.56
721	a	c	T(H30)P	0.01	0.05	0.07	0.02
733	a	g	M(H34)V	0.00	0.02	0.01	0.18
735	g	a	M(H34)I	0.00	0.02	0.05	0.04
771	c	a	D(H46)E	0.01	0.13	0.03	0.05
803	t	c	V(H56)A	0.00	0.01	0.00	0.06
847	g	t	V(H71)F	0.00	0.01	0.01	0.03
		a	V(H71)I	0.00	0.00	0.00	0.02
861	c	a,t,g		0.00	0.05	0.00	0.05
874	a	t,c	M(H80)L	0.90	0.99	0.99	0.86
889	t	a	L(H82C)M	0.62	0.71	0.78	0.50
955	t	c	Y(H102)H	0.00	0.00	0.00	0.03
		a	Y(H102)N	0.00	0.05	0.01	0.03
970	g	a	A(H107)T	0.01	0.08	0.11	0.00
1002	c	t		0.00	0.01	0.08	0.00
1018	t	g	S(Linker10)A	0.01	0.00	0.01	0.01
		c	S(Linker10)P	0.00	0.00	0.08	0.00
1039	a	g	I(L2)V	0.06	0.09	0.05	0.13
1069	g	a	A(L12)T	0.08	0.05	0.07	0.09
1078	c	a	L(L15)I	0.01	0.05	0.15	0.00
1086	a	g		0.00	0.00	0.05	0.01
1095	g	a		0.01	0.00	0.06	0.06
1099	t	c	S(L22)P	0.03	0.00	0.24	0.24
1108	t	a	S(L25)T	0.00	0.01	0.06	0.00
1110	c	a		0.00	0.01	0.00	0.07
1114	c	a	Q(L27)K	0.00	0.04	0.15	0.00
1209	c	t		0.06	0.07	0.04	0.02
1236	c	t		0.00	0.03	0.00	0.14
1257	a	g,t		0.00	0.02	0.13	0.00
1282	g	a	V(L78)M	0.00	0.03	0.00	0.16
1296	t	c		0.00	0.00	0.08	0.00
1350	g	t		0.00	0.00	0.06	0.01
1399	g	t	E404Stop	0.00	0.05	0.17	0.00
		a	E404K	0.00	0.01	0.00	0.00

*Nucleotide numbering scheme based on position 1 in pQE-80L as defined by the manufacturer (Qiagen) and adapted to the Lpp-OmpA-scFv variant plasmids.
†Numbering scheme same as in Table S3.

Table S6. Characterization of soluble scFvs: expression yields, binding kinetics, and amino acid replacement estimates.

AA Context	Met			Hpg			
Clone	Exp. Yield (mg/L)*	k_{on} (10^6 $M^{-1}s^{-1}$)†	k_{off} (10^{-3} s^{-1})	Exp. Yield (mg L^{-1})	% Met replacement‡	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})
Base	0.63	1.35 ± 0.08	1.85 ± 0.15				
Mut2	1.43	1.50 ± 0.16	1.72 ± 0.17	0.73	86 ± 4	2.23 ± 0.44	8.09 ± 0.53
Hpg3x1	1.17	1.29 ± 0.40	1.92 ± 0.29	1.47	90 ± 2	2.31 ± 0.34	6.54 ± 1.41
Hpg3x3	0.77	1.42 ± 0.12	0.93 ± 0.07	0.62	89 ± 0.4	1.80 ± 0.21	4.63 ± 0.34
Hpg4x3	2.44	1.45 ± 0.10	1.08 ± 0.09	0.26	87 ± 3	1.37 ± 0.12	6.61 ± 0.73
Hpg4x8	1.61	1.59 ± 0.05	1.03 ± 0.09	0.47	89 ± 4	1.31 ± 0.08	5.59 ± 1.42
AA Context	Met			Aha			
Clone	Exp. Yield (mg/L)	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})	Exp. Yield (mg L^{-1})	% Met replacement	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})
Base	0.63	1.35 ± 0.08	1.85 ± 0.15				
Mut2	1.43	1.50 ± 0.16	1.72 ± 0.17	0.98	84 ± 1	1.20 ± 0.11	0.96 ± 0.16
Aha3x2	1.59	1.72 ± 0.08	1.50 ± 0.14	1.34	80 ± 4	1.58 ± 0.18	0.81 ± 0.12
Aha4x4	2.31	1.41 ± 0.09	1.00 ± 0.19	1.21	Not determined	1.44 ± 0.28	0.70 ± 0.15
Aha4x5	1.44	1.55 ± 0.15	1.31 ± 0.29	0.86	83 ± 4	1.93 ± 0.47	0.68 ± 0.11
AA Context	Met			Nrl			
Clone	Exp. Yield (mg/L)	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})	Exp. Yield (mg L^{-1})	% Met replacement	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})
Base	0.63	1.35 ± 0.08	1.85 ± 0.15	Not determined	86	0.54 ± 0.36	2.31 ± 0.39
Mut2	1.43	1.50 ± 0.16	1.72 ± 0.17	1.12	88 ± 5	1.34 ± 0.07	2.07 ± 0.20
Nrl4x3	4.98	1.73 ± 0.06	1.08 ± 0.04	2.22	87 ± 5	1.90 ± 0.14	1.16 ± 0.10
Nrl4x4	4.07	1.39 ± 0.03	2.23 ± 0.19	1.53	93 ± 2	1.58 ± 0.16	2.51 ± 0.07
Nrl4x11	0.91	1.26 ± 0.12	0.93 ± 0.09	0.64	89 ± 3	1.19 ± 0.02	1.08 ± 0.07

*Expression yields were calculated based on bicinchoninic acid (BCA) assays of concentrated protein samples after size exclusion chromatography.

†Binding kinetics were determined using a Biacore T100 instrument. Digoxigenin (antigen) was immobilized on the surface of CM5 chips using a two-step immobilization process. First, antigen was conjugated to bovine serum albumin (BSA). This conjugate was then attached to the chip using standard amine coupling procedures. Multicycle kinetic assays using a range of scFv concentrations from 0.3125 to 20 nM were used to obtain kinetic parameters. The parameters reported here are the result of two independent assays performed on four chip surfaces having a range of ligand densities, displayed as averages plus or minus standard deviations. Kinetics were well-described by a standard 1:1 binding transport model. In the case of Hpg-containing fragments, a significant mass transport constant was invoked in the best fits of the data. All other data were found to be free of substantial mass transport limitations.

‡Amino acid replacement levels were estimated on the basis of matrix-assisted laser desorption ionization (MALDI) mass spectrometry of trypsinized scFv samples as described in Materials and Methods. Data are reported as the mean plus or minus the standard deviation of two or more independent trypsinizations and MALDI mass spectrum acquisitions.

Table S7. Amino acid mutations in clones produced in soluble form.

Colony	E (H1)	E (H10)	L (H11)	M (H20)	S (H24)	Y (H27)	F (H29)	T (H30)	M (H34)	H (H41)	D (H46)	Y (H53)	M (H80)	L (H82C)	Y (H102)	S (H113)	G (Linker9)	S (L25)	E (L68)	T (L74)	D (L76)	V (L78)
Hpg3x1	-	-	M	I	Y	-	-	-	-	N	-	-	L	M	-	-	R	-	-	S	N	-
Hpg3x3	-	-	S	I	F	-	-	S	-	-	-	-	L	M	-	-	-	-	-	-	-	-
Hpg4x3	-	-	-	I	F	N	-	-	-	-	-	-	L	M	-	-	-	-	-	-	-	M
Hpg4x8	-	-	-	I	Y	-	S	-	-	-	-	-	L	M	N	-	-	-	-	-	-	-
Aha3x2	-	-	-	I	-	-	S	-	-	-	-	-	L	M	-	-	-	-	D	-	-	-
Aha4x4	-	-	-	I	Y	-	-	-	I	-	-	-	L	M	-	-	-	T	-	-	-	-
Aha4x5	-	A	-	I	-	-	S	-	-	-	-	-	L	-	-	-	V	-	-	-	-	-
Nrl4x3	G	-	-	I	F	-	S	-	-	-	-	N	L	M	-	-	-	-	-	-	-	-
Nrl4x4	-	-	-	I	-	-	S	-	-	-	E	-	L	M	-	-	-	-	-	-	-	-
Nrl4x11	-	-	-	L	Y	-	S	-	-	-	-	-	L	-	-	F	-	-	-	-	-	-

Table S8. Dye labeling of Met- and Aha-containing proteins using **6**.

Protein	Aha3x2		Aha4x4		Aha4x5	
AA Context	Met	Aha	Met	Aha	Met	Aha
Dyes/Protein	0.00 ± 0.00	0.43 ± 0.15	0.00 ± 0.00	0.04 ± 0.05	0.00 ± 0.00	0.38 ± 0.20

Table S9. Kinetic characterization of scFvs before and after reaction with **6** (strain-promoted click chemistry)*.

AA Context	Met			
	No Click Chemistry		Click Chemistry	
Clone	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})
Aha3x2	1.72 ± 0.08	1.50 ± 0.14	1.03 ± 0.03	1.38 ± 0.14
Aha4x4	1.41 ± 0.09	1.00 ± 0.19	0.75 ± 0.15	0.92 ± 0.25
Aha4x5	1.55 ± 0.15	1.31 ± 0.29	1.16 ± 0.16	1.29 ± 0.23
AA Context	Aha			
	No Click Chemistry		Click Chemistry	
Clone	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})
Aha3x2	1.58 ± 0.18	0.81 ± 0.12	0.79 ± 0.07	0.70 ± 0.10
Aha4x4	1.44 ± 0.28	0.70 ± 0.15	0.93 ± 0.15	0.63 ± 0.17
Aha4x5	1.93 ± 0.47	0.68 ± 0.11	1.09 ± 0.08	0.71 ± 0.14

*Kinetic characterizations were performed as described in Table S6 and Materials and Methods. Clicked scFv samples were prepared for Biacore assays without separating unreacted dyes or quenching reagents from solution.

Table S10. Oligonucleotides used in study.

Name	Sequence
EcoRiElimFwd	5’-GTTTTTTTGGGCTAGCGTTTTCGAGCTCGGTACC-3’
EcoRIElimRev	5'-GGTACCGAGCTCGAAAACGCTAGCCCAAAAAAAC-3'
Lpp-OmpA-antidigFwd	5’-GAATTCGAGCTCGGTACCCGGGCTAGAG-3’
Lpp-OmpA-antidigRev1	5’-GTGATGAGAACCACGGTCCTCGGGGTCTTCCGGG-3’
Lpp-OmpA-antidigRev1	5’-TGCTCTAAGCTTACTAGTGATGGTGATGGTGATGAGAACCACGGTCCTC-3’
HindIIIElimFwd	5’-GAAACCAGGACAGCCACCCAAACTACTCATCTATAAGGTATCC-3’
HindIIIElimRev	5’-GGATACCTTATAGATGAGTAGTTTGGGTG GCTGTCCTGGTTTC-3’
LppHisRescue	5’- CGTGGTTCTCATCACCATCACCATCACGGCTCGTAGTAAGCTTAGAGCA-3’
80LLibFwd	5’-CAATTGTGAGCGGATAACAATTTCAC-3’
80LLibRev	5’-GCTCCTGAAAATCTCGCCAAGCTAGC-3’
PstIAddFwd	5’-CTAAAATCGATCAGGGAATTAACCTGCAGGTTGGCTTTGAAATGGGTTAC-3’
PstIAddRev	5’-GTAACCCATTTCAAAGCCAACCTGCAGGTTAATTCCCTGATCGATTTTAG-3’
PstIElimFwd	5’-GATGATGCTGCAATATATTATTGTAGCCAAACTACGCATGTTCC-3’
PstIElimRev	5’-GGAACATGCGTAGTTTGGCTACAATAATATATTGCAGCATCATC-3’
BglIIAddFwd	5’-GCCAGCCAGAACTCGCCCCGGAAGATCTCGAGGACCGTGGTTCTCATC-3’
BglIIAddRev	5’-GATGAGAACCACGGTCCTCGAGATCTTCCGGGGCGAGTTCTGGCTGGC-3’
AntidigpAK400Fwd	5'-ATATAAGGCCCAGCCGGCCATGGCGGGAATTCGTGAAGTTCAACTGCAAC-3'
AntidigpAK400Rev	5'-CTCGCCCCGGAAGATCTCGAGGACGCGGCCTCGGGGGCCAATATA-3'
AntidigpAK400FwdNrl4x3	5'-ATATAAGGCCCAGCCGGCCATGGCGGGAATTCGAGGAGTTCAACTGCAAC-3'

Figure S1. Fluorescence-activated cell sorting of Library 1 for clones that function when Met is replaced by Hpg. Expression of cells harboring the naïve library and sorted populations was performed in minimal media containing concentrations of Met and Hpg as indicated. After treatment of induced cells with **5**, the fluorescence of each sample was measured on the flow cytometer. Boxed populations were sorted, and events falling into the top ~0.1% of fluorescence measurements in the BODIPY FL channel were retained for the next round of sorting. Note that the naïve library was sorted after induction in medium containing 20 μM Met and 2.5 mM Hpg. The addition of Met to the media results in a population exhibiting an intermediate level of fluorescence compared to samples induced in Met or Hpg only, implying a partial replacement of Met with Hpg. AA, amino acid. μ, mean fluorescence.

Figure S2. Fluorescence-activated cell sorting of Library 2 for clones that function when Met is replaced by ncAAs. Expression of the naïve library and sorted populations was performed in minimal media containing concentrations of ncAAs as indicated. The fluorescence of each population was determined by flow cytometry. Starting with “Sort 2” in each amino acid context, cells were washed twice prior to flow cytometry. Each sample was subjected to labeling and kinetic competitions as outlined in Table S2; the top ~0.1%–1.0% of fluorescent events were retained. AA, amino acid. μ, mean fluorescence.

Figure S3. Strain-promoted click chemistry on scFvs using fluorescently labeled compound **6**. Solutions of the Met and Aha forms of scFv variants at 1.75 μM concentrations were subjected to reactions with 10 μM **6** for a period of 1 hour. After quenching the reactions with excess 3-azidopropanol, protein samples were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). A Typhoon Trio imaging system was used to confirm scFv labeling.

Figure S4. MALDI mass spectrometry on trypsinized fragments of the Aha forms of Aha4x5, Aha3x2, and Aha4x4 before and after reaction with **6**. A) MALDI mass spectrometry on trypsinized samples of the Aha form of Aha4x5 before and after reaction with **6**. The peak identified as "Unmodified" includes the Aha residue located at position H34 and has sequence SSGYISTDFYAhaNWVR ($[M+H]^+$ calculated: 1820.83 Da, $[M+H]^+$ observed (unclicked): 1820.65 Da). The peak labeled "Modified" is the same peptide after reaction with **6** ($[M+H]^+$ calculated: ~2654.96 Da, $[M+H]^+$ observed (clicked): 2656.22 Da), and is present in the mass spectrum only after click chemistry ("click" spectrum offset by 10 Da and 1000 counts for clarity). B) MALDI mass spectrometry on trypsinized samples of the Aha form of Aha3x2 before and after reaction with **6**. The peak identified as "Unmodified" includes the Aha residue located at position H34 and has sequence SSGYISTDFYAhaNWVR ($[M+H]^+$ calculated: 1820.83 Da, $[M+H]^+$ observed (unclicked): 1820.26 Da). The peak labeled "Modified" is the same peptide after reaction with **6** ($[M+H]^+$ calculated: ~2654.96 Da, $[M+H]^+$ observed (clicked): 2656.43 Da), and is present in the mass spectrum only after click chemistry ("click" spectrum offset by 10 Da and 1000 counts for clarity). C) MALDI mass spectrometry on trypsinized samples of the Aha form of Aha4x4 before and after reaction with **6**. The peak identified as "Unmodified" includes position H34 (Met (Aha) mutated to Ile) and has sequence YSGYIFTDFYINWVR ($[M+H]^+$ calculated: 1943.94 Da, $[M+H]^+$ observed: 1943.86 Da). No new peaks are visible in the mass spectrum of the "clicked" sample ("click" spectrum offset by 10 Da and 1000 counts for clarity).

Figure S5. Positions of Met residues and residues mutated to Met in scFv form of 26-10 antibody. Each site in the protein that is occupied by Met in at least one soluble variant is shown in space filling representation, with surface accessibilities calculated from GetArea listed beneath the residue of interest. The accessibility numbers reported here are the average of three calculations performed using PDB files 1IGI (1 Fab molecule/unit cell) and 1IGJ (2 Fab molecules/unit cell, complexed to ligand).[16] The backbone of the Fv portion of the Fab is outlined with a ribbon diagram (gray), with amino acids of interest shown in space filling representation and antigen shown as sticks in white. H, heavy chain. L, light chain. Space filling residues colored by elements: carbon, gray; oxygen, red; nitrogen, blue; sulfur, yellow.

Figure S6. Flow cytometry of the Met and Aha forms of cell surface-displayed Aha4x5 to probe binding function and chemical modification with strained alkynes. The two-dimensional dot plots show simultaneous measurements of the amounts of **5** (a measurement of binding function) and streptavidin-phycoerythrin (SA-PE; a measure of reactivity toward cyclooctyne **7**) bound to the cell surface. Mean fluorescence values for **5** (denoted "BD μ") and SA-PE (denoted "PE μ") are indicated on each plot. A) Cells displaying the Met form of Aha4x5 without treatment with fluorescent probes. B) Probes of binding function of cells using **5**, with amino acid context and treatment with **7** (i.e., strain-promoted click chemistry) as noted in panels. Some reduction in the amount of **5** bound per cell is noted upon treatment of the Aha form of Aha4x5 with **7**. This may be attributable to steric effects on the cell surface after treatment with **7** or the denaturation of some scFvs during the click reaction. C) Probes of the extent of modification of cells using SA-PE, with amino acid context and treatment with **7** as noted in panels. D) Use of **5** and SA-PE to probe binding and modification simultaneously, with amino acid con-

text and treatment with **7** as noted in panels. Simultaneous examination of these two properties on the cell surface gives results consistent with those of experiments in which binding and modification were assessed separately. E) Adding unlabeled digoxin to the Aha form of cells expressing Aha4x5 treated with **7** prior to treatment with **5** and SA-PE results in cells unable to bind to fluorescent antigen while retaining modification. F) Use of **5** and SA-PE to probe the function and modification of uninduced cells harboring plasmids encoding the Aha4x5 cell surface display construct. Amino acid context and treatment with **7** as noted in panels. The mean PE fluorescence of cells grown in medium supplemented with Aha and treated with **7** is less than one quarter of the mean PE fluorescence of cells displaying the Aha form of Aha4x5 and treated with **7**, suggesting that scFv labeling is responsible for approximately 75 percent of total cell labeling. G) Use of **5** and SA-PE to probe for antigen binding and biotin modification of cells lacking the plasmid for the cell surface-displayed Aha4x5 construct. Amino acid context and treatment with **7** as noted in panels. The mean PE fluorescence of cells lacking the display plasmid, grown in Aha, and treated with **7** is similar to the mean PE fluorescence of uninduced cells harboring the display plasmid, grown in Aha, and treated with **7**, suggesting that the observed increase in fluorescence observed in panels (F) and (G) comes from labeling of newly synthesized membrane proteins during the course of induction, not leaky expression from the display construct.