CHAPTER 4

Antibody Fragment Engineering with Noncanonical Amino Acids

Abstract

Noncanonical amino acids (ncAAs) can be used to modulate the physical and chemical properties of proteins. In this work, we examine how ncAAs can be used to engineer the binding properties and chemical reactivity of a model anti-digoxin antibody fragment in its single chain variable fragment (scFv) form. Experiments with scFvs displayed on the surface of *Escherichia coli* cells revealed that replacement of the methionine (Met) residues of the scFv with an analog containing an alkyne side chain reduced the fluorescence levels of cells treated with a fluorescently labeled antigen to background levels, indicating loss of binding function. Replacement of Met with analogs containing aliphatic and azide side chains left the fluorescence of cells unchanged and reduced by a factor of 0.6, respectively. Fluorescence-activated cell sorting of libraries of cell surface-displayed scFvs enabled the isolation of clones functional in multiple amino acid contexts. Cells displaying variants containing alkyne, azide, and aliphatic analogs and treated with fluorescently labeled antigen were more fluorescent than cells displaying the Met form of the parent scFv by factors of roughly 1.7, 3.5, and 1.3, respectively. Furthermore, the amino acid context used during high-throughput screening experiments appears to affect the frequencies of mutations occurring at various positions within the scFv construct. High-throughput sequencing revealed that populations isolated in different amino acid contexts exhibit mutational rates differing by greater than twenty percent at some residues in the protein.

Characterization of soluble scFvs indicated that each ncAA used in this study modulates the binding kinetics of scFvs in a distinct fashion. Perhaps most interestingly, scFvs containing the azide-containing analog azidohomoalanine (Aha) exhibit improved

binding kinetics relative to their methionine-containing counterparts. Replacement of Met by Aha in several variants lowers the dissociation constant of the fragment by up to a factor of two. Chemical conjugation of azide-containing scFvs to fluorescent dyes and biotin proved facile with strain-promoted cycloaddition reactions. Quantifications of the extent of reaction using fluorescent dyes revealed that approximately 0.4 dyes had been conjugated per protein, and the resulting conjugates were found to retain their binding function in kinetic and Western blotting assays. Experiments in which Aha-containing fragments were displayed on the surface of *Escherichia coli* cells and subjected to strain-promoted cycloadditions demonstrated that the extent of chemical modification and antigen binding can be monitored simultaneously and used to isolate cells displaying functional, modified proteins. These experiments demonstrate how ncAAs can be used to modulate multiple properties of antibody fragments and illustrate the feasibility of developing and screening libraries of chemically modified proteins. Evolved, functional bioconjugates may be applicable to a variety of outstanding diagnostic and therapeutic problems.

Introduction

The vastness of protein sequence space provides scientists with ample possibilities for mutating and engineering proteins in order to improve their existing functions or properties and to impart them with nonnatural characteristics (1-3). Biosynthetic incorporation of noncanonical amino acids (ncAAs) into proteins is a powerful approach to augmenting or altering the functions of full-length proteins (4-7). Residue-specific incorporation of ncAAs into proteins by the global replacement of one or more of the twenty canonical amino acids can facilitate large changes in protein properties, either by

introducing new chemical functionalities into proteins or by employing subtle side chain structures that change the stability, aggregation, or solvation characteristics of proteins (8-12). Site-specific incorporation approaches also afford opportunities for changing local protein properties by making atomic-level changes to amino acid side chains or polypeptide backbones (6, 7, 13, 14). However, property changes that result from ncAA incorporation are not always readily predictable, and can lead to insignificant or even deleterious changes to proteins of interest.

The techniques of directed evolution provide a powerful set of tools for engineering proteins with more desirable properties. Our laboratory and others have demonstrated that directed evolution can be combined with ncAAs to enable the creation of ncAA-containing proteins with user-defined characteristics. Using high-throughput screening approaches, our laboratory has previously evolved fluorinated proteins to have biophysical properties comparable to the properties of conventional proteins. These efforts resulted in the isolation of a thermostable fluorinated enzyme and a fluorinated green fluorescent protein with folding kinetics comparable to the parent protein (15, 16). The Schultz laboratory has combined the use of phage display and antibody fragments containing site-specifically incorporated ncAAs to target specific antigens. They were able to create libraries of antibody fragments containing the ncAAs sulfotyrosine and *p*-boronophenylalanine and successfully screen these libraries for binders against the HIV protein gp120 (17, 18) and the acyclic sugar glucosamine (19), respectively. The results of directed evolution experiments performed with libraries of ncAA-containing proteins illustrate the feasibility and power of using high-throughput screening techniques to explore unnatural sequence spaces in search of molecules with properties of interest (i.e., fluorinated, functional

molecules or high-affinity binders against a particular target). The ability to tailor the properties of molecules containing ncAAs via directed evolution may be particularly useful in the context of protein therapeutics, where candidate molecules must have many favorable properties. The chemical reactivity of protein therapeutics is especially critical, often simultaneously requiring selective conjugation of proteins to other molecules while preserving the original molecule through the reduction or elimination of all undesirable reactions with native side chains (20-23). Some recent studies have demonstrated how unique chemical handles can be introduced into proteins in the form of ncAAs and exploited through selective chemistries (24). However, little work has been done to understand how the functional properties of these proteins change after incorporation of ncAAs or how directed evolution can be used to manipulate these proteins.

In this work, we focus on changing the chemistries of antibody fragments through the global replacement of methionine (Met, **1**, scheme 4.1) with homopropargylglycine (Hpg, **2**), azidohomoalanine (Aha, **3**), or norleucine (Nrl, **4**). Substituting Hpg or Aha for Met introduces residues useful for performing bioorthogonal chemical reactions (scheme 4.2) including copper-catalyzed azide-alkyne cycloaddition (CuAAC) chemistry (Hpg, Aha) and copper-free, strain-promoted cycloadditions (Aha) (25). While replacement of Met by Nrl does not introduce any new chemical functionality, it eliminates the thioether moiety of Met, an oxidizable group that can pose problems for long-term protein storage (26-28).

We investigate the directed evolution of antibody fragments containing ncAAs and the effects of ncAA incorporation on the binding and chemical properties of these therapeutically relevant proteins. Replacing Met with ncAAs changes the genetic code,

which can be thought of as changes to the amino acid context or sequence space in which a protein exists. We have adapted a previously reported *E. coli* cell surface display anchor for use with ncAAs and used it to engineer ncAA-containing anti-digoxin single chain variable fragments (scFvs). This platform enabled us to perform screens of antibody fragment libraries using fluorescence-activated cell sorting (FACS) in a number of amino acid contexts and study how different ncAA side chains distinctly affect the course of directed evolution experiments. Production and characterization of several soluble scFv variants containing the different Met analogs highlighted the effects of each analog on antigen binding kinetics. Furthermore, examination of the chemical reactivities of scFvs showed that azide-containing fragments can be used to create functional conjugates via strain-promoted click chemistry. Finally, we demonstrate the feasibility of using *E. coli* cell surface display for screening libraries of proteins for functional, chemically modified protein variants via flow cytometry. Our findings illustrate how ncAAs can be used to manipulate the antigen binding and chemical reactivity of antibody fragments simultaneously and suggest new methods for developing functional bioconjugates.

Results and Discussion

Cell surface display and flow cytometry. The Lpp-OmpA' *E. coli* cell surface display system was adapted to be compatible with engineering antibody fragments containing ncAAs. This system has previously been used to screen libraries of anti-digoxin single chain variable fragments (scFvs) using flow cytometry (29-31). We placed the display construct under the control of the isopropyl-β-D-1-thiogalactopyranoside (IPTG)-inducible T5 promoter in order to avoid potential incompatibilities with residue-specific ncAA

incorporation techniques (31). Tight repression was maintained by cotransforming cells with two plasmids containing constitutively expressed *lacI*: derivatives of pQE-80L containing display constructs and the repressor plasmid pREP4.

Flow cytometry experiments were used to investigate the function of the scFv form of a variant of the high affinity murine 26-10 anti-digoxin antibody (32) in multiple amino acid contexts. Display of functional copies of the scFv (referred to here as the base construct or Base) on *E. coli* cells after IPTG induction was confirmed by probing induced cells with BODIPY FL digoxigenin **5** and measuring fluorescence levels on a flow cytometer (figure 4.1A) (33, 34). In contrast, replacement of Met by Hpg resulted in cells with near-background levels of fluorescence (figure 4.1B), suggesting that the substitution of the alkyne amino acid for Met greatly reduced or eliminated the binding function of the displayed scFv. This apparent loss of binding activity served as a starting point for investigating whether anti-digoxin scFv variants that function in the Hpg context could be identified using directed evolution.

Screening for Hpg-tolerant scFv variants. We used flow cytometry to investigate whether Hpg-tolerant variants of Base could be isolated from a library of scFvs. Error-prone PCR was employed to introduce mutations throughout the majority of the Lpp-OmpA'-scFv gene (excluding only the N-terminal Lpp portion and C-terminal histidine tag of the gene), and the resulting genetic mutants were used to construct a library (Lib1_1a) consisting of approximately 5.3×10^5 transformants. Introducing mutations throughout the construct allowed us to simultaneously explore the possibilities that either i) display of the scFv is impaired by replacing Met with Hpg in the display anchor, or ii) the binding

function of the scFv is impaired upon substitution of Met by Hpg. The display anchor contains four Met residues (including the initiator Met), and Base codes for four additional Met residues. Although few functional binders appeared to be present in the naïve library when Met was replaced by Hpg (figure 4.2), screening the library under expression conditions in which 20 μM Met was added to the Hpg expression medium (16) enabled isolation of a population of mutants tolerant of at least some level of Hpg substitution. Two more rounds of sorting for cells able to bind to **5**, in which the population to be screened was expressed under conditions of near-complete replacement of Met by Hpg, resulted in the isolation of a population of cells able to bind to substantial amounts of labeled antigen in the Hpg context (figure 4.2).

Individual clones were isolated, tested for their ability to bind antigen in the Hpg context when displayed on cells using flow cytometry, and sequenced. All ten clones randomly selected from the population were functional when Met was replaced with Hpg, and every clone contained at least one mutation in the scFv portion of the protein (table 4.1). The majority of observed mutations involve either the elimination or addition of a Met codon within the scFv, and the mutation of the Met at position 80 of the scFv heavy chain (H80, Kabat numbering) to leucine appears in every fragment sequenced. Another striking result is the frequent isolation of clones containing the same set of three mutations involving Met (M(H20)I, M(H80)L, and L(H82C)M); these five clones were found to be identical at the genetic level and will be referred to as Mut2. While the majority of sequenced clones contained at least one amino acid mutation in the display anchor portion of the construct, all isolated clones bound to comparable amounts of digoxigenin probe regardless of whether the mutants contained amino acid mutations in their display anchors.

A complete list of every amino acid mutation observed in sequenced clones can be found in table 4.2.

After three rounds of sorting, the isolated population (referred to here as Lib1_1a Hpg3x; see (35) for additional nomenclature) retained binding function in the Met context while gaining the ability to bind to antigen in a new amino acid context. These results suggested that perhaps isolated clones would retain their binding activity after replacement of Met with other ncAAs in addition to Hpg. In order to examine this possibility, we studied Mut2's ability to bind to digoxigenin after replacement of Met with Hpg, Aha, and Nrl. Flow cytometry revealed that cells displaying Mut2 consistently exhibit higher levels of fluorescence than Base in all of the amino acid contexts investigated (figure 4.1), confirming the improved function of this clone in ncAA contexts, particularly in the Hpg and Aha contexts. Because Mut2 was isolated in the Hpg context in screens aiming only to maximize the fluorescence of cells displaying scFvs, we suspected that further improvement of the kinetic properties of clones in multiple amino acid contexts might be possible with the generation and screening of an additional library of scFv variants.

Screening in multiple ncAA contexts for variants with improved binding function. Lib1_1a Hpg3x was amplified using additional error-prone PCR and then used to construct a library consisting of 5.6 million transformants, Lib2 (see materials and methods). Lib2 was subjected to screening after replacement of Met with Hpg, Aha, or Nrl. The use of multiple ncAAs during screening allowed us to investigate how different analogs affect scFv properties, and whether these effects would be large enough to result in context-dependent mutations. Four rounds of FACS were performed under increasingly stringent

conditions (table 4.3) after library expression in each amino acid context, including the use of a kinetic competition procedure in the last two rounds of screening (29) in an attempt to isolate clones with the most favorable kinetic properties. After one round of sorting, the majority of clones in the sorted populations exhibited high levels of fluorescence, and the mean fluorescence of the sorted populations remained high after each round of sorting (figure 4.3). Individual clones were picked randomly after three (Hpg and Aha contexts) and four (Hpg, Aha, and Nrl contexts) rounds of enrichment, sequenced, and assessed for digoxigenin binding function using a method to estimate dissociation constants with cells displaying scFvs (29). The majority of clones isolated appeared to have on-cell dissociation rates comparable to or lower than that of Mut2 in the same amino acid context (table 4.4), and these clones exhibit numerous mutations that add or eliminate Met (ncAA) residues, convert small, aliphatic residues to aromatic residues, or convert aromatic residues to small, aliphatic residues (table 4.5). Although the mutations observed were similar regardless of the amino acid context used during screening, the frequencies with which these mutations occurred appeared to be context dependent.

High-throughput sequencing of sorted populations. Differences in the sequences of isolated clones prompted us to more fully investigate whole populations of sorted clones. We used high-throughput sequencing to characterize samples of Lib1_1a Hpg3x, Lib2 Hpg4x, Lib2 Aha4x, and Lib2 Nrl4x. Alignment of the sequencing output from each sample using the DNA sequence of Base as a reference enabled the calculation of frequencies of mutation at each position of the gene. Many positions within the populations were mutated at a frequency greater than five percent in at least one sample (table 4.6), and

a number of these nucleotide changes result in amino acid mutations within the scFv portion of the display construct such as Met additions and eliminations and conversions between small aliphatic and large aromatic residues (figure 4.4 and table 4.6).

Certain positions within the fragment appear to show ncAA context-dependent mutational frequencies in the sorted Lib2 populations (figure 4.4). For example, the observed frequencies of M(H20)I, M(H80)L, and L(H82C)M mutations increase from Lib1_1a Hpg Sort 3 to the Lib2 Hpg4x and Lib2 Aha4x populations, while these mutations are found to have similar or decreased frequencies in the Lib2 Nrl4x population. Mutations at position H24, which do not involve the Met codon, are found in over half of the reads covering this position in Lib2 Aha4x and Lib2 Nrl4x, while they occur at a much lower rate of twenty to thirty percent in Lib2 Hpg4x. We should note that the FACS procedure of isolating bright cells from a population does not ensure that all cells displaying scFvs with favorable binding properties will be isolated during the course of screening, nor does it ensure that bright cells will be isolated with a frequency proportional to their function. However, within our ability to limit these factors by oversampling populations to be sorted by at least tenfold, we observe amino acid–dependent changes in mutational frequencies. These results suggest that the mutations observed here affect scFv function in an amino acid context dependent fashion.

Frequent amino acid mutations. Investigation of previously reported crystal structures of the Fab form of the 26-10 anti-digoxin antibody (32) revealed that most of the mutations identified in this work are located in regions far from the binding pocket of the protein (figure 4.4B). These results are consistent with the high affinity of the parent fragment for

digoxin and the surface complementarity observed between the binding pocket and the steroid portion of the antigen (32). Some of the mutations observed in this study are striking because of the frequent addition or elimination of Met (ncAA) residues within the scFv. Mutations M(H20)I and M(H80)L have been observed in previous engineering experiments with the 26-10 scFv (in the Met context only) (29), albeit only in one out of several sequence variants reported. Examinations of V_H protein sequences from mouse germlines indicate the frequent presence of Met, Leu, and Ile at position H20, and predominantly Met and Leu residues at position H80 (36, 37). Perhaps mutations to convert Met to other commonly occurring residues at these framework positions become more favorable in ncAA contexts due to differences in Met and ncAA side chain properties. In particular, the differences between Met and Hpg side chains are quite striking because the Hpg side chain has one less rotatable bond than Met, Aha, and Nrl due to the presence of the terminal alkyne. The restriction of side chain conformation experienced upon replacement of Met by Hpg may partially explain why the clone Mut2 was isolated so frequently from Lib1_1a. The Hpg side chain has far fewer possible conformations than the Met side chain and may prevent adequate packing of the hydrophobic core upon substitution at positions H20 and H80. The mutation of Leu to Met (Hpg) at position H82C is more surprising given the structural differences between the side chains of Leu and Hpg and the lack of murine germline sequences possessing this mutation (36, 37). Perhaps the unique side chain character of Hpg or the position of this residue within a complementarity determining region (CDR) makes its insertion at position H82C more favorable or tolerable. However, we cannot rule out the possibility that this mutation is not

advantageous given its frequent occurrence in parallel with Met elimination mutations that appear to be much more beneficial to the scFv.

Mutations at positions H20, H80, and H82C seem to be accommodated quite well in the Aha and Nrl contexts. However, their slight deenrichment in the Nrl4x population suggests that these mutations may not be as favorable with the aliphatic side chain compared to their occurrence with other amino acid side chains. The infrequent eliminations of Met observed in sorted populations at positions H34 and H100B show that not all Met positions within the scFv are intolerant of ncAA side chains. In fact, Met(H100B) directly contacts the antigen in the crystal structure of the Fab-digoxin complex (32), suggesting that the substituted ncAAs also directly contact the antigen upon incorporation. This direct contact with the antigen may be responsible for some of the changes in binding behavior observed in soluble scFvs (see below). Mutations at positions H24, H27, H29, and H30 are within a region known as the upper core of the scFv (38) that is disordered within the uncomplexed 26-10 Fab crystal structure (32). Mutations in this region have previously been observed to occur frequently in evolution experiments with the 26-10 scFv (29). Given the importance of residues within this region for packing the upper core, ensuring a stable scFv framework, and ensuring proper orientation of the framework relative to complementarity determining regions (CDRs) (38), it seems plausible that the mutations observed here have beneficial effects on scFv stability or folding. Destabilization or poor protein folding oftentimes accompany ncAA substitutions (15, 16, 39) and may make mutations within the upper core very favorable in the amino acid contexts investigated here.

Kinetic characterization of scFvs produced in soluble form. We further studied the functional characteristics of scFvs containing ncAAs by expressing and purifying a number of mutants in soluble form (figure 4.5, table 4.7). Mutants for further characterization were chosen based on their on-cell dissociation kinetics (table 4.4): we selected the two to three mutants with the best k_{off} values measured on-cell and one mutant exhibiting kinetics within the central range of k_{off} values observed on-cell. Expression and purification resulted in the production of approximately 0.5–5 mg/L pure, monomeric protein for all scFvs produced (table 4.8). Each scFv variant isolated from Lib2 was produced in the ncAA context in which it was isolated and in its Met form, and Mut2 was produced in all four amino acid contexts of interest. Base was also expressed in all four amino acid contexts, but could only be purified in Met and Nrl forms. The extent of replacement of Met by ncAAs was estimated by MALDI mass spectrometry on trypsinized fragments of scFvs and found to be approximately 80%–90% in all samples analyzed (see materials and methods, figure 4.6, and table 4.8).

Systematic characterization of the kinetic properties of soluble scFvs using surface plasmon resonance revealed that the isolated fragments possess a modest range of kinetic characteristics that are influenced by amino acid context and amino acid mutations (table 4.8, table 4.9). These differences manifest themselves almost entirely in the dissociation constants of the fragments. The sensitivity of the kinetics to replacement of Met is evident in the dissociation constant of Mut2. Incorporation of Hpg in Mut2 increases the dissociation rate constant by a factor of four, substitution of Aha lowers (improves) the dissociation constant by a factor of two, and incorporation of Nrl leaves the dissociation constant essentially unchanged. Interestingly, the dissociation constant of Mut2 in the Met

context is the same as that of Base, confirming that screening for scFv mutants functional in the Hpg context led to the isolation of clones retaining function in the methionine context while adapting to this new amino acid. Variants isolated from Lib2 were found to exhibit improved kinetics, with the k_{off} values of the best variants isolated in Hpg and Nrl contexts reduced by twofold. Interestingly, the dissociation rate constants of these clones were also improved in the Met context compared to Mut2 and Base (also by twofold), suggesting that the mutations introduced into these clones are beneficial in multiple amino acid contexts; this trend is also borne out in the kinetic properties of other scFvs characterized in soluble form (table 4.8). The Aha forms of clones isolated in the Aha context had dissociation constants comparable to the k_{off} values of Mut2 containing Aha, and in the case of Aha4x4, the Met form was found to have superior kinetic properties to the Met forms of Mut2 and Base. The lack of observed affinity maturation in the Aha context may be a result of the high affinity of Mut2 and other parent proteins in Aha form.

These kinetic characterizations reveal general trends in the effects of substituting ncAAs for Met in scFvs: i) compared to Met, Hpg is somewhat detrimental to the binding properties of all sequence variants characterized, ii) Aha improves binding affinity modestly, and iii) Nrl leaves the binding properties of scFvs unchanged. In all three ncAA contexts, we were successful in isolating sequence variants with favorable binding properties, demonstrating that sequence spaces containing ncAA analogs of Met still include numerous functional proteins. These results also suggest that affinity maturation experiments in multiple amino acid contexts can result in the isolation of clones with improved functional properties, as demonstrated by our ability to isolate Hpg- and Nrl-containing fragments with lowered dissociation rates.

Chemical modification of scFvs. We attempted to exploit the chemical reactivities of Hpg- and Aha-containing scFvs using copper-catalyzed and copper-free click chemistries. An initial assessment of the surface accessibility (40) of all potential modification sites indicated that many appear to be buried in the interior of the protein (Figure 4.7). In our hands, copper-catalyzed azide-alkyne cycloadditions (CuAAC) of ncAA-containing proteins with fluorescent dyes appear to be inefficient (figure 4.8, table 4.10), with CuAAC on scFvs proceeding in lower yields than CuAAC with control proteins containing Aha or Hpg. These results are in line with previous reports suggesting that surface accessibility is critical for such reactions (41). However, strain-promoted click modification of azide-containing proteins with dibenzocyclooctyne-functionalized Alexa Fluor 488 (DIBO 488, **6**) (42, 43) was more successful (Figure 4.9). Quantification of the extent of modification with the Aha forms of Aha3x2, Aha4x4, and Aha4x5 showed that roughly 0.4 dyes had reacted per protein in the case of Aha3x2 and Aha4x5 (table 4.11). The lack of labeling in the case of Aha4x4 may be explained by the absence of a potential modification site at position H34; this residue has been mutated from Met to isoleucine in Aha4x4. MALDI characterizations of intact and trypsinized protein samples before and after click chemistry confirmed labeling and enabled identification of position H34 as a frequent site of modification (figure 4.9, figure 4.10). The high number of proteins having multiple modifications may result from enhanced reactivity after a single protein site is modified.

Multiple experiments confirmed that proteins modified with **6** retain binding function. Kinetic characterizations revealed no substantial changes to the kinetic properties of chemically modified proteins (table 4.12). Western blotting with the clicked scFvs verified that these conjugates can detect the presence of digoxigenin-labeled proteins

(figure 4.11, figure 4.12). Probing nitrocellulose membranes with the Aha form of **6**-labeled Aha4x5 enabled fluorescence detection of the BSA-Dig samples down to approximately 5 nanograms. Interestingly, when the presence of BSA-Dig was detected through the hexahistidine tag of the Aha and Met forms of Aha4x5, the Aha form enabled fluorescence detection down to 5 nanograms BSA-Dig, while the Met form was only able to detect quantities of 50 nanograms BSA-Dig or greater. The higher sensitivity of the Aha form is consistent with the improved kinetic properties of scFvs containing Aha in place of Met.

Simultaneous detection of scFv function and modification. The successful modification of Aha-containing scFvs with **6** prompted us to investigate the feasibility of using flow cytometry to detect function and modification of cell surface-displayed scFvs. Cells displaying Aha4x5 in Met and Aha forms were subjected to click reactions with biotin cyclooctyne **7** (44) and probed for binding with **5** and modification with streptavidin-phycoerythrin (SA-PE). Control experiments allowed us to establish that i) many copies of the displayed scFvs remain functional after click chemistry, ii) whole-cell labeling with **7** is selective for cells displaying Aha-containing fragments, iii) the majority of signal associated with SA-PE detection of **7** is due to the display of Aha-bearing scFvs and not other cellular proteins, and iv) treating cells with unlabeled digoxin prior to treatment with **5** blocks binding of the fluorescent probe to the scFvs (figure 4.13). Having shown that binding and modification can be simultaneously assessed on whole cells, we used a subset of samples to demonstrate separation of various cellular populations via FACS (figure 4.14). Four-way sorting of a mixture of cells resulted in the isolation of populations of cells

binding antigen and possessing modifications, cells binding antigen without possessing modifications, cells possessing modifications without binding labeled antigen, and cells possessing neither modifications nor binding antigen. These results illustrate the feasibility of using FACS to screen libraries of chemically modified proteins.

Conclusions

The combined use of ncAAs and *E. coli* cell surface display provides a powerful platform for engineering antibody fragments. Our results demonstrate that highly functional scFvs tolerant of ncAA side chains can be isolated from large protein libraries. These proteins appear to have gained functionality after ncAA incorporation while retaining their original functionality and can be thought of as amino acid "generalists," in analogy to how enzymes tend to retain their activity toward an established substrate while evolving new substrate recognition capabilities (45, 46). Previous work has attributed the property of ncAA tolerance to robust folding properties (16, 39). In this work, the frequent mutation of amino acids located far away from the antigen binding pocket and within flexible regions of the protein chain suggests that the protein variants we have isolated may be favored for similar reasons.

We find that the binding properties of scFvs can be affected by the replacement of Met with ncAAs. The side chain properties of Hpg are apparently distinct enough from those of Met that two rounds of directed evolution were insufficient to isolate mutants that function equally well in Met and Hpg contexts. All of the Nrl-containing scFvs investigated here exhibit kinetic properties almost identical to their Met counterparts, suggesting that creating and evolving Nrl-containing proteins may provide an alternative route to

oxidation-resistant proteins for use in therapeutic settings (26-28). The functional consequences of replacing of Met with Aha in anti-digoxin scFvs are striking in that the kinetic properties of the scFvs examined here appear to be improved in their Aha forms. The Met residue that directly contacts the antigen in the Fab-digoxin complex, M(H100B) (32), is conserved in all of the sequence variants characterized in this work. This conservation suggests the incorporation of the azide functionality into the binding pocket may be responsible for the improvements in the kinetic properties we have observed. This view is further supported by the fact that removal of each of the commonly occurring Mets (ncAAs) from two other structural locations (H34, H82C), in the scFv variants Aha4x4 and Aha4x5 does not appear to eliminate the kinetic improvement accompanying substitution of Aha for Met. Employing ncAAs in library designs aimed toward heavily patterning the binding pockets of proteins with combinations of a few amino acids (47) may reveal that unnatural side chains such as that of Aha have unique molecular recognition capabilities. Site-specific incorporation of ncAAs into antibody fragments is known to endow these binding proteins with advantages in recognizing targets where specific molecular interactions are known to be favorable (17-19), but a more expansive exploration of chemical space in binding pockets may yield new classes of binding reagents or general insights into how molecular recognition events are mediated.

The chemical reactivity of scFvs containing ncAAs provides mild routes to the preparation of functional bioconjugates. We achieved reasonably efficient labeling of azide-containing scFvs with strained alkynes even in the absence of azide groups predicted to be surface accessible, and soluble proteins modified in this fashion retain their function. The cell surface display platform utilized to isolate functional sequence variants also

provides a means for monitoring the chemical modification state of scFvs. Our FACS data suggests that libraries of proteins should be able to be screened for proteins that are both functional and amenable to chemical conjugation. The ability to improve protein function and chemistry simultaneously in a high-throughput manner may provide an efficient means to develop protein-small molecule (20, 21) and protein-polymer conjugates (22, 23) with properties useful for biopharmaceutical applications. This approach may be particularly useful for systematically studying how the number and location of modification sites within a protein impact protein function. Genetically encoding and screening libraries of chemically modified proteins may also be applicable to areas including the chemical construction of bi- or multivalent protein structures (48), switchable sensors (49, 50), or cyclized peptides (51, 52).

Materials and Methods

Materials. The plasmids pB18D and pB30D, which encode for the cell surface-displayed anti-digoxin single chain variable fragment (scFv) in pBAD18 and pBAD30 vector backbones, were generous gifts from Professor George Georgiou (31). The plasmid pAK400, which is a standard periplasmic expression vector used with scFvs, was a generous gift from Professor Andreas Plückthun (53). 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 the *E. coli* strain DH10B (54, 55). 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 using the method of Link et al. (56) with minor modifications. BODIPY FL Digoxigenin **5** used during flow cytometry experiments was a generous gift of 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. DIBO 488 **6** and TAMRA-Alkyne were from Invitrogen. Synthesis of the lissamine-rhodamine azide (57) and biotin cyclooctyne **7** (44) have been described previously. Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) and 3-azidopropanol were gifts from Dr. Janek Szychowski. THPTA was synthesized as described previously (58), and 3-azidopropanol was prepared by azide displacement of bromine in 3-bromopropanol. GFPrm_AM (55) in alkyne and azide forms were provided by Alborz Mahdavi. 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 columns 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 EcoRI site in pB30D was eliminated via site-directed mutagenesis using the primer EcoRIElimFwd (sequence given in table 4.13) and its reverse complement EcoRIElimRev. This resulted in the generation of plasmid pJAV2, which was sequence verified. The Lpp-OmpA-scFv fusion was then amplified in two steps from pB18D. In the first step, an internal HindIII 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 using XmaI and HindIII and ligated. After transformation of electrocompetent XL-1 Blue cells and selection on plates containing chlorampenicol, sequences of clones containing the correctly sized insert were verified with 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 fusion with Lpp-OmpA-antidigFwd and LppHisRescue. Double digestion of pQE-80L and the resulting PCR product, purification, and ligation resulted in the construction of the sequence-verified plasmid pQE-80L-antidig-HisGS. 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. This PCR product was gel purified and doubly digested along with pQE-80L, both with the enzymes HindIII and XmaI. The two digest products were gel purified, ligated, transformed into electrocompetent *E. coli* XL-1 Blue cells, and plated on ampicillin plates. Colonies were tested for the proper insert size and sequence verified for the desired restriction site modifications (PstI and BglII), 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.

Libraries were constructed using pQE-80L-antidig-HisGS-Base (Lib1_1a) or the sorted population Lib1_1a Hpg3x (Lib2) as a template for error-prone PCR. The polymerase Mutazyme II from the Stratagene GeneMorph II kit was used to introduce errors throughout the length of 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 base construct were doubly digested using PstI and BglII. After gel purification and ligation, the resulting DNA was used to transform ~200 μL of electrocompetent *E. coli* TYJV2 cells containing pREP4. These cells were rescued for one hour in 20–25 mL super optimized broth with catabolic repression (SOC), followed by inoculation into a large amount (0.25–1.0 L) of 2×YT medium 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 large amounts of culture volume (~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 the first library (Lib1_1a), 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 independent 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 and grown on plates containing ampicillin and kanamycin. Individual colonies were randomly picked from plates, subjected 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 SfiI. 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 one percent glucose. 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. Protein 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 (59). 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 (6000 × g for 7 minutes in a fixed angle rotor or 3000–4000 × g for 10 minutes in a variable angle rotor) and 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 ncAAs as suitable for the given experiments. These cells were induced with 1 mM IPTG and allowed to grow at 25 °C for six hours. At the end of this time period, cells were pelleted and prepared for flow cytometry according to procedures described below.

Flow cytometry, Lib1_1a. 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 at 37 °C. This culture was then diluted 1:100 into fresh M9 KA Glucose 20AA (~10 mL) the next morning. In the initial round of screening, equal volumes of transformations 1 and 1a were mixed, diluted 1:30 into 60 mL M9 KA Glucose 20AA, and grown at 37 °C. After reaching suitable OD_{600}s (0.7–0.8) cells were shifted into new medium as described above. Cells bearing the plasmid coding for Base were aliquoted (3–4 mL/aliquot) and induced with 1 mM IPTG in M9 KA Glucose containing 20 amino acids or 19 amino acids (−Met) supplemented with 2.5 mM Hpg. Cells bearing Lib1_1a were aliquoted into several small samples and induced in medium containing twenty canonical amino acids or 19 amino

acids (−Met) with 2.5 mM Hpg supplemented with 0, 10, 20, or 30 μM Met. Upon completion of the expression (6 hours), all 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 $Na_2HPO_4•H_2O$, 0.20 g/L KH_2PO_4), and resuspended in PBS to an OD_{600} of 1.0. Cells were then treated with 200 nM **5** in 0.25–1.0 mL aliquots for at least 45 minutes with gentle agitation at room temperature, diluted 1:20 into PBS, and filtered using 25 mm filters containing 5 μm Acrodisc Supor membranes (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 Met resulted cells exhibiting high fluorescence after exposure to **5**, while cells expressed in 2.5 mM Hpg exhibited low fluorescence levels after exposure to **5**, confirming that the medium shift was successful. Lib1_1a expressed in 2.5 mM Hpg and 20 μM Met 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 h at 37 °C in 2.5 mL SOC, diluted with 10.5 mL 2×YT medium 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 remainder of the rescued cells was miniprepped in order to isolate the plasmid DNA from the sort. Subsequent rounds of expression, scanning, and Lib1_1a 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 2.5 mM Hpg (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 one hour with SOC and plated on agar containing ampicillin and kanamycin. Ten colonies were chosen randomly for sequencing and characterization. Following inoculation into overnight cultures of M9 KA Glucose 20AA grown 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 1 mM IPTG in 2.5 mM Hpg for six hours, and harvested. Cells were treated with 200 nM **5** 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, Lib2. Rounds one and two of Lib2 sorting in various amino acid contexts were performed essentially as described for sorting Lib1_1a. Aliquots of the library were thawed and diluted 1:30 into fresh M9 KA Glucose 20AA, grown, shifted into fresh medium, and subjected to six-hour induction of protein expression in medium containing either 0.27 mM Met, 2.5 mM Hpg, 0.28 mM Aha, or 0.30 mM Nrl. In round 1, cells were sorted in Purify 1 mode after treatment with 200 nM **5**, dilution, and filtering. A total set of events greater 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 100 nM **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 as before. After completion of the induction period, 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 expressed. The library fraction was treated with 100 nM **5**, with control samples (i.e., library fractions expressed in Met and Base in Met and Hpg) resuspended and treated with 100 nM **5** 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 exclude dead cells from isolation. Samples to be sorted were treated with 100 nM **5** for at least 45 minutes before being washed twice in PBS, diluted, filtered, and scanned. Cells were also treated with 5 μM propidium iodide (PI) 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 (29). 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 to a final concentration of 2 μM. Total competition duration was determined empirically by monitoring the competition progress with scans of the more dilute cells every five 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. Again, 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. These transformed cells were then allowed to grow on agar plates supplemented with ampicillin and kanamycin. Individual colonies were picked and used to inoculate overnight liquid cultures in M9 KA Glucose 20AA and grown 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 100 nM **5** for at least 45 min at an OD_{600} of 0.2 and stored on ice. While on ice, all samples to be used in kinetic dissociation experiments (30) 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 1 μM unlabeled digoxin competitor, mixed, and scanned every 60 seconds over a period of ten minutes.

Data from timed scans was analyzed using FlowJo (Tree Star, Ashland, OR) and Excel (Microsoft, Redmond, WA) using a previously described approach (30). 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 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 min. The resulting background-corrected, normalized data were fit to a first-order exponential equation of the form

$$F = e^{-k_{off} t}, \quad (4.1)$$

where F is the relative fluorescence of the sample compared to the sample at time t = 0 min and k_{off} is the dissociation rate of the antibody fragment. 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 was analyzed using the program Igor (Wavemetrics, Lake Oswego, OR), and the errors reported are the 95% confidence intervals of the fits to equation (4.1).

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 as described above and induced with 1 mM IPTG after using the standard cell surface display medium shift procedure, with induction performed in medium containing either 0.27 mM Met or 1 mM Aha. 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 six-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 100 μM biotin cyclooctyne **7** and incubated at 37 °C for 16 hours (60). 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 100 nM **5**, a 1:100 dilution of 0.2 mg/mL streptavidin-phycoerythrin (SA-PE), or both dyes simultaneously for a period of 45 minutes or longer. Cells displaying the Aha form of Aha4x5 were also treated with 10 μM digoxin 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 with scans of many different samples (see Figure 4.13). Once the ability to simultaneously monitor antigen binding and chemical modification was established, a model sort was set up by combining four populations of cells bearing the plasmid coding for Aha4x5 that were all reacted with **7** and exposed to **5** and SA-PE: 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. Elliptical regions were established in the two-dimensional dot plot of BODIPY versus phycoerythrin fluorescence intending to capture populations of cells with interesting 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 cells bearing individual scFv sequence variants in the pAK400 backbone were inoculated into 2 mL M9 minimal medium containing twenty amino acids and 0.4% glucose supplemented with chloramphenicol (M9 Chlor Glucose 20AA). Cells were allowed to grow at 30 °C for four or more hours, followed by 1:20 dilution into 20–40 mL fresh M9 Chlor Glucose 20AA and growth at 30 °C overnight. Saturated cultures were then diluted 1:20 into 0.25–1.0 L M9 Chlor Glucose 20AA and grown at 30 °C until reaching an OD_{600} of approximately 0.9–1.0. At this time, the cells were pelleted (15 min at 5000 × *g*, 4 °C), washed three times in ice-cold 0.9% NaCl, and resuspended in fresh M9 medium containing 19 amino acids (minus Met), 0.4% glycerol, and chloramphenicol (M9 Chlor 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 thirty minutes. Cultures were then induced with 1 mM IPTG and allowed to grow for four hours at 25 °C.

At completion of expression, cells were pelleted (15 min at 5000 × *g*, 4 °C) and resuspended vigorously in 32 mL per liter culture volume of ice-cold 0.75 M sucrose, 0.1 M Tris, pH 8.0 (Tris/sucrose buffer). 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 (61) was used to isolate the periplasmic fractions of

cells, which contain the expressed scFvs. After resuspension, the Tris/sucrose buffer containing cells was supplemented with 3.2 mL per liter culture volume 10 mg/mL ice-cold lysozyme in Tris/sucrose buffer. Pellets were rotated at 140 RPM on ice during dropwise addition of 64 mL per liter culture volume of 1 mM ethylenediaminetetraacetic acid (EDTA). After a ten-minute incubation (still with rotation), 3 mL per liter culture volume 0.5 M magnesium chloride was added dropwise to the cell suspension. After another ten-minute incubation period, the samples were pelleted (10,000 × *g*, 20 minutes, 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 25 mm Acrodisc Supor filters with 5 μm pores and incubated for multiple 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 flowthrough was collected. After two washes using 20 mL per liter culture volume Ni-NTA wash buffer (0.3 M NaCl, 0.05 M NaH_2PO_4, 0.02 M imidazole, pH 8.0), samples were eluted using 4 × 4.0 mL per liter culture volume Ni-NTA elution buffer (0.3 M NaCl, 0.05 M NaH_2PO_4, 0.25 M imidazole, pH 8.0). 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 refrigerated at 4 °C. A 1.0 mL HiTrapQ XL ion exchange column 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 approximately 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 using HBS +EDTA at a flow rate of 1 mL/min. 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 sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. After Ni-NTA, ion exchange, and size exclusion chromatography, all samples were found to be greater than ninety percent 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 multiple 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. 20–40 μL of sample was reduced by adding tris(2-carboxyethyl)phosphine (TCEP) to a final concentration of 3.75 mM and incubating the sample at room temperature for 15 minutes. In the case of azide-containing samples, dithiothreitol (DTT) was used to reduce some samples by adding the reducing agent to a final concentration of 2.5 mM and incubating samples at 55−60 °C for 15 minutes. All reduced samples were then alkylated by adding iodoacetamide to a final concentration of 12.5 mM and incubation at room temperature in the dark for 15 minutes. Following alkylation, 10 volumes of 50 mM ammonium bicarbonate, pH 7.8, were added to each solution along with 2.5–5 μL of trypsin (Promega, 0.1 μg/μL), and cleavage was allowed to proceed overnight at 37 °C. Trifluoroacetic acid (TFA) was added to a final concentration of 0.1% to each cleaved sample, 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 complexed with α-cyano-4-hydroxycinnamic acid, dried on a MALDI target, and assayed using 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. Azide-containing peptides were observed to frequently lose dinitrogen and gain two hydrogen atoms during acquisition of MALDI mass spectra (62).

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 solution of 8 M Urea, 0.1 M Tris, pH 8.0. To each solution was added ten volumes of 50 mM ammonium bicarbonate, pH 7.8, followed by adjustment of the sample to 0.1% TFA. 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** was added to a final concentration of 10 μM to 1.75 μM solutions of Aha- or Met-

containing scFvs in HBS +EDTA buffer. Solutions were vortexed briefly, and reaction was allowed to proceed for one hour at room temperature. Reactions were quenched by adding 3-azidopropanol to a final concentration of 10–20 mM and vortexing. Copper-catalyzed azide-alkyne cycloadditions (CuAAC) were performed with reference to the conditions outlined by Hong et al. (58). All reactions were performed on protein solutions in PBS, pH 7.4, having concentrations of roughly 1.0 to 1.75 μM protein. scFv protein samples were buffer exchanged into PBS prior to reaction using one or two Zeba Spin desalting columns (7000 Dalton MWCO) from Pierce. Modification of alkyne-containing proteins was performed using a lissamine rhodamine azide dye described previously (57), and modification of azide-containing proteins was performed with TAMRA-alkyne (Invitrogen). Final reaction mixtures contained 100 μM cupric sulfate, 500 μM THPTA, 5 mM aminoguanidine, and 5 mM sodium ascorbate (added last). Fluorescent dyes were preincubated with cupric sulfate and THPTA and added to the reactions at final concentrations of 20–200 μM (azide-containing proteins) or 20 μM (alkyne-containing proteins). Reaction mixtures were capped, vortexed briefly, and incubated for one hour at room temperature. All reactions were quenched with 10 mM 3-azidopropanol. GFPrm_AM (55) in azide and alkyne forms was used as a positive control.

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 the presence of 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. Gels 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 a single batch 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, approximately 10.2 mg of BSA was dissolved in 1 mL 0.1 M NaH_2CO_3, pH 9.0 and stirred in a scintillation vial. A small amount (≤ 1 mg) of 3-amino-3-deoxydigoxigenin hemisuccinamide, succinimidyl ester dissolved in 150 μL *N,N*-dimethylformamide was added dropwise to the BSA solution. The mixture was stirred at room temperature for five minutes, diluted to 2.5 mL using PBS, pH 7.4, 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, pH 7.4.

Detection of the presence 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 onto nitrocellulose membranes. Membranes were blocked for at least one hour in PBS, pH 7.4, containing 0.1% TWEEN-20 and either 3% (w/v) BSA or 5% (w/v) powdered milk solution (Nestle, Glendale, CA). Membranes were washed in PBS, pH 7.4 containing 0.1% TWEEN-20 (PBS-TWEEN) before exposure to labeled scFvs. 100–250 μL of 1.75 μM antibody fragment samples reacted with compound **6** as described above were diluted into approximately 10 mL of PBS-TWEEN containing either 3% BSA or 5% powdered milk and exposed to membranes for approximately 1 hour. Membranes were washed with multiple changes of PBS-TWEEN and imaged using the Typhoon Trio. In some cases, membranes were placed back into PBS-TWEEN containing 5% milk and exposed to Alexa Fluor 647-labeled anti-penta-his antibodies added to solution at a 1:10,000 dilution for one hour, washed, and imaged. These experiments enabled secondary detection of digoxigenin using the histidine tag of the scFvs and served to probe the binding function of scFvs regardless of whether the antibody fragments in question appeared to be labeled with fluorescent dye.

Kinetic characterizations 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 used to assay kinetics contained one flow 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 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 reacted with **6**) were diluted to the final concentrations need for kinetic assays using HBS EP+, without dialysis or other buffer exchange into the Biacore running buffer.

Before assaying experimental samples, and in between different scFv samples, the chip used in assays 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 (63). 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 ~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 (64). 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 Lib1_1a Hpg3x, Lib2 Hpg4x, Lib2 Aha4x, and Lib2 Nrl4x were thawed, diluted 1:200 into 200 mL 2×YT KA, and grown at 37 °C until the OD_{600}s of each culture exceeded 1.0. The cultures were pelleted (5000 × *g*, 15 min, 4 °C), decanted, and frozen at −20 °C for at least 12 hours. Plasmid DNA from each sorted population was then isolated using a Maxiprep kit. All DNA was digested using restriction enzymes BglII and PstI, 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. Approximately 1–3 μg DNA of each library population 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 30 million sequencing reads on an estimated 3.5–4.2 $\times$ 10^5 clusters. All data was aligned to the DNA sequence of Base, and the total number of calls of A, C, G, T, and N, at every position of each population were tallied using a script developed at the genomics laboratory. Every position within the construct was read at least 8,500 times in all populations. Substantial variations in the total numbers of reads at each position were observed (ranging from approximately 8,500 to greater than 1 million), most likely due to nonuniform fragmentation of the restriction fragment. However, even the minimum coverage of 8500 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 was 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 Base with the additional criteria 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.

Acknowledgements

We thank Mona Shagoli for MALDI mass spectrometry assistance at the Caltech Chemistry Mass Spectroscopy facility, Igor Antoshechkin and the Millard and Muriel Jacobs Genetics and Genomics Laboratory for assistance with high-throughput sequencing, Cynthia Shuman of GE Healthcare for Biacore kinetic assay guidance, Dr. Janek Szychowski and Alborz Mahdavi for materials, and Dr. Maren Buck for helpful comments on the manuscript. This work was supported by NIH grant GM62523. J. A. V. was supported in part by a National Defense Science and Engineering Graduate (NDSEG) fellowship, and K. P. Y. was supported in part by a National Science Foundation (NSF) Graduate Fellowship.

References

1. Jackel C, Kast P, & Hilvert D (2008) Protein design by directed evolution. *Ann. Rev. Biophys.* **37,** 153-173.
2. Romero PA & Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10,** 866-876.
3. Brustad EM & Arnold FH (2011) Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* doi:10.1016/j.cbpa.2010.1011.1020.
4. Johnson JA, Lu YY, Van Deventer JA, & Tirrell DA (2010) Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14,** 774-780.
5. Voloshchuk N & Montclare JK (2010) Incorporation of unnatural amino acids for synthetic biology. *Mol. Biosyst.* **6,** 65-80.

6. Liu CC & Schultz PG (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413-444.

7. Dougherty DA (2009) In Vivo Studies of Receptors and Ion Channels with Unnatural Amino Acids. *Protein Eng.,* Nucleic Acids and Molecular Biology, eds Kohrer C & RajBhandary UL (Springer-Verlag, Berlin), Vol 22, pp 231-254.

8. Van Deventer JA, Fisk JD, & Tirrell DA (2011) Homoisoleucine: a translationally active leucine surrogate of expanded hydrophobic surface area. *ChemBioChem* **12,** 700-702.

9. Kwon OH*, et al.* (2010) Hydration dynamics at fluorinated protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17101-17106.

10. Edwardraja S, Sriram S, Govindan R, Budisa N, & Lee SG (2011) Enhancing the thermal stability of a single-chain Fv fragment by in vivo global fluorination of the proline residues. *Mol. Biosyst.* **7,** 258-265.

11. Wolschner C*, et al.* (2009) Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7756-7761.

12. Tang Y & Tirrell DA (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J. Am. Chem. Soc.* **123,** 11089-11090.

13. Mendel D*, et al.* (1992) Probing protein stability with unnatural amino acids. *Science* **256,** 1798-1802.

14. Lummis SCR*, et al.* (2005) Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438,** 248-252.

15. Montclare JK & Tirrell DA (2006) Evolving proteins of novel composition. *Angew. Chem. Int. Ed.* **45,** 4518-4521.

16. Yoo TH, Link AJ, & Tirrell DA (2007) Evolution of a fluorinated green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 13887-13890.

17. Liu CC, Choe H, Farzan M, Smider VV, & Schultz PG (2009) Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* **48,** 8891-8898.

18. Liu CC*, et al.* (2008) Protein evolution with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 17688-17693.

19. Liu CC*, et al.* (2009) Evolution of proteins with genetically encoded "chemical warheads." *J. Am. Chem. Soc.* **131,** 9616-9617.

20. Alley SC, Okeley NM, & Senter PD (2010) Antibody-drug conjugates: targeted drug delivery for cancer. *Curr. Opin. Chem. Biol.* **14,** 529-537.

21. Ducry L & Stump B (2010) Antibody-Drug Conjugates: Linking Cytotoxic Payloads to Monoclonal Antibodies. *Bioconj. Chem.* **21,** 5-13.

22. Harris JM & Chess RB (2003) Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* **2,** 214-221.

23. Haag R & Kratz F (2006) Polymer therapeutics: Concepts and applications. *Angew. Chem. Int. Ed.* **45,** 1198-1215.

24. de Graaf AJ, Kooijman M, Hennink WE, & Mastrobattista E (2009) Nonnatural amino acids for site-specific protein conjugation. *Bioconj. Chem.* **20,** 1281-1295.

25. Sletten EM & Bertozzi CR (2009) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* 6974-6998.

26. Nguyen TH (1994) Oxidation degradation of protein pharmaceuticals. *Formulation and Delivery of Proteins and Peptides,* Acs Symposium Series, eds Cleland JL & Langer R (Amer Chemical Soc, Washington), Vol 567, pp 59-71.

27. Chu JW*, et al.* (2004) A comprehensive picture of non-site specific oxidation of methionine residues by peroxides in protein pharmaceuticals. *J. Pharm. Sci.* **93,** 3096-3102.

28. Liu D*, et al.* (2008) Structure and stability changes of human IgG1 Fc as a consequence of methionine oxidation. *Biochemistry* **47,** 5088-5100.

29. Daugherty PS, Chen G, Iverson BL, & Georgiou G (2000) Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2029-2034.

30. Daugherty PS, Chen G, Olsen MJ, Iverson BL, & Georgiou G (1998) Antibody affinity maturation using bacterial surface display. *Protein Eng.* **11,** 825-832.

31. Daugherty PS, Olsen MJ, Iverson BL, & Georgiou G (1999) Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface. *Protein Eng.* **12,** 613-621.

32. Jeffrey PD*, et al.* (1993) 26-10 Fab-digoxin complex - affinity and specificity due to surface complementarity. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 10310-10314.

33. Schildbach JF*, et al.* (1991) Altered hapten recognition by 2 antidigoxin hybridoma variants due to variable region point mutations. *J. Biol. Chem.* **266,** 4640-4647.

34. The 26-10 antibody is known to recognize digoxin through the lactone and cardenolide portions of the hapten. Removal of the sugar residues (resulting in the

compound digoxigenin) does not affect the affinity of the antibody for the hapten (33).

35. The sorted populations will be named using the convention Lib* AA#x, where * refers to the library (1_1a or 2), AA refers to the amino acid context in which clones were expressed prior to sorting, and # refers to the number of times the population was sorted from the original library. For example, the thrice-sorted Lib2 population from the Hpg context is called "Lib2 Hpg3x." Single clones isolated from Lib2 will be designated with an additional number at the end of the population name. For example, the 5th clone isolated from the 4th sort in the Nrl context will be referred to as "Nrl4x5."

36. Almagro JC, Hernandez I, Ramirez MD, & Vargas-Madrazo E (1998) Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. *Immunogenetics* **47,** 355-363.

37. de Bono B, Madera M, & Chothia C (2004) V-H gene segments in the mouse and human genomes. *J. Mol. Biol.* **342,** 131-143.

38. Ewert S, Honegger A, & Pluckthun A (2004) Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. *Methods* **34,** 184-199.

39. Nagasundarapandian S*, et al.* (2010) Engineering Protein Sequence Composition for Folding Robustness Renders Efficient Noncanonical Amino acid Incorporations. *ChemBioChem* **11,** 2521-2524.

40. Fraczkiewicz R & Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319-333.

41. van Kasteren SI*, et al.* (2007) Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446,** 1105-1109.

42. Ning XH, Guo J, Wolfert MA, & Boons GJ (2008) Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angew. Chem. Int. Ed.* **47,** 2253-2255.

43. Debets MF*, et al.* (2010) Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun.* **46,** 97-99.

44. Link AJ & Tirrell DA (2003) Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition. *J. Am. Chem. Soc.* **125,** 11164-11165.

45. Aharoni A*, et al.* (2005) The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37,** 73-76.

46. Khersonsky O & Tawfik DS (2010) Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79,** 471-505.

47. Fellouse FA, Barthelemy PA, Kelley RF, & Sidhu SS (2006) Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J. Mol. Biol.* **357,** 100-114.

48. Enever C, Batuwangala T, Plummer C, & Sepp A (2009) Next generation immunotherapeutics-honing the magic bullet. *Curr. Opin. Biotechnol.* **20,** 405-411.

49. Mills JH, Lee HS, Liu CC, Wang JY, & Schultz PG (2009) A Genetically Encoded Direct Sensor of Antibody-Antigen Interactions. *ChemBioChem* **10,** 2162-2164.

50. Sloan DJ & Hellinga HW (1998) Structure-based engineering of environmentally sensitive fluorophores for monitoring protein-protein interactions. *Protein Eng.* **11,** 819-823.

51. Heinis C, Rutherford T, Freund S, & Winter G (2009) Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **5,** 502-507.

52. Sako Y, Morimoto J, Murakami H, & Suga H (2008) Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* **130,** 7232-7234.

53. Krebber A*, et al.* (1997) Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods* **201,** 35-55.

54. Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640-6645.

55. Yoo TH & Tirrell DA (2007) High-throughput screening for Methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. *Angew. Chem. Int. Ed.* **46,** 5340-5343.

56. Link AJ, Vink MKS, & Tirrell DA (2007) Preparation of the functionalizable methionine surrogate azidohomoalanine via copper-catalyzed diazo transfer. *Nat. Protoc.* **2,** 1879-1883.

57. Beatty KE & Tirrell DA (2008) Two-color labeling of temporally defined protein populations in mammalian cells. *Biorg. Med. Chem. Lett.* **18,** 5995-5999.

58. Hong V, Presolski SI, Ma C, & Finn MG (2009) Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem. Int. Ed.* **48,** 9879-9883.

59. Link AJ, Vink MKS, & Tirrell DA (2004) Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* **126,** 10598-10602.

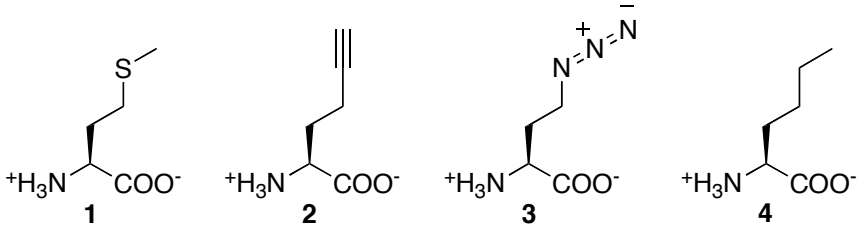
60. Link AJ*, et al.* (2006) Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10180-10185.

61. Hayhurst A*, et al.* (2003) Isolation and expression of recombinant antibody fragments to the biological warfare pathogen Brucella melitensis. *J. Immunol. Methods* **276,** 185-196.

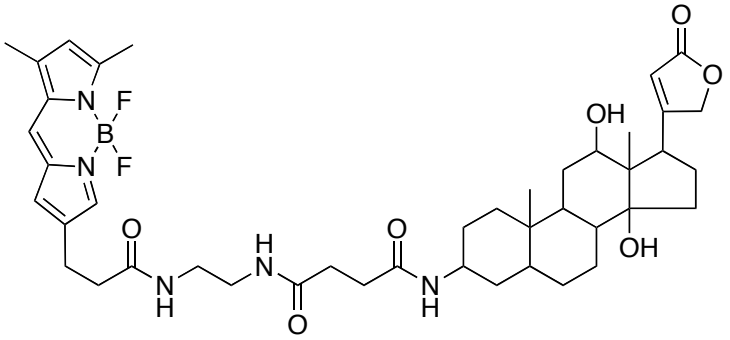
62. Kiick KL, Saxon E, Tirrell DA, & Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 19-24.

63. Karlsson R (1999) Affinity analysis of non-steady-state data obtained under mass transport limited conditions using BIAcore technology. *J. Mol. Recognit.* **12,** 285-292.

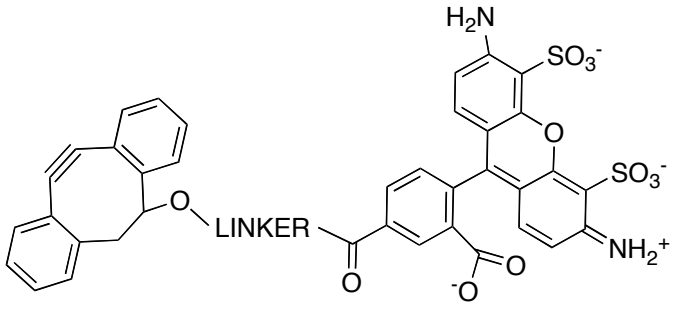
64. Onell A & Andersson K (2005) Kinetic determinations of molecular interactions using Biacore - minimum data requirements for efficient experimental design. *J. Mol. Recognit.* **18,** 307-317.



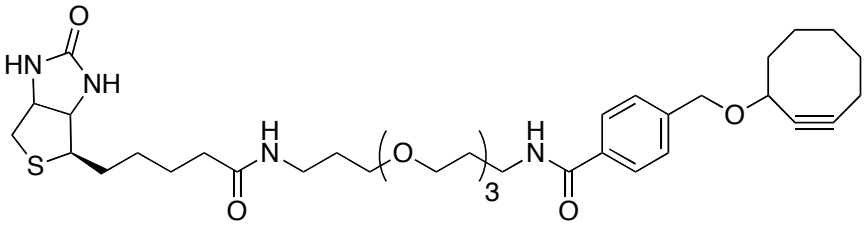
1 **2** **3** **4**



5

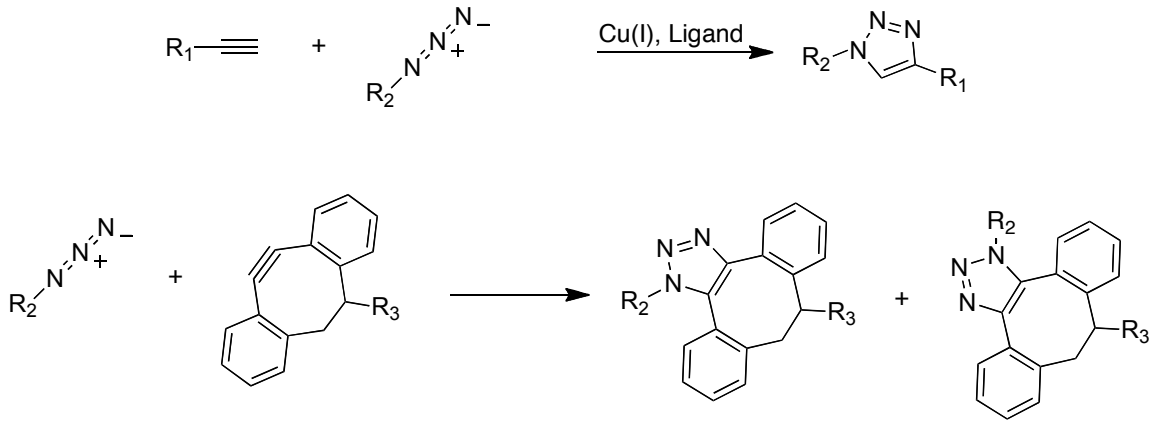


6



7

Scheme 4.1. Compounds used in study. **1**, methionine (Met). **2**, homopropargylglycine (Hpg). **3**, azidohomoalanine (Aha). **4**, norleucine (Nrl). **5**, BODIPY FL Digoxigenin. **6**, Alexa Fluor 488 dibenzocyclooctyne (DIBO 488). The structure of the linker has not been disclosed by Invitrogen. **7**, biotin cyclooctyne.



Scheme 4.2. Chemistries used in modifying azide- and alkyne-containing ncAAs. Copper-catalyzed azide-alkyne cycloaddition reaction (CuAAC, top) and strain-promoted cycloaddition reaction (bottom).

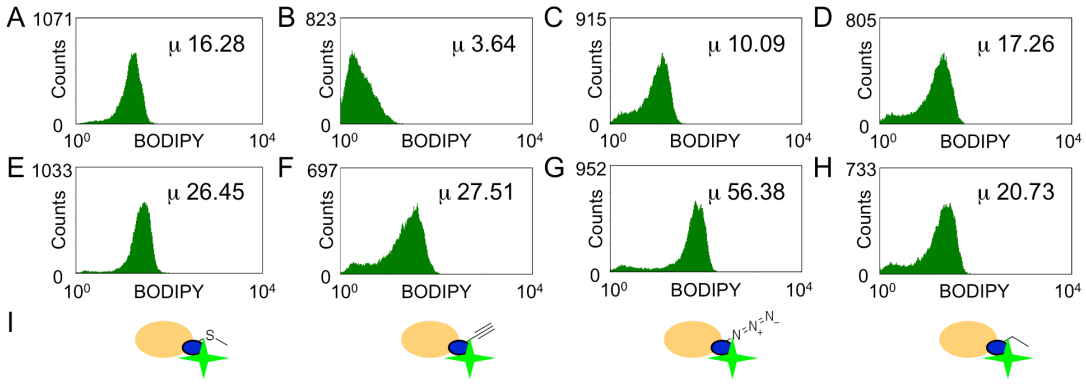


Figure 4.1. Flow cytometry studies of the binding properties of cell surface displayed scFvs. (*A–D*) Base construct expression in Met (*A*), Hpg (*B*), Aha (*C*), and Nrl (*D*) contexts, followed by exposure of cells to fluorescently labeled antigen **5**. (*E–H*) Mut2 construct expression in Met (*E*), Hpg (*F*), Aha (*G*), and Nrl (*H*) contexts, followed by exposure of cells to fluorescently labeled antigen **5**. (*I*) Schematic representation of cells displaying scFvs containing Met or ncAA analogs of Met and binding to labeled antigen. μ, mean fluorescence.

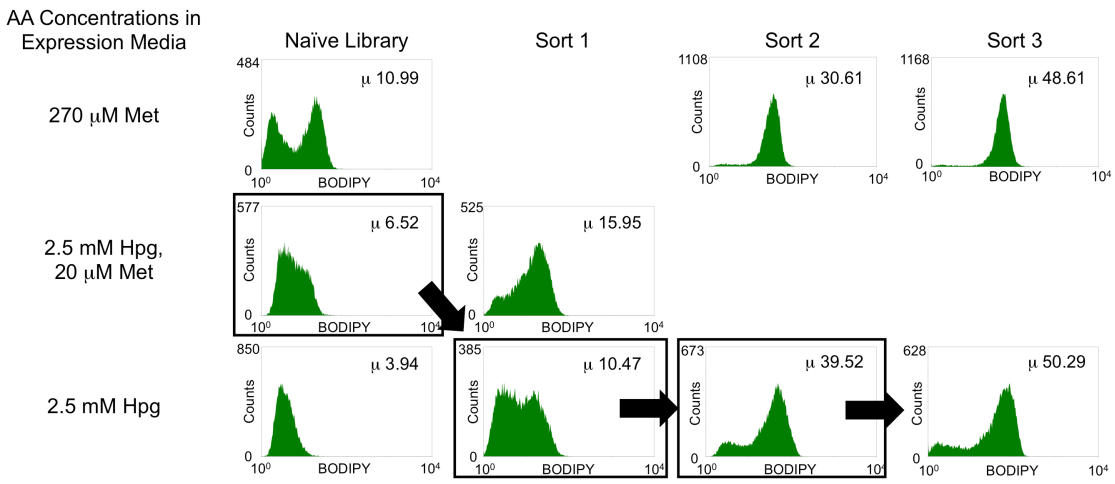


Figure 4.2. Fluorescence activated cell sorting of Lib1_1a for clones that function when Met is replaced by Hpg. Expression of cells harboring the naïve library and sorted populations was performed in minimal media containing concentrations of Met and Hpg as indicated. After treatment of induced cells 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 channel were retained for the next round of sorting. Note that the naïve library was sorted after induction of expression in medium containing a small amount of Met in addition to Hpg (16). AA, amino acid. μ, Mean fluorescence.

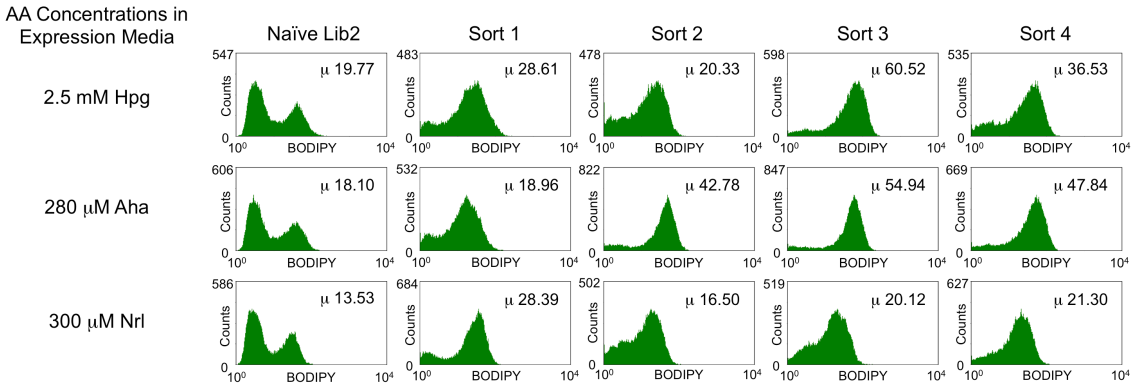


Figure 4.3. Fluorescence activated cell sorting of Lib2 for clones that function when Met is replaced by ncAAs. Expression of the naïve library and sorted populations was performed in minimal media containing concentrations of ncAAs as indicated. The fluorescence of each population was measured on the flow cytometer. 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 4.3; the top ~0.1%–1.0% of fluorescent events were retained. AA, amino acid. μ, mean fluorescence.

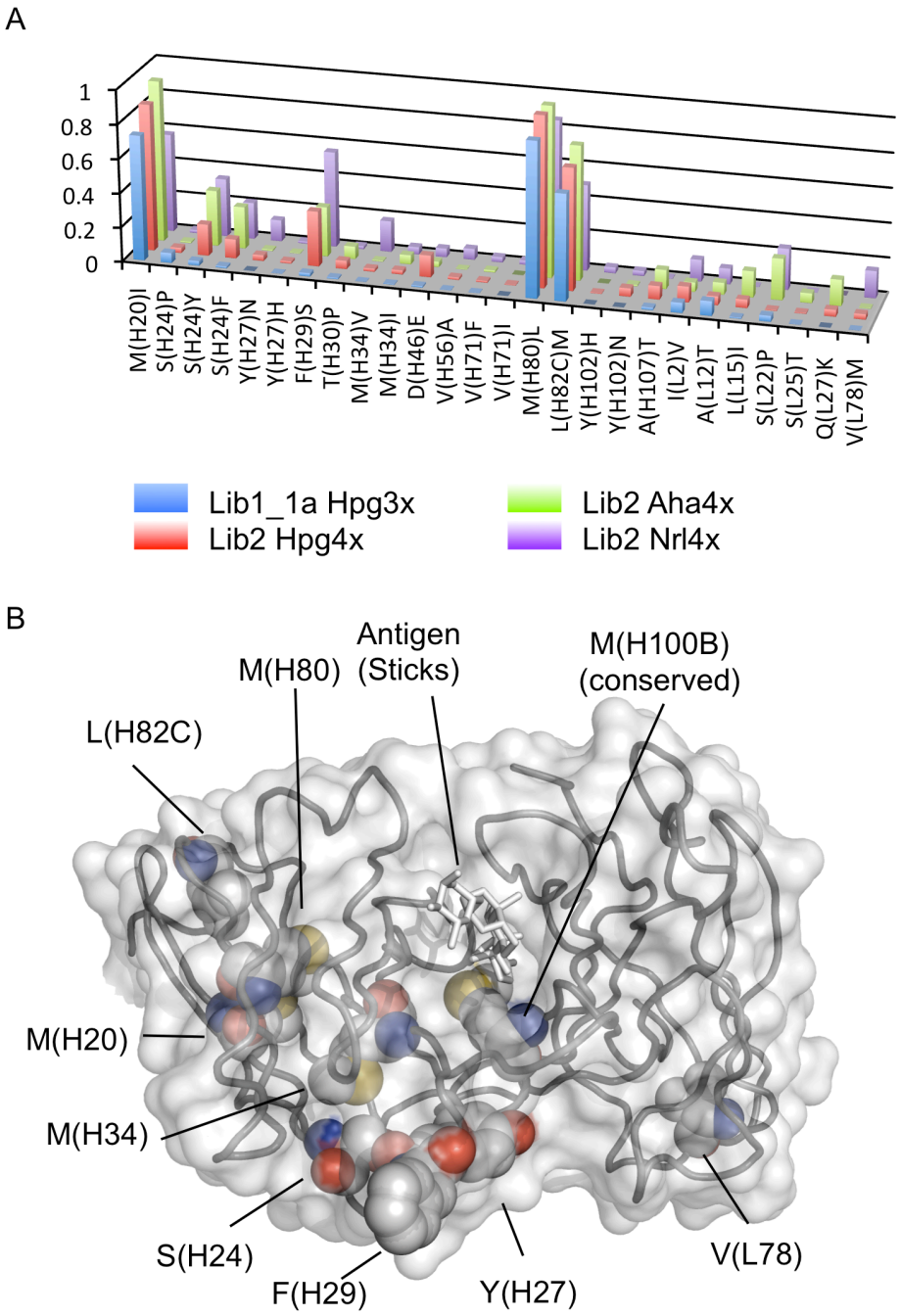


Figure 4.4. Population-level sequence characterization of scFv mutants using high-throughput sequencing. (*A*) Frequently mutated amino acids in sorted scFv populations (Kabat numbering). (*B*) Structural positions of mutations to scFv in crystal structure of 26-10 Fab-digoxin complex solved by Jeffrey et al. (32). The backbone of the Fv portion of

the Fab is outlined with a ribbon diagram (gray), with amino acids of interest shown as space filling and the antigen shown as sticks in white. Unless otherwise noted, labeled residues are frequently mutated. H, heavy chain. L, light chain. Space filling colored by elements: carbon, gray; oxygen, red; nitrogen, blue; sulfur, yellow. Structure produced from the “A” and “B” chains of PDB structure 1IGJ with MacPyMOL.

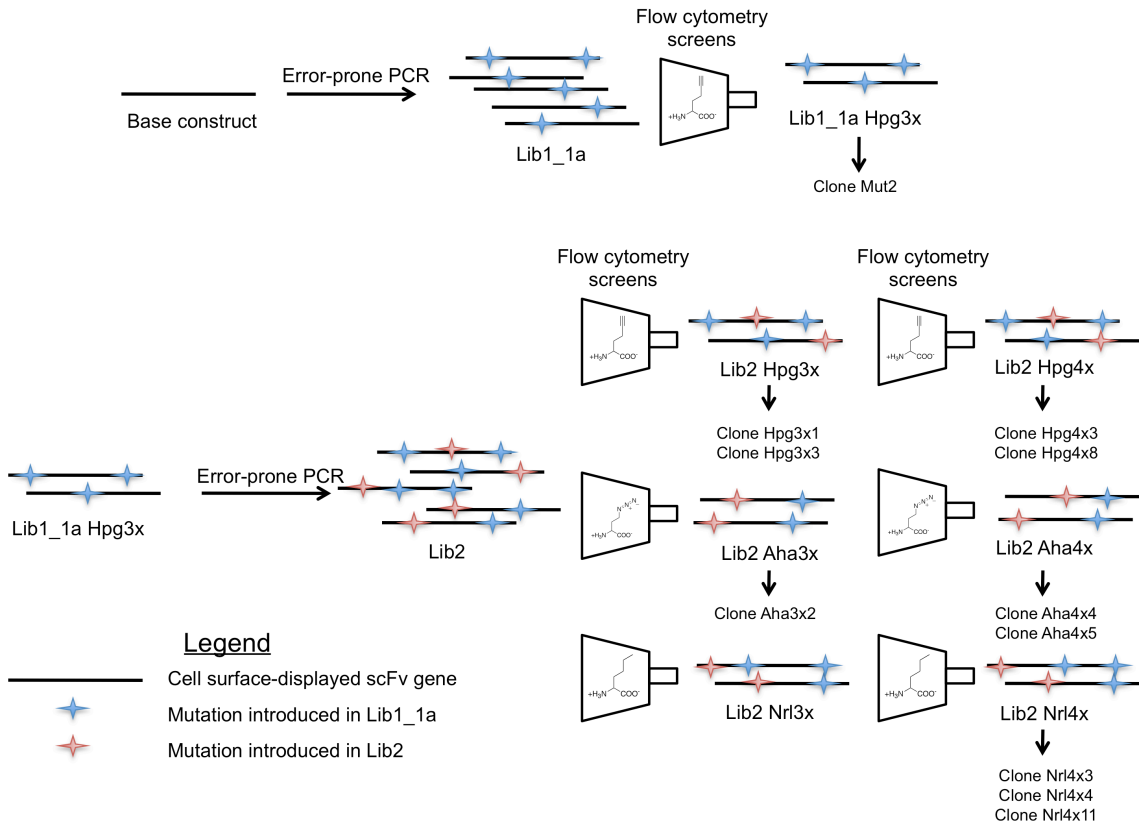


Figure 4.5. Summary of directed evolution of cell surface-displayed scFvs. Error-prone PCR using the DNA of Base (both display anchor and scFv) enabled construction and expression of Lib1_1a on the surface of *E. coli* cells. The library was screened for clones retaining binding function after the replacement of Met with Hpg. After three rounds of cell sorting, Lib1_1a Hpg3x contained a large fraction of Hpg-tolerant clones. This population was used as the basis for the construction of Lib2. Lib2 was screened for functional scFvs after the replacement of Met with Hpg, Aha, and Nrl in separate screens. Individual clones characterized in this work are noted below the populations from which they were isolated.

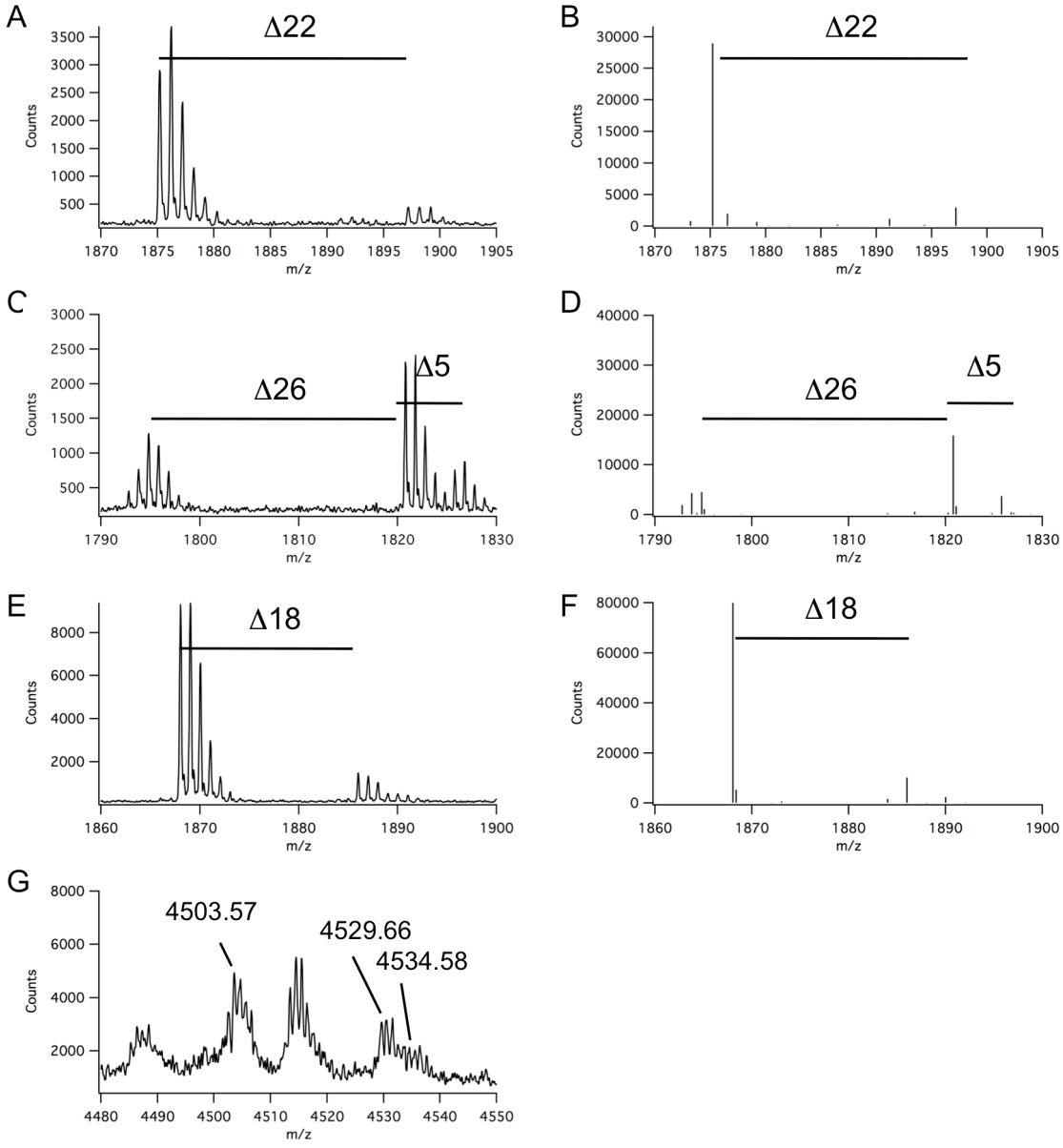


Figure 4.6. Examples of data used in estimating ncAA incorporation levels in scFvs with matrix-assisted laser desorption ionization (MALDI) mass spectrometry. (*A*, *B*) Portion of MALDI spectrum of trypsinized Hpg4x3 containing Hpg before (*A*) and after (*B*) processing that includes the peptide FSGNIFTDFY**M**NWVR. Replacement of Met by Hpg causes a peptide mass shift of approximately 22 Da. $[M+H]^+$'s in unprocessed spectrum: Met form, 1896.87 Da calculated, 1897.17 Da observed. Hpg form, 1874.88 Da calculated,

1875.18 Da observed. (*C*, *D*) Portion of MALDI spectrum of trypsinized Aha3x2 containing Aha before (*C*) and after (*D*) processing that includes the peptide SSGYISTDFY**M**NWVR. Replacement of Met by Aha results in a peptide mass shift of approximately 5 Da. The peak with a mass 26 Da lower than the substituted peak is characteristic of azide-containing peptides in MALDI and corresponds to the Aha-containing peptide after loss of dinitrogen and gain of two hydrogen atoms (62). $[M+H]^+$'s in unprocessed spectrum: Met form, 1825.82 Da calculated, 1825.77 Da observed. Aha form, 1820.82 and 1794.84 Da (loss of dinitrogen and gain of two hydrogen atoms) calculated, 1820.77 and 1794.82 Da observed, respectively. (*E*, *F*) Portion of MALDI spectrum of trypsinized Mut2 containing Nrl before (*E*) and after (*F*) processing containing the peptide SSGYIFTDFY**M**NWVR. Replacement of Met by Nrl results in a peptide mass shift of approximately 18 Da. $[M+H]^+$'s in unprocessed spectrum: Met form, 1885.86 Da calculated, 1885.98 Da observed. Nrl form, 1867.90 Da calculated, 1868.04 Da observed. Prior to estimating incorporation levels, all spectra were baseline corrected and deisotoped, with results shown as in (*B*), (*D*), and (*F*). The fraction of peptides bearing the ncAA substitution was calculated and used as a means of estimating amino acid replacement levels. In most cases, the deisotoping procedure resulted in the presence of single peaks at the substituted and unsubstituted positions. In cases where the deisotoping still resulted in multiple peaks (Aha-containing peptides or imperfect deisotoping), the counts of all peaks corresponding to unsubstituted and substituted peptides were summed prior to calculating the fraction of substituted peaks. (*G*) Portion of MALDI spectrum of trypsinized Aha4x4 containing Aha that includes the peptide WA**M**DYWGHGASVTVSSGGGGSGGGGSGGGGSDIVLTQSPASLAVSLGQR. The

only peptide containing Met (Aha) within the range of the higher-resolution acquisition used to estimate ncAA incorporation levels of all other proteins did not appear at high enough intensities to allow for quantification of substitution with Aha4x4 containing Aha. However, the pattern of substituted peak intensities observed in the unprocessed spectrum shown in (*G*) is similar to the pattern observed in other Aha-substituted peaks (*C*), suggesting a high level of substitution. The deisotoping algorithm could not resolve the unsubstituted and substituted peaks at observed m/z values of 4534.58 and 4529.66 Da, respectively, making quantification of substitution levels unfeasible. $[M+H]^+$'s: Met form, 4535.23 Da calculated, 4534.58 Da observed. Aha form, 4530.14 and 4504.14 Da (loss of dinitrogen and gain of two hydrogen atoms) calculated, 4529.66 and 4503.57 Da observed.

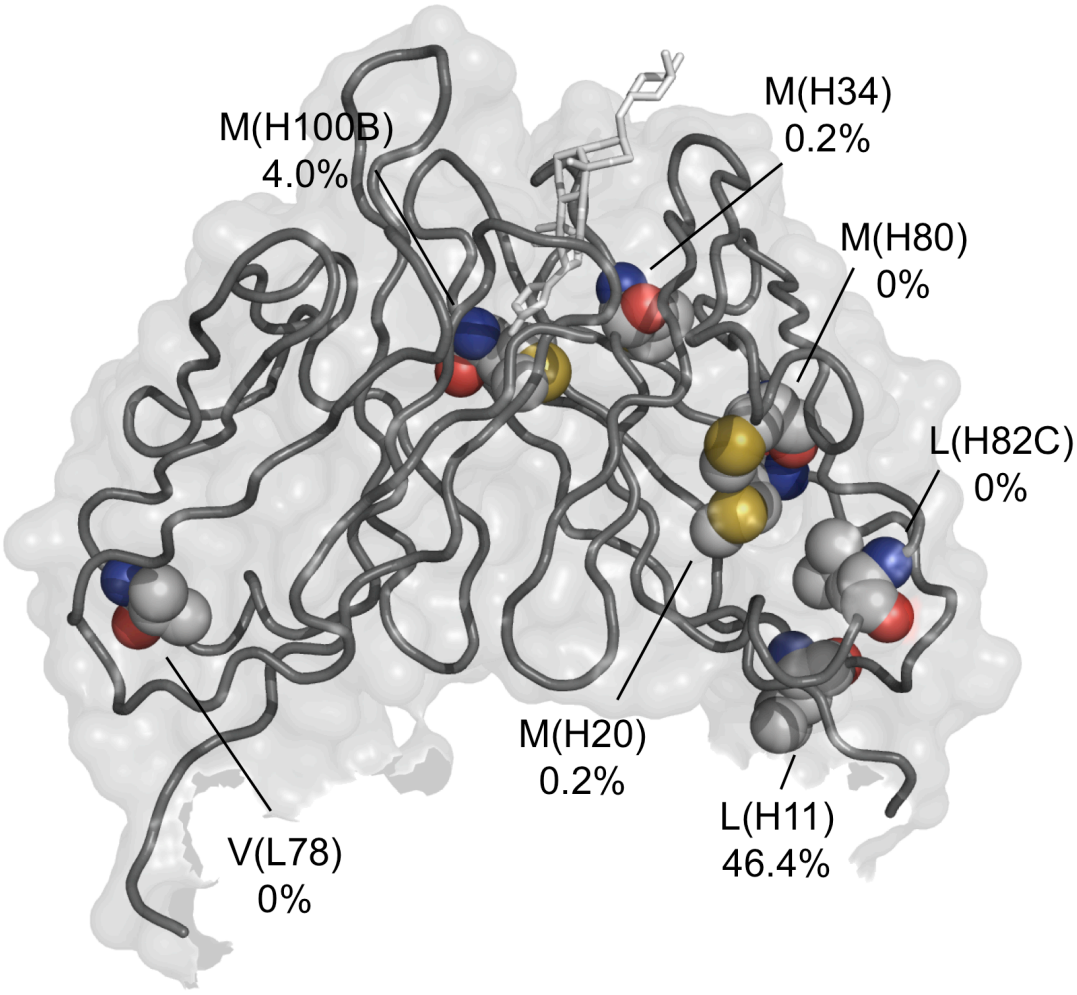


Figure 4.7. Positions of Met residues and residues mutated to Met in scFv. Each residue in the protein that exists as Met in at least one solubly produced variant is shown in space filling, with calculated surface accessibilities from GetArea (40) listed beneath the residue of interest. The accessibility 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) (32). The backbone of the Fv portion of the Fab is outlined with a ribbon diagram (gray), with amino acids of interest shown as space filling and the antigen is shown as sticks in white. Unless otherwise noted, labeled residues are frequently

mutated. H, heavy chain. L, light chain. Space filling colored by elements: carbon, gray; oxygen, red; nitrogen, blue; sulfur, yellow.

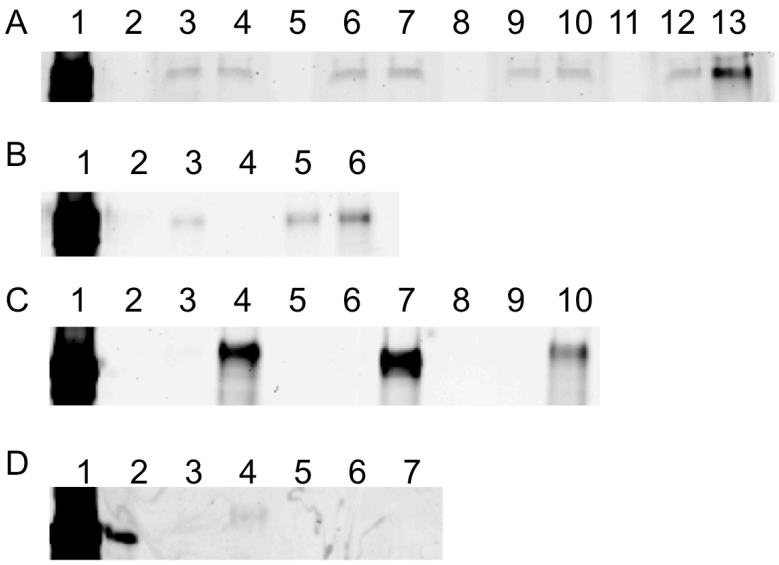


Figure 4.8. Copper-catalyzed azide-alkyne cycloadditions (CuAAC). All reactions were performed with solutions of 1.0–1.75 μM protein using reaction conditions as recommended by Hong et al. (58). After one hour, room temperature reactions, all reactions were quenched using 10 mM 3-azidopropanol. Reacted samples were run on SDS-PAGE gels and imaged for fluorescence. The quantity of fluorescence and protein in each lane was quantified using ImageQuant software (table 4.10) (*A*) Reactions of azide-containing proteins with a fluorescent TAMRA-alkyne dye (Invitrogen). EDTA removal from scFv solutions was performed using a single desalting column (Zeba desalting column, Thermo Fisher). The dye was added to the reaction mixtures at a final concentration of 200 μM. Lanes: 1, Aha form of GFPrm_AM (positive control), a variant of green fluorescent protein coding for seven Met (Aha) residues (55). 2, blank. 3, Met form of Mut2. 4, Aha form of Mut2. 5, blank. 6, Met form of Aha3x2. 7, Aha form of Aha3x2. 8, blank. 9, Met form of Aha4x4. 10, Aha form of Aha4x4. 11, blank. 12, Met form of Aha4x5. 13, Aha form of Aha4x5. (*B*) Dye labeling of Aha-containing scFvs was not substantially improved with multiple buffer exchanges. scFvs were exchanged once or twice as noted and reacted with

TAMRA-alkyne added to a final concentration of 20 μM. Lanes: 1, Aha form of GFPrm_AM (positive control). 2, blank. 3, Aha form of Aha4x5, EDTA removal with one desalting column. 4, blank. 5, Met form of Aha4x5, EDTA removal with two desalting columns. 6, Aha form of Aha4x5, EDTA removal with two desalting columns. (*C*, *D*) Reactions of alkyne-containing proteins with a fluorescent lissamine-rhodamine dye (57). EDTA removal from scFv solutions was performed using two desalting columns, and the dye was added to the reaction mixture at a final concentration of 20 μM. Lanes in (*C*): 1, Hpg form of GFPrm_AM (positive control) (55). 2, blank. 3, Met form of Mut2. 4, Hpg form of Mut2. 5, blank. 6, Met form of Hpg3x1. 7, Hpg form of Hpg3x1. 8, blank. 9, Met form of Hpg3x3. 10, Hpg form of Hpg3x3. Lanes in (*D*): 1, Hpg form of GFPrm_AM (positive control). 2, blank. 3, Met form of Hpg4x3. 4, Hpg form of Hpg4x3. 5, blank. 6, Met form of Hpg4x8. 7, Hpg form of Hpg4x8.

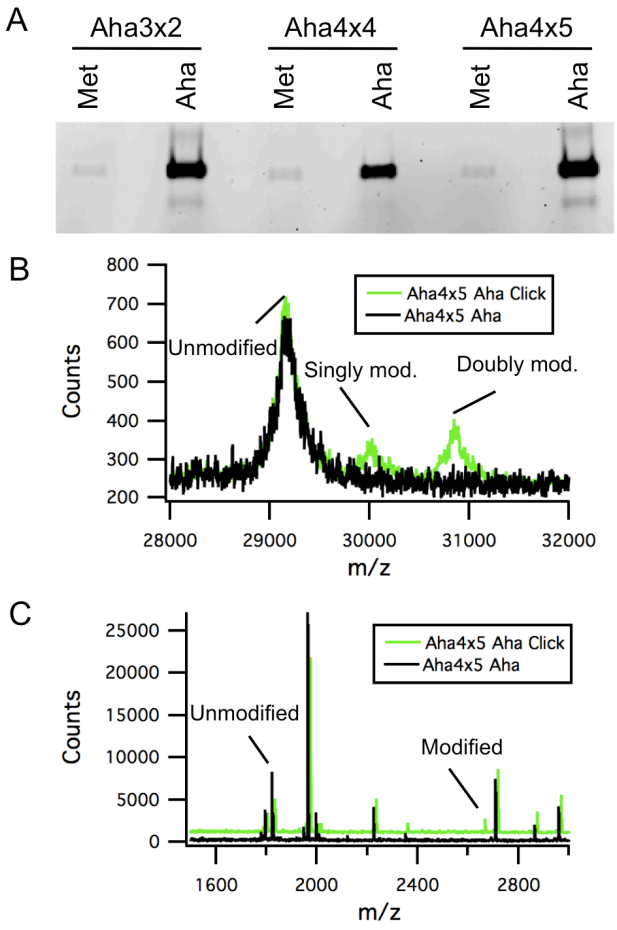


Figure 4.9. Strain-promoted click chemistry on scFvs using fluorescently labeled compound **6**. (*A*) SDS-PAGE on Met and Aha forms of scFv variants after reactions with **6**. (*B*) MALDI mass spectrometry on the Aha form of intact Aha4x5 before and after reaction with **6**. The ladders of signals detected are spaced apart by approximately 834 Da, corresponding to the molecular weight of **6** (the baseline and counts of the unmodified spectrum have been adjusted so that the two unmodified peaks overlay; no changes were made to m/z). $[M+H]^+$'s: unmodified: 29202 Da calculated, ~29130–29210 Da observed. Singly modified: ~30036 Da calculated, ~29981–30057 Da observed. Doubly modified: ~30870 Da calculated, ~30813–30888 Da observed. (*C*) 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 SSGYISTDFY**Aha**NWVR ($[M+H]^+$ 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 only present in the mass spectrum after performing click chemistry (“click” spectrum offset by 10 Da and 1000 counts for clarity).

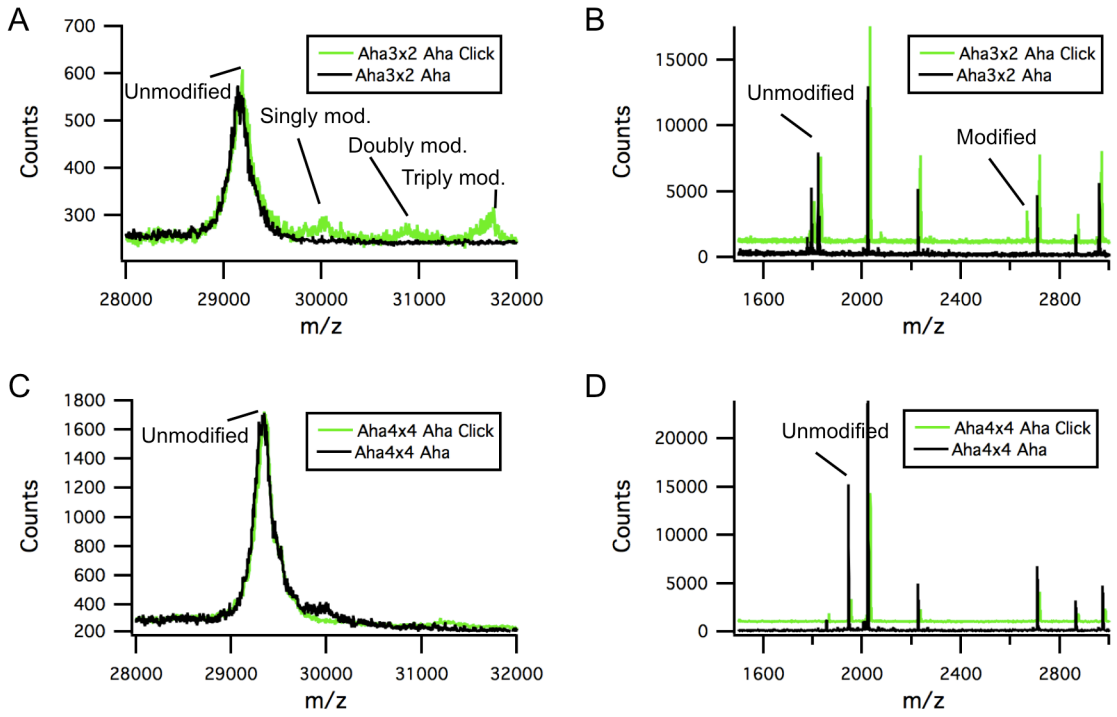


Figure 4.10. MALDI mass spectrometry on scFvs before and after strain-promoted click chemistry. (*A*) MALDI mass spectrometry on the Aha form of intact Aha3x2 before and after reaction with **6**. The ladders of signals detected are spaced apart by approximately 834 Da, corresponding to the molecular weight of **6** (the unmodified sample spectrum has been adjusted in intensity and position so that the two unmodified peaks overlay; no changes were made to m/z). $[M+H]^+$'s: unmodified: 29207 Da calculated, ~29127–29188 Da observed. Singly modified: ~30041 Da calculated, ~29986–30048 Da observed. Doubly modified: ~30860 Da calculated, ~30862–30897 Da observed. Triply modified: ~31694 Da calculated, 31694–31760 Da observed. (*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 SSGYISTDFY**Aha**NWVR ($[M+H]^+$ 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 only present in the mass spectrum after performing click chemistry ("click" spectrum offset by 10 Da and 1000 counts for clarity). (*C*) MALDI mass spectrometry on the Aha form of intact Aha4x4 before and after reaction with **6** (the unmodified spectrum has been adjusted in intensity and position so that the two unmodified peaks overlay; no changes were made to m/z). Consistent with the lower extents of modification observed via SDS-PAGE, modifications to the Aha form of Aha4x4 are not evident in mass spectrometry samples. $[M+H]^+$'s: unmodified: 29359 Da calculated, ~29290–29366 Da observed. Singly modified: ~30193 Da calculated, none observed. Doubly modified: ~31027 Da calculated, none observed. (*D*) 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.

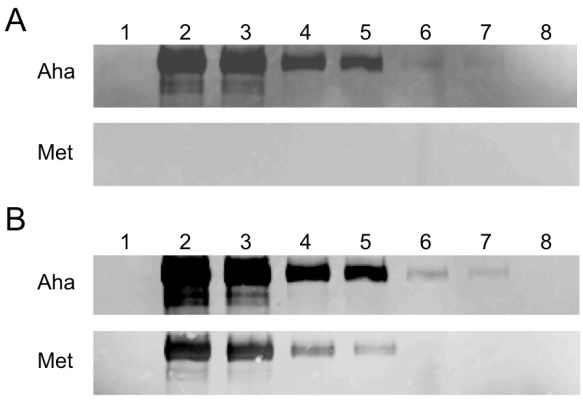


Figure 4.11. Western blotting using fluorescently labeled scFvs. (*A*) Probing nitrocellulose membranes for digoxigenin-labed BSA (BSA-Dig) using Aha and Met forms of Aha4x5 subjected to reaction with **6** and detection of Alexa-Fluor 488 fluorescence. Each lane of the SDS-PAGE gel transferred onto the blot was loaded with ~5 μg *E. coli* lysate and the following protein samples: Lane 1, 1000 ng BSA (unlabeled). Lanes 2–8: 1000, 500, 100, 50, 10, 5, and 1 ng BSA-Dig, respectively. (*B*) Probing for BSA-Dig using labeled scFvs subjected to reaction with **6** and secondary detection of scFvs using Alexa Fluor 647-labeled anti-Penta-His antibodies. These controls confirm that scFvs do not appear to lose their binding function after reaction with **6**. Lanes same as in (*A*).

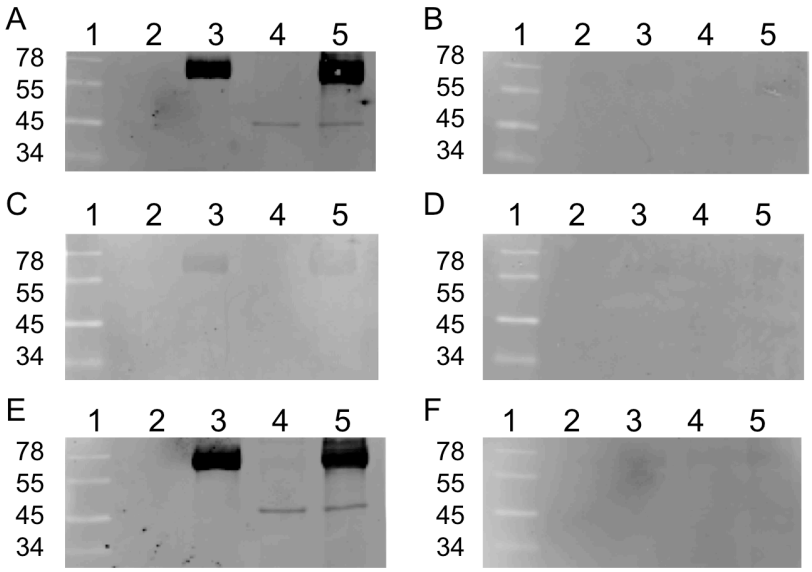


Figure 4.12. Western blotting using fluorescently labeled scFvs, part 2. 1000 ng of BSA or digoxigenin-labeled BSA (BSA-Dig) were run on SDS-PAGE gels, with 5.5 μg *E. coli* lysates added to protein samples as specified. Gels were transferred to nitrocellulose membranes, blocked, and probed for the presence of digoxigenin using scFvs labeled with **6**. Lanes in all panels: 1, protein standard. 2, BSA. 3, BSA-Dig. 4, BSA plus lysate. 5, BSA-Dig plus lysate. Blots were probed with (*A*) Aha form of Aha3x2, (*B*) Met form of Aha3x2, (*C*) Aha form of Aha4x4, (*D*), Met form of Aha4x4, (*E*) Aha form of Aha4x5, (*F*), Met form of Aha4x5 after quenching dye labeling reactions, but without separating unreacted **6** from protein samples. Molecular weights of marker proteins are given in kilodaltons.

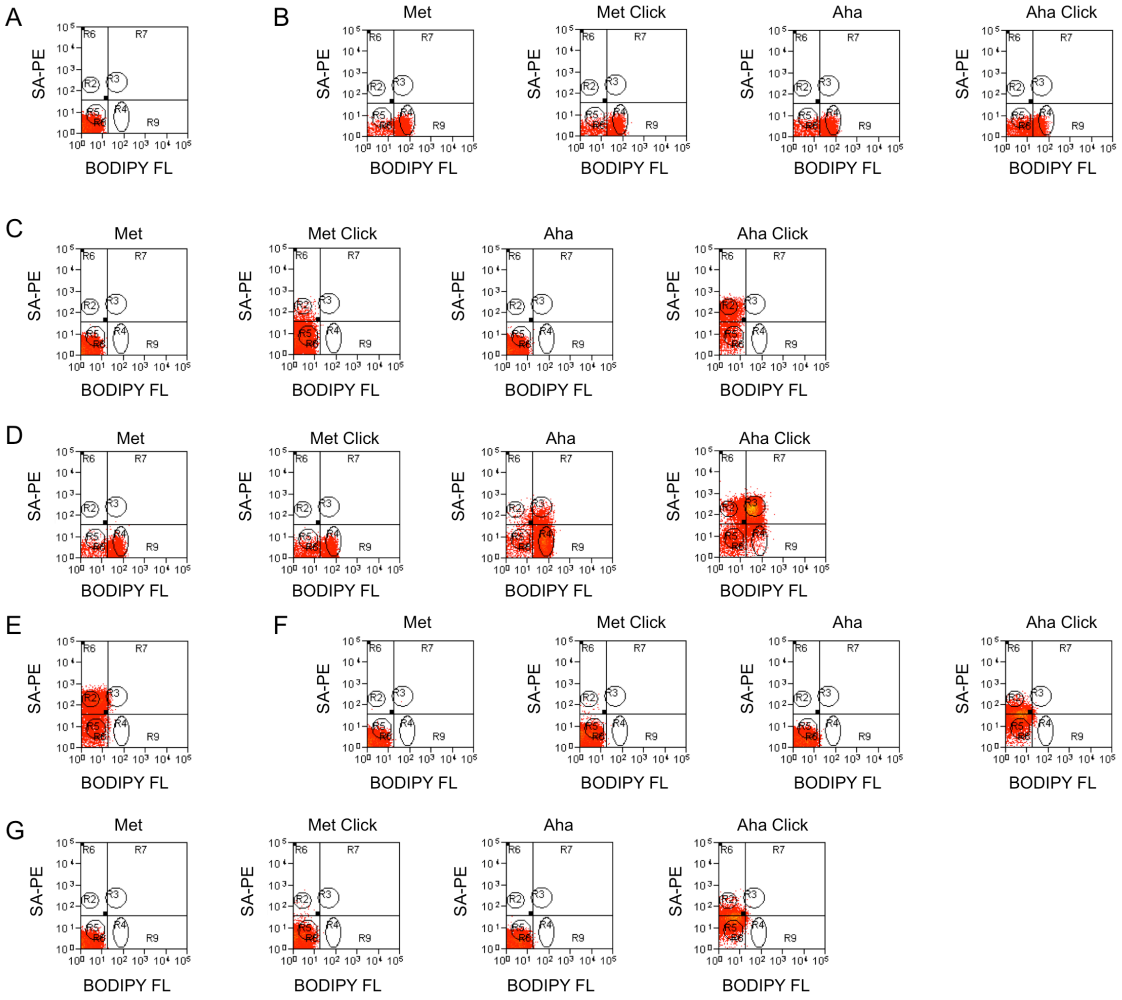


Figure 4.13. Flow cytometry of cell surface-displayed Aha4x5 to probe binding function and chemical modification with strained alkynes. The two-dimensional dot plots show the simultaneous measurements of the amount of **5** bound to cells (a measurement of binding function) and streptavidin-phycoerythrin (SA-PE) bound, an indirect estimation of the amount of **7** that has been chemically attached to cells. (*A*) Cells displaying the Met form of Aha4x5 without any dyes. (*B*) Probes of binding function of cells using **5**, with amino acid context and exposure to **7** (i.e., strain-promoted click chemistry) as noted in panels. (*C*) Probes of extent of modification of cells using SA-PE, with amino acid context and exposure to **7** as noted in panels. (*D*) Use of **5** and SA-PE to probe binding and function

simultaneously, with amino acid context and exposure to **7** as noted in panels. Simultaneous examination of these two properties on cell surfaces gives the same results as when these properties are assessed separately. (*E*) Adding unlabeled digoxin to the Aha form of cells expressing Aha4x5 clicked with **7** prior to exposure to **5** and SA-PE results in cells unable to bind to fluorescent antigen while retaining their modification. (*F*) Use of **5** and SA-PE to probe the function and modification of uninduced cells harboring plasmids bearing the Aha4x5 cell surface display construct. Amino acid context and exposure to **7** as noted in panels. Cells grown in Aha and reacted with **7** bind to a moderate amount of SA-PE, although far less than when the scFv construct is expressed. (*G*) Use of **5** and SA-PE to probe for the binding and modification of cells lacking the plasmid for the cell surface-displayed Aha4x5 construct. Amino acid context and exposure to **7** as noted in panels. Uninduced cells and cells lacking copies of the cell surface display vector behave identically, suggesting that labeling of cells incubated with Aha with **7** is the result of incorporation of Aha into other cellular membrane proteins, albeit at a level far lower than when scFv constructs are present on the cell surface.

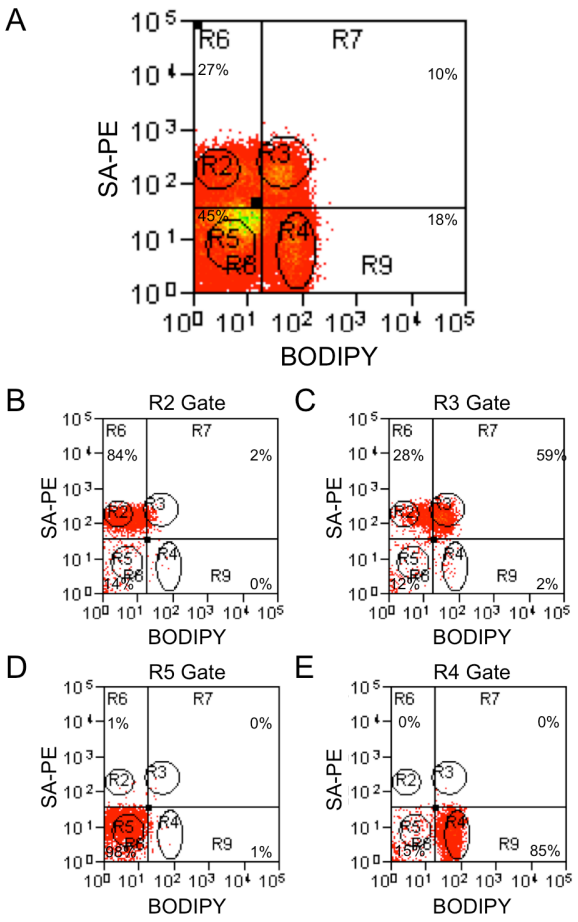


Figure 4.14. Fluorescence activated cell sorting for isolation of functional, modified proteins. (*A*) Two-dimensional dot plot of a mixture of four populations of cells. Each cellular population was reacted with **7** and exposed to streptavidin-phycoerythrin (SA-PE) and **5** for detection of chemical modification and binding, respectively, prior to flow cytometry. The four cell populations in the mixture are Aha4x5 expressed in Met (binding but no chemical modification), Aha4x5 expressed in Aha (binding and modification), Aha4x5 expressed in Aha blocked with nonfluorescent digoxin (chemical modification but no binding), and Aha4x5 grown in Aha without induction (no chemical modification or binding). (*B–E*) Sorted populations of cells. Stringent sorting using elliptical regions R2–R5 led to the isolation of populations of cells having distinct fluorescence characteristics

based on their functional and chemical modification characteristics. The quadrant defining regions R6–R9 was set based on the fluorescence properties of cells sorted from R5. In all panels, the reported percentages correspond to the fraction of cells appearing within regions R6–R9.

Table 4.1. Amino acid mutations in clones isolated from Lib1_1a Hpg3x

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 4.2. Amino acid mutations in cell surface-displayed scFvs isolated from Lib1_1a including mutations in display anchor*

Clone	E54	V63	F109	T118	W129	M(H20)	S(H20)	M(H80)	L(H82C)	I(L2)	A(L12)
1	-	A	-	-	-	-	P	L	-	V	-
3	-	-	-	-	-	-	-	L	-	-	-
6	D	-	L	I	-	I	-	L	-	-	T
7	-	-	-	-	-	-	-	L	-	-	-
10	-	-	-	-	R	-	-	L	-	-	-
2	-	-	Y	-	-	I	-	L	M	-	-
4	-	-	Y	-	-	I	-	L	M	-	-
5	-	-	Y	-	-	I	-	L	M	-	-
8	-	-	Y	-	-	I	-	L	M	-	-
9	-	-	Y	-	-	I	-	L	M	-	-

*Numbering scheme: numbers in parentheses are located within the scFv and are numbered according to the Kabat numbering scheme (H, heavy chain. L, light chain). All other mutations are numbered according to their position in the cell surface display construct with position 1 signifying the initiator Met in the signal sequence.

Table 4.3. Summary of conditions used in flow cytometry sorting of Lib2

	Expression in Medium Containing		
Population	Hpg (2.5 mM)	Aha (0.28 mM)	Nrl (0.30 mM)
Naïve Lib2	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 min. competition	Fluorescence, Viability 40 min. competition	Fluorescence, Viability 8 min. competition
3x sort	Fluorescence, Viability 40 min. competition	Fluorescence, Viability 100 min. competition	Fluorescence, Viability 15 min. competition
4x sort			

Table 4.4. ScFv 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. (30) (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 4.5. Amino acid mutations in clones isolated from Lib2*

Colony	Notes	V35	M39	Y41	D42	P48	E54	N55	G56	K59	Q61	V63	L65	T66
Hpg4x1		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x3		-	-	-	-	-	-	-	-	-	-	-	-	S
Hpg4x4		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x5	†	-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x6		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x7		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x8		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x9		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg4x10		-	-	-	-	L	-	-	-	-	R	-	-	-
Hpg4x11		-	-	-	-	-	-	-	-	-	-	-	-	S
Hpg4x12	‡	-	-	-	-	-	D	-	-	-	-	-	-	-
Hpg3x1		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x2		-	-	N	-	-	-	-	-	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-	-	-	-	-	I	-	-
Hpg3x4		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x5		-	-	-	-	-	-	-	-	-	-	-	-	-
Hpg3x6		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x1		-	-	-	Y	-	-	-	-	-	-	-	-	-
Aha3x2		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x3		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x4		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x5		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha3x6		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x1		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x2	†	-	-	-	-	-	-	-	-	-	R	-	-	-
Aha4x4		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x5		-	-	-	-	-	-	-	-	-	-	-	-	-
Aha4x6		-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x1		-	-	-	-	-	-	-	-	-	-	-	-	S
Nrl4x2		D	V	-	-	-	-	-	-	-	-	-	-	-
Nrl4x3		-	-	-	-	-	-	-	-	-	-	-	P	-
Nrl4x4		-	-	-	-	-	D	S	N	Q	-	-	-	-
Nrl4x5		-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x6		-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x7		-	-	-	-	-	-	-	-	-	-	A	-	-
Nrl4x8		-	-	-	-	-	-	-	-	-	-	-	-	S
Nrl4x9		-	-	-	-	-	-	-	-	-	-	-	-	-
Nrl4x10		-	-	-	-	-	-	-	-	-	-	-	-	S
Nrl4x11		-	-	-	-	-	-	-	-	-	-	A	-	-

*Numbering scheme: numbers in parentheses are located within the scFv and are numbered according to the Kabat numbering scheme (H, heavy chain. L, light chain). Positions in between the heavy and light chains are in the linker region and are numbered accordingly. All other mutations are numbered according to their position in the cell surface display construct (position 1: initiator Met in the signal sequence).

†Truncated fragment.

‡Missing amino acids 400−404.

Colony	Notes	A67	P72	D76	D78	Y80	T81	G84	G85	D91	N95	K99
Hpg4x1		-	-	-	-	-	A	-	-	-	-	-
Hpg4x3		-	-	-	-	-	-	-	-	-	-	-
Hpg4x4		-	-	-	-	-	-	-	-	-	-	-
Hpg4x5	†	-	-	-	-	-	-	-	-	-	-	-
Hpg4x6		-	-	N	-	-	-	-	-	-	-	-
Hpg4x7		-	-	-	-	-	-	-	-	-	-	-
Hpg4x8		-	-	-	-	-	-	S	-	-	-	-
Hpg4x9		-	-	-	-	-	-	-	-	-	-	-
Hpg4x10		-	-	-	-	-	-	-	-	-	-	-
Hpg4x11		-	-	-	-	-	-	-	-	-	-	-
Hpg4x12	‡	-	-	-	-	-	-	-	-	-	-	-
Hpg3x1		-	-	-	-	-	-	-	-	-	-	-
Hpg3x2		-	-	-	-	-	-	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-	-	-	-	-	-
Hpg3x4		-	S	-	-	-	-	-	-	-	-	-
Hpg3x5		-	-	-	-	H	-	-	-	-	-	-
Hpg3x6		-	-	-	-	-	-	-	-	-	-	-
Aha3x1		-	-	-	-	-	-	-	-	-	-	-
Aha3x2		-	-	-	-	-	-	-	-	-	-	-
Aha3x3		-	-	-	-	-	-	-	-	-	-	-
Aha3x4		-	-	-	-	-	-	-	-	-	-	-
Aha3x5		-	-	-	-	-	-	-	-	-	Y	-
Aha3x6		-	-	-	-	-	-	-	-	-	-	-
Aha4x1		-	-	-	-	-	-	-	-	-	-	-
Aha4x2	†	-	-	-	-	-	-	-	-	-	-	-
Aha4x4		-	-	-	-	-	-	-	-	-	-	-
Aha4x5		-	-	-	-	-	-	-	-	-	-	-
Aha4x6		-	-	-	-	-	-	-	-	-	-	-
Nrl4x1		-	-	-	-	-	-	-	-	-	-	-
Nrl4x2		-	-	-	-	-	-	-	-	-	-	-
Nrl4x3		-	-	-	-	-	-	-	-	-	-	-
Nrl4x4		-	-	-	-	-	-	-	-	-	-	-
Nrl4x5		-	-	-	-	-	-	-	D	-	-	-
Nrl4x6		T	-	-	A	-	-	-	-	-	-	-
Nrl4x7		-	-	-	-	-	-	-	-	-	-	-
Nrl4x8		-	-	-	-	-	-	-	-	-	-	-
Nrl4x9		-	-	-	-	-	A	-	-	-	-	-
Nrl4x10		-	-	-	-	-	-	-	-	-	-	-
Nrl4x11		-	-	-	-	-	-	-	-	Y	-	N

†Truncated fragment.

‡Missing amino acids 400−404.

Colony	Notes	N100	F109	E114	I117	T118	E120	A122	Q128
Hpg4x1		-	-	-	-	-	-	-	-
Hpg4x3		-	-	G	-	-	-	-	-
Hpg4x4		-	-	-	-	-	-	-	-
Hpg4x5	†	-	Y	-	-	-	-	-	-
Hpg4x6		-	-	-	-	-	-	-	-
Hpg4x7		-	-	-	-	-	-	-	-
Hpg4x8		-	-	-	-	-	-	-	-
Hpg4x9		Y	-	-	-	-	-	-	-
Hpg4x10		-	-	-	T	-	-	-	-
Hpg4x11		-	-	G	-	-	-	-	-
Hpg4x12	‡	K	Y	-	-	-	-	-	-
Hpg3x1		-	Y	-	-	-	-	-	-
Hpg3x2		-	-	-	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-	-	-
Hpg3x4		-	-	-	-	-	-	-	-
Hpg3x5		-	Y	-	-	-	F	-	-
Hpg3x6		-	-	-	-	-	-	-	-
Aha3x1		-	-	-	-	-	-	-	-
Aha3x2		-	Y	-	-	-	-	-	-
Aha3x3		-	Y	-	-	-	-	-	-
Aha3x4		-	L	-	-	-	-	-	K
Aha3x5		-	-	-	-	-	-	-	-
Aha3x6		-	-	-	-	-	-	-	-
Aha4x1		-	Y	-	-	-	-	-	-
Aha4x2	†	-	-	-	-	-	-	-	-
Aha4x4		-	-	-	-	-	-	-	-
Aha4x5		-	-	-	-	-	-	-	-
Aha4x6		-	-	-	-	P	-	-	-
Nrl4x1		-	-	G	-	-	-	-	-
Nrl4x2		-	-	-	-	-	D	-	-
Nrl4x3		-	-	-	-	-	-	G	-
Nrl4x4		-	-	-	-	-	-	-	-
Nrl4x5		-	Y	-	-	-	-	-	H
Nrl4x6		-	Y	-	-	-	-	-	-
Nrl4x7		-	-	-	-	-	-	-	-
Nrl4x8		-	-	G	-	-	-	-	-
Nrl4x9		-	Y	-	-	-	-	-	-
Nrl4x10		-	-	G	-	-	-	-	-
Nrl4x11		-	Y	-	-	-	-	-	-

†Truncated fragment.

‡Missing amino acids 400−404.

Colony	Notes	W129	N131	A136	G140	T141	E149
Hpg4x1		-	-	-	-	-	-
Hpg4x3		-	-	-	-	-	-
Hpg4x4		-	-	-	-	-	-
Hpg4x5	†	-	-	-	S	-	-
Hpg4x6		R	-	V	-	-	-
Hpg4x7		R	-	-	-	-	-
Hpg4x8		R	-	-	-	-	-
Hpg4x9		-	-	-	-	-	-
Hpg4x10		R	K	-	-	-	-
Hpg4x11		-	-	-	-	-	-
Hpg4x12	‡	-	-	-	-	-	-
Hpg3x1		-	-	-	-	-	-
Hpg3x2		R	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-
Hpg3x4		-	-	-	-	-	-
Hpg3x5		-	-	-	-	-	-
Hpg3x6		R	-	-	-	-	-
Aha3x1		R	-	-	-	-	-
Aha3x2		-	-	-	-	-	-
Aha3x3		R	-	-	-	-	-
Aha3x4		-	-	-	-	-	-
Aha3x5		R	-	-	D	-	-
Aha3x6		-	-	-	-	-	-
Aha4x1		R	-	-	-	P	-
Aha4x2	†	R	-	-	-	-	-
Aha4x4		R	-	-	-	-	-
Aha4x5		R	-	-	-	-	-
Aha4x6		R	-	-	-	-	-
Nrl4x1		-	-	-	-	-	-
Nrl4x2		-	-	-	-	-	-
Nrl4x3		-	-	-	-	-	G
Nrl4x4		-	-	-	-	-	-
Nrl4x5		-	-	-	-	-	-
Nrl4x6		-	-	-	-	-	-
Nrl4x7		-	-	-	-	-	-
Nrl4x8		-	-	-	-	-	-
Nrl4x9		-	-	-	-	-	-
Nrl4x10		-	-	-	-	-	-
Nrl4x11		-	-	-	-	-	-

†Truncated fragment.

‡Missing amino acids 400−404.

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

†Truncated fragment.

‡Missing amino acids 400−404.

Colony	Notes	M(H80)	L(H82C)	Y(H102)	A(H107)	S(H113)	G(Linker1)	G(Linker7)	G(Linker9)	I(L2)	A(L12)	S(L22)
Hpg4x1		L	M	-	-	-	-	-	-	-	-	-
Hpg4x3		L	M	-	-	-	-	-	-	-	-	-
Hpg4x4		L	M	-	-	-	-	-	-	-	-	-
Hpg4x5	†	L	M	-	-	-	S	-	-	-	-	-
Hpg4x6		L	M	-	-	-	-	-	-	-	-	-
Hpg4x7		L	M	-	-	-	-	-	-	-	-	F
Hpg4x8		L	M	N	-	-	-	-	-	-	-	-
Hpg4x9		L	M	-	-	-	-	-	-	-	-	-
Hpg4x10		L	M	-	T	-	-	-	-	-	-	-
Hpg4x11		L	M	-	-	-	-	-	-	-	-	-
Hpg4x12	‡	L	M	-	-	-	-	-	-	-	-	-
Hpg3x1		L	M	-	-	-	-	-	R	-	-	-
Hpg3x2		L	M	-	-	-	-	-	-	-	-	-
Hpg3x3		L	M	-	-	-	-	-	-	-	-	-
Hpg3x4		L	M	-	-	-	-	-	-	-	-	-
Hpg3x5		L	M	-	-	-	-	-	-	-	-	-
Hpg3x6		L	M	-	-	-	-	-	-	-	-	-
Aha3x1		L	M	-	-	-	-	-	-	-	-	-
Aha3x2		L	M	-	-	-	-	-	-	-	-	-
Aha3x3		L	M	-	-	-	-	-	-	-	-	-
Aha3x4		L	M	-	-	-	-	-	-	-	-	-
Aha3x5		L	-	-	-	-	-	-	-	-	T	-
Aha3x6		L	M	-	-	-	-	-	-	-	-	-
Aha4x1		L	M	-	-	-	-	-	-	-	-	-
Aha4x2	†	L	M	-	-	-	-	-	-	-	T	-
Aha4x4		L	M	-	-	-	-	-	-	-	-	-
Aha4x5		L	-	-	-	-	-	-	V	-	-	-
Aha4x6		L	M	-	-	-	-	-	-	-	-	-
Nrl4x1		L	M	-	-	-	-	-	-	-	-	-
Nrl4x2		L	-	-	-	-	-	-	-	-	-	-
Nrl4x3		L	M	-	-	-	-	-	-	-	-	-
Nrl4x4		L	M	-	-	-	-	-	-	-	-	-
Nrl4x5		L	-	-	-	-	-	-	-	V	-	-
Nrl4x6		L	-	-	-	-	-	-	-	-	-	-
Nrl4x7		L	-	-	-	-	-	-	-	V	-	-
Nrl4x8		L	M	-	-	-	-	-	-	-	-	-
Nrl4x9		L	M	-	-	-	-	R	-	-	-	-
Nrl4x10		L	M	-	-	-	-	-	-	-	-	-
Nrl4x11		L	-	-	-	F	-	-	-	-	-	-

†Truncated fragment.

‡Missing amino acids 400−404.

Colony	Notes	S(L25)	Q(L27)	P(L40)	N(L53)	S(L64)	E(L68)	F(L71)	T(L74)
Hpg4x1		-	-	-	-	-	-	-	-
Hpg4x3		-	-	-	-	-	-	-	-
Hpg4x4		-	-	-	-	-	-	-	-
Hpg4x5	†	-	-	-	-	-	-	-	-
Hpg4x6		-	-	-	-	-	-	-	-
Hpg4x7		-	-	-	-	-	-	-	-
Hpg4x8		-	-	-	-	-	-	-	-
Hpg4x9		-	-	-	-	-	-	-	-
Hpg4x10		-	-	-	Y	-	-	-	-
Hpg4x11		-	-	-	-	-	-	-	-
Hpg4x12*	‡	-	-	-	-	-	-	-	-
Hpg3x1		-	-	-	-	-	-	-	S
Hpg3x2		-	-	-	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-	-	-
Hpg3x4		-	-	-	-	-	-	-	-
Hpg3x5		-	-	-	-	-	-	-	-
Hpg3x6		-	-	-	-	-	-	-	-
Aha3x1		-	-	-	-	-	-	-	-
Aha3x2		-	-	-	-	-	D	-	-
Aha3x3		-	-	-	-	-	-	-	-
Aha3x4		-	-	-	-	-	-	-	-
Aha3x5		-	-	-	-	-	-	-	-
Aha3x6		-	-	-	-	-	-	Y	-
Aha4x1		-	-	-	-	-	-	-	-
Aha4x2	†	-	K	-	-	-	-	-	-
Aha4x4		T	-	-	-	-	-	-	-
Aha4x5		-	-	-	-	-	-	-	-
Aha4x6		-	-	-	-	-	-	-	-
Nrl4x1		-	-	-	-	-	-	-	-
Nrl4x2		-	-	S	-	-	-	-	-
Nrl4x3		-	-	-	-	-	-	-	-
Nrl4x4		-	-	-	-	-	-	-	-
Nrl4x5		-	-	-	-	T	-	-	-
Nrl4x6		-	-	-	-	-	-	-	-
Nrl4x7		-	-	-	-	-	-	-	-
Nrl4x8		-	-	-	-	-	-	-	-
Nrl4x9		-	-	-	-	-	-	-	-
Nrl4x10		-	-	-	-	-	-	-	-
Nrl4x11		-	-	-	-	-	-	-	-

†Truncated fragment.

‡Missing amino acids 400−404.

Colony	Notes	D(L76)	V(L78)	I(L85)	394	R395	404
Hpg4x1		-	-	-	-	-	-
Hpg4x3		-	M	-	-	-	-
Hpg4x4		-	-	-	-	-	-
Hpg4x5	†	-	-	-	Stop	-	-
Hpg4x6		-	-	-	-	-	-
Hpg4x7		-	-	-	-	-	-
Hpg4x8		-	-	-	-	-	-
Hpg4x9		-	-	-	-	-	-
Hpg4x10		-	-	-	-	-	-
Hpg4x11		-	M	-	-	-	-
Hpg4x12*	‡	N	-	-	-	H	-
Hpg3x1		N	-	-	-	-	-
Hpg3x2		-	-	-	-	-	-
Hpg3x3		-	-	-	-	-	-
Hpg3x4		-	-	-	-	-	-
Hpg3x5		-	-	-	-	-	-
Hpg3x6		-	-	-	-	-	-
Aha3x1		-	-	-	-	-	-
Aha3x2		-	-	-	-	-	-
Aha3x3		-	-	-	-	-	-
Aha3x4		-	-	-	-	-	-
Aha3x5		-	-	-	-	-	-
Aha3x6		-	-	-	-	-	-
Aha4x1		-	-	-	-	-	-
Aha4x2	†	-	-	-	-	-	Stop
Aha4x4		-	-	-	-	-	-
Aha4x5		-	-	-	-	-	-
Aha4x6		-	-	-	-	-	-
Nrl4x1		-	M	-	-	-	-
Nrl4x2		-	-	K	-	-	-
Nrl4x3		-	-	-	-	-	-
Nrl4x4		-	-	-	-	-	-
Nrl4x5		-	-	-	-	-	-
Nrl4x6		-	-	-	-	-	-
Nrl4x7		-	-	-	-	-	-
Nrl4x8		-	M	-	-	-	-
Nrl4x9		-	-	-	-	-	-
Nrl4x10		-	M	-	-	-	-
Nrl4x11		-	-	-	-	-	-

†Truncated fragment.

‡Missing amino acids 400−404.

Table 4.6. Frequently mutated positions (>5%) of scFvs identified in high-throughput sequencing of sorted populations

Position	Base*	Frequent Mutation	Amino Acid Mutation†	Mutational Frequency			
				Lib1_1a Hpg Sort 3	Lib2 Hpg4x	Lib2 Aha4x	Lib2 Nrl4x
342	c	t		0.01	0.06	0.01	0.02
351	a	c	E54R	0.10	0.11	0.03	0.06
357	t	a		0.01	0.04	0.00	0.16
371	a	g	Q61R	0.00	0.03	0.13	0.00
		t	Q61L	0.00	0.00	0.09	0.00
377	t	c	V63A	0.06	0.02	0.01	0.16
385	a	t	T66S	0.00	0.03	0.00	0.16
390	t	a		0.00	0.03	0.14	0.00
414	c	t		0.63	0.47	0.59	0.36
439	g	a	G84S	0.01	0.06	0.00	0.02
443	g	a	G85D	0.02	0.02	0.01	0.10
446	t	c	M86T	0.00	0.02	0.00	0.06
460	g	t	D91Y	0.00	0.02	0.00	0.05
514	t	c	F109L	0.09	0.05	0.02	0.02
515	t	a	F109Y	0.64	0.29	0.30	0.37
530	a	g	E114G	0.00	0.04	0.01	0.16
542	c	t	T118I	0.09	0.03	0.05	0.01
543	t	c		0.04	0.32	0.38	0.08
555	t	c		0.04	0.33	0.39	0.09
574	t	a,c	W129R	0.05	0.38	0.52	0.11
601	a	g	T138A	0.01	0.01	0.10	0.03
624	c	t		0.09	0.05	0.02	0.02
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 identity in DNA sequence of Base at specified position in plasmid.
†Numbering scheme same as in table 4.2.

Table 4.7. Amino acid mutations in scFvs studied in soluble form (Kabat numbering)

	E(H1)	E(H10)	L(H11)	M(H20)	S(H24)	Y(H27)	F(H29)	T(H30)	M(H34)	H(H41)	D(H46)	Y(H53)
Base	-	-	-	-	-	-	-	-	-	-	-	-
Mut2	-	-	-	I	-	-	-	-	-	-	-	-
Nrl4x3	G	-	-	I	F	-	S	-	-	-	-	N
Nrl4x4	-	-	-	I	-	-	S	-	-	-	E	-
Nrl4x11	-	-	-	L	Y	-	S	-	-	-	-	-
Hpg3x1	-	-	M	I	I	-	-	-	-	N	E	-
Hpg3x3	-	-	S	I	F	-	-	S	-	-	-	-
Hpg4x3	-	-	-	I	F	N	-	-	-	-	-	-
Hpg4x8	-	-	-	I	-	-	S	-	-	-	-	-
Aha3x2	-	-	-	I	-	-	S	-	-	-	-	-
Aha4x4	-	-	-	I	Y	-	-	-	I	-	-	-
Aha4x5	-	A	-	I	-	-	S	-	-	-	-	-

	M(H80)	L(H82C)	Y(H102)	S(H113)	G(Linker9)	S(L25)	E(L68)	T(L74)	D(L76)	V(L78)
Base	-	-	-	-	-	-	-	-	-	-
Mut2	L	M	-	-	-	-	-	-	-	-
Nrl4x3	L	M	-	-	-	-	-	-	-	-
Nrl4x4	L	M	-	-	-	-	-	-	-	-
Nrl4x11	L	-	-	F	-	-	-	-	-	-
Hpg3x1	L	M	-	-	R	-	-	S	N	-
Hpg3x3	L	M	-	-	-	-	-	-	-	-
Hpg4x3	L	M	-	-	-	-	-	-	-	M
Hpg4x8	L	M	N	-	-	-	-	-	-	-
Aha3x2	L	M	-	-	-	-	D	-	-	-
Aha4x4	L	M	-	-	-	T	-	-	-	-
Aha4x5	L	-	-	-	V	-	-	-	-	-

Table 4.8. 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) in order to form BSA-Dig. 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 was found to be free of substantial mass transport limitations.

‡Amino acid replacement levels were estimated based on matrix-assisted laser desorption ionization (MALDI) mass spectrometry of trypsinized scFv samples as described in figure 4.6 and materials and methods. Data is reported as the mean plus or minus the standard deviation of two or more independent trypsinizations and MALDI mass spectrum acquisitions.

Table 4.9. Dissociation kinetic rate constants of selected scFvs in Met and ncAA forms

Clone	k_{off} (10^{-3} s^{-1})			
AA Context	Met	Hpg	Aha	Nrl
Base	1.85 ± 0.15			2.31 ± 0.39
Mut2	1.72 ± 0.17	8.09 ± 0.53	0.96 ± 0.16	2.07 ± 0.20
Hpg3x3	0.93 ± 0.07	4.63 ± 0.34		
Aha4x4	1.00 ± 0.19		0.70 ± 0.15	
Nrl4x11	0.93 ± 0.09			1.08 ± 0.07

Table 4.10. Copper-catalyzed click chemistry (CuAAC) on Aha- and Hpg-containing proteins with TAMRA-alkyne and lissamine-rhodamine azide dyes, respectively*

TAMRA-Alkyne Functionalization									
Experiment 1**									
Protein	GFPrm_AM	Mut2		Aha3x2		Aha4x4		Aha4x5	
AA context	Aha	Met	Aha	Met	Aha	Met	Aha	Met	Aha
Normalized Fluorescence†	15.05	0.03	0.03	0.03	0.04	0.02	0.03	0.03	0.29
Experiment 2‡									
Protein	GFPrm_AM	Aha4x5							
AA context	Aha	Aha, 1 ex.	Met, 2 ex.	Aha, 2 ex.					
Normalized Fluorescence	1.88	0.02	0.07	0.19					
Lissamine-Rhodamine Azide Functionalization§									
Protein	GFPrm_AM	Mut2		Hpg3x1		Hpg3x3			
AA context	Hpg	Met	Hpg	Met	Hpg	Met	Hpg		
Normalized Fluorescence	3.18	0.00	0.48	0.00	0.51	0.00	0.16		
Protein	GFPrm_AM	Hpg4x3		Hpg4x8					
AA context	Hpg	Met	Hpg	Met	Hpg				
Normalized Fluorescence	3.33	0.05	0.24	0.04	0.05				

*All reactions were performed at room temperature for one hour following CuAAC conditions outlined by Hong et al. (58) (see figure 4.8 caption and materials and methods for details). GFPrm_AM was obtained in phosphate-buffered saline. ScFvs were purified in HEPES-buffered saline containing 3 mM EDTA and were buffer exchanged into phosphate-buffered saline prior to chemical reactions as noted.

†Samples were run on SDS-PAGE gels and imaged using a Typhoon Trio imaging system to assess functionalization. First, fluorescence detection of dyes was used to interrogate the efficiency of reaction. Second, all gels were stained in colloidal blue and imaged to assess protein quantities. The normalized data reported here is the intensity of the fluorescence of a sample divided by the intensity of colloidal blue staining in the same sample.

**Single buffer exchange to remove EDTA prior to reaction, 200 μM dye used during reaction.

‡One or two buffer exchanges to remove EDTA prior to reaction as noted (1 ex. or 2 ex.), 20 μM dye used during reaction.

§Two buffer exchanges to remove EDTA prior to reaction, 20 μM dye used during reaction.

Table 4.11. Dye labeling of Met- and Aha-containing proteins with **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 4.12. 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 4.8 and materials and methods. Clicked scFv samples were prepared for Biacore assays without separating unreacted dyes or quenching reagents from solution.

Table 4.13. 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'