Biophysics and Protein Engineering with Noncanonical Amino Acids

Thesis by

James A. Van Deventer

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2011

(Defended April 19, 2011)

For My Family

ACKNOWLEDGEMENTS

I am grateful to numerous individuals for their help and support during the pursuit of my doctoral degree. First, I would like to thank my advisor, Professor David Tirrell, for providing such comprehensive support, financial and otherwise, during my time in his laboratory. He presented me with an environment in which I was free to boldly pursue problems at the interface of chemistry and biology. I am extremely thankful for Dave's rigorous approach to scientific analysis, as his approach has trained me to be a very careful scientist. Moreover, the scientific freedom I have had during my time in the Tirrell laboratory has allowed me to develop an interest and passion for science and engineering that will continue to inspire me for years to come.

Many other members of the academic community have been very supportive of my work. My thesis committee, Professors Frances Arnold, William Clemons, and Mark E. Davis, have all provided helpful suggestions during my studies. I would especially like to thank Professor Arnold for helping me to understand some of the intricacies of natural evolution, directed evolution, and high-throughput screening. Professor Anand Asthagiri provided helpful comments as a member of my candidacy committee, and Professor Jim Swartz of Stanford University has been very generous with his time and advice over the past few years. Last, I would like to thank my undergraduate research advisor, Professor Stacey Bent, and her former student, Professor Michael Filler, for providing me with my earliest research opportunities as an undergraduate student at Stanford. Both Stacey and Mike were hugely supportive of me from the moment I started research, and I am extremely grateful for their encouragement to pursue a career in scientific research.

I have had the opportunity to collaborate with a number of colleagues during the course of my work. I thank Professor Ahmed Zewail and his laboratory members, Dr. Oh-Hoon Kwon and Dr. Christina Othon, for the pleasure of studying hydration dynamics with them. Collaborations with a number of former and current Tirrell lab members have also been extremely enjoyable over the years. Dr. Tae Hyeon Yoo was responsible for starting the hydration dynamics work with the Zewail Laboratory and provided support during its completion. He was also instrumental in training and advising me during early flow cytometry experiments, library constructions, and cell strain development. Professor A. James Link helped me perform enzymatic assays with radioactive chemicals and taught me some of the basics of molecular biology in support of studies on aliphatic ketone-containing amino acids. Professor Nick Fisk provided large-scale syntheses of homoisoleucine and other large, aliphatic noncanonical amino acids for use in biophysical studies. Professors Yi Tang and Pin Wang provided helpful comments during the completion of a manuscript on aliphatic ketone-containing amino acids. Kai P. Yuet has been instrumental during the completion of the antibody fragment engineering work. His enthusiasm, diligence, and attention to detail during the expression and characterization of more than thirty antibody fragments have been invaluable. It has been a pleasure to work with him during these efforts.

Life at Caltech has been very enjoyable thanks to the people I have been lucky enough to be surrounded with during my stay at the Institute. The Tirrell laboratory has been a wonderful place to work over the years, and I have appreciated conversing and interacting with so many thoughtful and kind scientists during my time in the lab. I especially thank Alborz, Alison, Beverly, Brad, Brett, Caglar, Chethana, Eileen, Frank,

Inchan, Jamie, Janek, J. D., Jeremiah, Jin, John, Julie Champion, Julie Liu, Kai, Kechun, Kevin, Kimberly, Larry, Maren, Nick Ball, Nick Fisk, Phoebe, Rebecca, Shelly, Soojin, Stacey, Tae Hyeon, and Yvette for conversations about life and the lab. My chemical engineering classmates are another group of people that has made life more fun over the years. I especially thank Chase Beisel for many enjoyable conversations about science and science politics and Jason Gamba for discussions of science and sports over lunch. I also thank my friends from outside chemical engineering for their help and encouragement during my scientific studies. These include Alberto Cueto, Katie and Ted Brenner, and countless additional friends I have made through the church communities of Warehouse and Westside Christian Fellowship.

My final thanks go out to my family for their unending support and love. My parents Allen and Debbie Van Deventer have inspired and encouraged me at every step. My brother Paul has also been a great backer of my studies throughout my time at Caltech. I am so fortunate to have had them all so close by while working on my degree, and their presence has enriched my life greatly over the past few years. Last, I thank my wife Allison for being a steadfast companion to me ever since we started dating. I am so lucky to have found an amazing woman to spend the rest of my life with during my time in Pasadena. Her advocacy and encouragement have been invaluable in helping me to persist and persevere through all of the challenges of completing a doctoral degree at Caltech. I am a much better person and scientist today thanks to her constant love and loyalty.

ABSTRACT

Noncanonical amino acids are tools for expanding and altering the chemical functionalities available within proteins. Much recent work has focused on developing biosynthetic means for incorporating noncanonical amino acids into proteins, and applications of noncanonical amino acids to many problems in science and engineering are emerging. The first portion of this thesis describes established methods to incorporate noncanonical amino acids into proteins and efforts to exploit the properties of noncanonical amino acids in areas such as protein structure determination, protein and organism evolution, modulation of the immune system, and proteomics. Researchers' creative and successful use of this growing toolkit suggests that noncanonical amino acids will continue to be a valuable asset for dissecting biological problems and imparting proteins with new chemical and physical properties.

Biophysical studies with noncanonical amino acids provide a platform for studying the effects of atom-by-atom manipulations of amino acid side chains on protein properties. The middle portions of this thesis describe work to better understand how protein properties are affected by subtle amino acid side chain manipulations. This work was aided greatly by the establishment of homoisoleucine as a translationally active analog of leucine in bacterial cells. The small side chain differences between leucine, homoisoleucine, and the fluorinated amino acid trifluoroleucine allow for detailed studies on how amino acid side chain size and fluorination affect protein stability and hydration dynamics. Replacement of leucine by homoisoleucine in coiled-coil peptides stabilizes these proteins, as shown by elevation of the coiled coil thermal denaturation temperature. The stabilization observed when homoisoleucine replaces leucine in the peptides is greater than when

trifluoroleucine replaces leucine, suggesting that expansion of side chain volume may play a role in protein stabilization irrespective of hydrocarbon or fluorocarbon character.

Studies of water-protein interactions using designed coiled coils containing surface-exposed leucine, homoisoleucine, or trifluoroleucine residues enabled systematic examination of the roles that side chain size and fluorination play in dictating solvation dynamics. Fluorinated side chains appear to exert a large electrostatic drag on nearby water molecules. These results have important implications for the design and engineering of fluorinated proteins due to the critical role water-protein interactions play in many protein properties and functions.

The final portion of this thesis details efforts to engineer the binding properties and chemical reactivity of antibody fragments with noncanonical amino acids. The properties of the single chain variable fragment form of a model anti-digoxin antibody have been studied after replacement of the protein's methionine residues with methionine analogs containing alkyne, azide, and aliphatic side chains. Experiments with antibody fragments displayed on the surface of *Escherichia coli* cells revealed that replacement of the methionine residues of the fragment 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 methionine 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 antibody fragments 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 methionine form of the parent antibody fragment 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 antibody fragment 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 fragments indicated that each noncanonical amino acid used in this study modulates the binding kinetics of antibody fragments in a distinct fashion. Perhaps most interestingly, fragments containing the azide-containing analog azidohomoalanine exhibit improved binding kinetics relative to their methionine-containing counterparts. Replacement of methionine by azidohomoalanine in several variants lowers the dissociation constant of the fragment by up to a factor of two. Chemical conjugation of azide-containing fragments 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 azide-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 noncanonical amino acids 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.

TABLE OF CONTENTS

List of Schemes xiii
List of Figures xiv
List of Tables xvii

Chapter 1 Expanding the Scope of Protein Science and Engineering with Noncanonical Amino acids 1
Abstract 2
Introduction 2
Biosynthetic Methodologies for the Incorporation of Genetically Encoded Noncanonical Amino Acids into Proteins 4
Applications 22
Outlook 63
Acknowledgements 67
References 67

Chapter 2 Homoisoleucine: A Translationally Active Leucine Surrogate of Expanded Hydrophobic Surface Area 129
Abstract 130
Introduction 130
Results and Discussion 131
Materials and Methods 133
Acknowledgements 143
References 143

Chapter 3 Hydration Dynamics at Fluorinated Protein Surfaces 155
Abstract 156
Introduction 156
Results 158
Discussion 165
Conclusions 167
Materials and Methods 167
Acknowledgements 179
References 180

Chapter 4 Antibody Fragment Engineering with Noncanonical Amino Acids 197
Abstract 198
Introduction 199
Results and Discussion 202
Conclusions 214
Materials and Methods 216
Acknowledgements 241

References 241

LIST OF SCHEMES

Number *Page*

1.1 Noncanonical amino acids 108

2.1 Amino acids used in study 147

3.1 Amino acids used in study 187

4.1 Compounds used in study 249

4.2 Chemistries used in modifying azide- and alkyne-containing ncAAs 250

LIST OF FIGURES

Number *Page*

1.1 Overview of strategies for incorporating noncanonical amino acids into proteins....112

1.2 Methods for residue-specific incorporation of noncanonical amino acids into proteins....113

1.3 Orthogonality requirements for adding tRNAs and aminoacyl-tRNA synthetases (aaRSs) to a translation system....115

1.4 Methods for site-specific incorporation of noncanonical amino acids into proteins....116

1.5 NMR studies of ligand binding employing noncanonical amino acids as isotopically labeled probes....118

1.6 Secondary structure analysis and thermal denaturation of recombinant human prion protein ($rhPrP^{C}$) containing methionine analogs....119

1.7 Unnatural amino acid mutagenesis on the 5-hydroxytryptamine type 3 receptor leads to a new model for receptor gating....120

1.8 Use of noncanonical amino acids to investigate cation-π interactions in the muscle-type nicotinic acetylcholine receptor (nAChR) using patch-clamp experiments....121

1.9 Use of ncAAs to breaking immunochemical self-tolerance....123

1.10 Flow cytometric analysis of cells expressing green fluorescent protein (GFP) and GFP variants during the course of evolving fluorinated GFPs....125

1.11 Cell-selective labeling in mixtures of bacterial and mammalian cells....127

2.1 A1 peptide sequence and helical wheel representation of A1 homodimers....148

2.2 Circular dichroism spectra of Leu-A1 and Hil-A1....149

Number *Page*

2.3 MALDI spectra of tryptic fragments 150
2.4 LC/MS results obtained on trypsinized sample of Hil-A1 151
3.1 Protein sequence and structure 188
3.2 Circular dichroism 189
3.3 Normalized plots from the Sedfit $c(s)$ analysis 190
3.4 Steady-state UV-visible absorption and fluorescence emission spectra 191
3.5 Time-resolved anisotropy, $r(t)$, of the proteins 192
3.6 Hydration dynamics 193
3.7 Hydration energy relaxation 194
3.8 LC/MS/MS of trypsinized A1m-H 195
4.1 Flow cytometry studies of the binding properties of cell surface displayed scFvs 251
4.2 Fluorescence activated cell sorting of Lib1_1a for clones that function when Met is replaced by Hpg 252
4.3 Fluorescence activated cell sorting of Lib2 for clones that function when Met is replaced by ncAAs 253
4.4 Population-level sequence characterization of scFv mutants using high-throughput sequencing 254
4.5 Summary of directed evolution of cell surface-displayed scFvs 256
4.6 Examples of data used in estimating ncAA incorporation levels in scFvs with matrix-assisted laser desorption ionization (MALDI) mass spectrometry 257
4.7 Positions of Met residues and residues mutated to Met in scFv 260
4.8 Copper-catalyzed azide-alkyne cycloadditions (CuAAC) 262
4.9 Strain-promoted click chemistry on scFvs using fluorescently labeled compound **6** 264
4.10 MALDI mass spectrometry on scFvs before and after strain-promoted click chemistry 266

Number *Page*

4.11 Western blotting using fluorescently labeled scFvs 268
4.12 Western blotting using fluorescently labeled scFvs, part 2.....269
4.13 Flow cytometry of cell surface-displayed Aha4x5 to probe binding function and chemical modification with strained alkynes.....270
4.14 Fluorescence activated cell sorting for isolation of functional, modified proteins 272

LIST OF TABLES

Number *Page*

2.1 Kinetic parameters for activation of Leu and Hil by LeuRS 152

2.2 Stabilization of A1 by replacement of Leu with noncanonical amino acids.....153

2.3 Incorporation levels of Hil in A1 samples determined from multiple series of substituted peptides 154

3.1 Fluorescence emission maxima (λ_{max}), hydration-correlated energy relaxation [$\Delta E_s(t)$], and depolarization dynamics [$r(t)$].....196

4.1 Amino acid mutations in clones isolated from Lib1_1a Hpg3x 274

4.2 Amino acid mutations in cell surface-displayed scFvs isolated from Lib1_1a including mutations in display anchor 275

4.3 Summary of conditions used in flow cytometry sorting of Lib2 276

4.4 ScFv off rate estimates performed using cell surface-displayed scFvs ... .277

4.5 Amino acid mutations in clones isolated from Lib2 279

4.6 Frequently mutated positions (>5%) of scFvs identified in high-throughput sequencing of sorted populations.....287

4.7 Amino acid mutations in scFvs studied in soluble form (Kabat numbering) 288

4.8 Characterization of soluble scFvs: expression yields, binding kinetics, and amino acid replacement estimates 289

4.9 Dissociation kinetics of selected scFvs in Met and ncAA forms.....290

4.10 Copper-catalyzed click chemistry (CuAAC) on Aha- and Hpg-containing proteins with TAMRA-alkyne and lissamine-rhodamine azide dyes, respectively 291

4.11 Dye labeling of Met- and Aha-containing proteins with **6** 292

4.12 Kinetic characterization of scFvs before and after reaction with **6** (strain-promoted click chemistry).....293

Number *Page*

4.13 Oligonucleotides used in study 294

CHAPTER 1

Expanding the Scope of Protein Science and Engineering with Noncanonical Amino Acids

This chapter has been adapted from a review in preparation by James A. Van Deventer, Ying Y. Liu, Jeremiah A. Johnson, and David A. Tirrell.

Abstract

Noncanonical amino acids (ncAAs) have the potential to greatly expand the chemical functionalities available within proteins. Methodologies for the genetic encoding and incorporation of ncAAs into proteins using the protein synthesis machinery from living organisms are now quite common, and applications of proteins containing ncAAs are emerging in a variety of fields. In this review, we highlight the most widely used methodologies for biosynthetic incorporation of ncAAs into proteins. We then discuss applications of ncAA incorporation to areas that include protein structure determination, protein and organism evolution, modulation of the immune system, and proteomics. Numerous successes in these and other fields suggest that biosynthetic ncAA incorporation will continue to be a valuable tool for biological science and engineering in the future. The ease of incorporation and large functional toolkit available with ncAAs will continue to enable researchers from many disciplines to tailor the use of ncAAs to their specific needs.

Introduction

Proteins perform an astonishingly broad range of functions in biological systems, yet they are usually synthesized from no more than twenty amino acid monomers. While the templated synthesis of polypeptide chains provides access to an enormous sequence space (1), and posttranslational modifications significantly expand the available functional space (2), the chemical diversity of natural amino acid side chains is rather limited from a chemist's perspective. Thus, researchers have focused much effort on the development of new methodologies for incorporation of amino acids that are not normally specified by the genetic code, or noncanonical amino acids (ncAAs), into peptides and proteins.

Methods for the incorporation of ncAAs into peptides and proteins are either chemical, biosynthetic, or some combination of the two. In the area of chemical synthesis, solid-phase peptide synthesis has enabled the routine incorporation of virtually any suitably protected amino acid into synthetic peptides (3, 4). Small peptides can then be coupled together to give full-length synthetic proteins via techniques such as native chemical ligation (5, 6). The combination of chemical ligation strategies and biological protein production methods in the form of techniques such as expressed protein ligation (7-9) further enhances researchers' abilities to incorporate new chemical functionalities into proteins. These nontemplated chemical approaches for incorporation of ncAAs into proteins continue to grow in importance to the scientific community.

Despite the successes of chemical methods, the natural protein biosynthetic machinery of living organisms remains unparalleled in its ability to produce complex, genetically templated polypeptide chains in large quantities. For many years scientists have imagined that a reworking of the genetic code or the protein biosynthetic machinery could enable the biological production of proteins containing a nearly endless variety of ncAAs (10) and perhaps even other non-amino acid monomer structures (11). Sequence-specific, monodisperse polymers, which are currently challenging or impossible to synthesize using traditional synthetic techniques, would be expected to display a wide range of chemical and physical properties. Methodological developments over more than fifty years have facilitated the biosynthetic incorporation of an ever-increasing number of ncAAs and other monomers into proteins. These techniques have begun to allow scientists to investigate a vast range of subjects in biology, biological chemistry, and engineering.

In this review, we aim to highlight some of the ways in which scientists have employed biosynthetically produced proteins containing ncAAs to study and manipulate proteins and biological systems. The section "Biosynthetic Methodologies for the Incorporation of Genetically Encoded Noncanonical Amino Acids into Proteins" will summarize some important historical and methodological underpinnings of ncAA incorporation into proteins using enzymatic machinery found in living organisms. In the "Applications" section, we will discuss a number of research topics that have been impacted substantially by the use of ncAAs. The subjects highlighted in this review are meant to emphasize the creative approaches enabled by manipulation of the genetic code and are not intended to cover all areas of application. Finally, in the "Outlook" section, we will briefly speculate about potential future uses of ncAAs.

Biosynthetic Methodologies for the Incorporation of Genetically Encoded Noncanonical Amino Acids into Proteins

Methodologies for biosynthetic incorporation of ncAAs into proteins can be divided into "residue-specific" and "site-specific" strategies. An overview of these strategies is presented in figure 1.1. In a "residue-specific" strategy, a single canonical amino acid is replaced by a ncAA wherever the mRNA encoding the protein specifies the canonical amino acid. Figure 1.1a illustrates the resulting "reprogramming" of protein translation. In contrast, "site-specific" strategies (figure 1.1b) involve the replacement of a single amino acid with a ncAA at a desired location within the polypeptide chain. Usually, this approach utilizes suppression of a stop, or nonsense, codon with an appropriately charged suppressor tRNA, although some additional approaches have also been explored.

Background: key advances that enabled modern ncAA incorporation methodologies. Characterization of the components of the biosynthetic protein translation apparatus has provided insights into how the genetic code functions and how the code can be manipulated to perform unnatural functions. Among the key results in this area are Chapeville and coworkers' "Raney Nickel" experiments (12), which provided strong support for Francis Crick's adaptor hypothesis (13). The adaptor hypothesis proposed that small nucleic acids (now called transfer RNAs, or tRNAs) could provide a means by which a sequence of nucleic acids could be translated into a corresponding polypeptide sequence, with the specific pairing of amino acids and tRNAs ensuring the fidelity of the genetic code. Chapeville and coworkers were able to show for the first time that chemical manipulation of aminoacyl-tRNAs can alter the way in which an RNA transcript is decoded. The researchers treated cysteinyl-$tRNA^{Cys}$ with Raney Nickel to convert cysteinyl-$tRNA^{Cys}$ to alanyl-$tRNA^{Cys}$ and compared the results of polypeptide synthesis after supplementing ribosomal preparations with cysteinyl- or alanyl-$tRNA^{Cys}$ and an RNA transcript coding for cysteine. Polypeptide synthesis was supported in each case, and polypeptides produced in reactions containing cysteinyl-$tRNA^{Cys}$ and alanyl-$tRNA^{Cys}$ contained only cysteine and alanine, respectively. These results showed conclusively that misacylated tRNAs remain substrates of the ribosome and that they support protein synthesis. The ability of the ribosome to use misacylated tRNAs in translation implies that ncAA incorporation into protein is possible in theory. If a ncAA can be attached to a particular tRNA, the protein translation apparatus will likely accept the "misacylated" tRNA as a substrate and place the ncAA into a growing polypeptide chain in response to a particular codon.

The "Raney Nickel" experiments also demonstrated the feasibility of performing chemistry directly on tRNA molecules, and many groups have since extended the types of chemistry that can be performed on tRNAs. In the 1970s, experiments performed in the Menninger group demonstrated that (i) Lys-tRNA^{Lys} could be acetylated specifically at the ε-amino group of lysine, and (ii) AcLys-tRNA^{Lys} could be incorporated into proteins in a cell-free protein synthesis system in response to lysine codons (14). These experiments were the first to show that chemically synthesized aminoacyl-tRNAs containing a ncAA could be utilized by the biosynthetic protein synthesis machinery. Hecht and coworkers expanded on the work of Menninger and coworkers by developing a general route to the chemical acylation of tRNAs (15, 16). Their method involved chemical synthesis of the acylated dinucleotide pCpA, the generation of tRNA missing its last two RNA bases, and enzymatic attachment of the acylated pCpA to the truncated tRNA. Further improvements to chemical acylation techniques were elucidated in the 1980s (17, 18). These synthetic methods have proven crucial in the generation of tRNAs bearing a wide variety of ncAAs for use with site-specific ncAA incorporation into proteins.

The identification and detailed characterization of aminoacyl-tRNA synthetases (aaRSs), the enzymes responsible for joining together amino acids and tRNAs in living organisms, have also been crucial for enabling incorporation of ncAAs into proteins. Early work (19, 20) by Berg and others established that aaRSs are responsible for catalyzing the attachment of tRNA and amino acids using the free energy gained from hydrolysis of adenosine triphosphate (ATP) to adenosine monophosphate (AMP) in a two-step process:

$$\text{Amino Acid (AA)} + \text{ATP} \xrightleftharpoons{\text{aaRS}} \text{AA-AMP} \quad \text{(Activation)} \qquad (1.1)$$

$$\text{AA-AMP} + \text{tRNA}^{AA} \xrightleftharpoons{\text{aaRS}} \text{AA-tRNA}^{AA} + \text{AMP} \quad \text{(Aminoacylation)} \qquad (1.2)$$

Characterizations of the enzymatic activities of aaRSs have established that these enzymes are extremely important for maintaining the fidelity of protein translation (21) and that ncAAs can serve as kinetically efficient substrates of aaRSs (22). These findings imply that understanding how to manipulate the enzymatic activities of aaRSs can facilitate the specific, enzymatic linkage of ncAAs and tRNAs.

Additional characterizations of aaRSs and tRNAs laid the groundwork for adding amino acids to the genetic code by adding new translational components to living cells. One of the important ways in which aaRSs maintain the fidelity of the genetic code is by selectively recognizing and aminoacylating their cognate tRNAs. For example, the methionyl-tRNA synthetase (MetRS) must conjugate methionine specifically to $tRNA^{Met}$ without aminoacylating any other tRNA (21). This specific recognition is critical for maintaining genetic code fidelity, but the mechanisms of specificity are not conserved from species to species. As early as 1963, researchers started to realize that the aaRSs from different species recognize their cognate tRNAs by distinct mechanisms. Doctor and Mudd first observed this phenomenon when they discovered that an aaRS from one species cannot always aminoacylate the corresponding tRNA from another species (23). Smith and coworkers proposed that the structural basis for these observations is aaRS recognition of structural features in tRNA molecules that differ from species to species (24). This hypothesis has since been investigated thoroughly using structural and biochemical experiments, resulting in the detailed characterization of how cognate aaRSs and tRNAs are properly paired in different organisms (25). Another early breakthrough that eventually facilitated adding new amino acids to the genetic code was the discovery and characterization of suppressor mutations in *E. coli* (26-28). These mutations, which result

in the conversion of a "stop" message into a message coding for an amino acid, were found to mutate the tRNA anticodon sequence from nucleotides specifying a sense codon into nucleotides specifying a termination codon (24). These findings suggested that using codons normally reserved for the termination of protein synthesis to instead specify an amino acid could be a way to add an amino acid to the genetic code. The combined exploitation of the species specificity of aaRS-tRNA recognition and suppressor tRNAs eventually enabled researchers to dictate the coding of a 21st amino acid, first in vitro and later in living cells and organisms (see below).

Residue-specific incorporation strategies. ***Conceptual advances.*** Even before the components of the translation apparatus were characterized, scientists recognized that ncAAs could be incorporated into the proteins of multi- and single-celled organisms. These earliest examples employed ncAAs that were isosteric or structurally similar to canonical amino acids. During the 1950s and 1960s, ncAAs were employed extensively in studies that involved microorganisms; this work was reviewed thoroughly in 1962 (29). These early experiments, especially those performed in the laboratory of Georges Cohen, laid the foundation for later advances in the field. Cohen and colleagues were the first to use a medium replacement strategy that we now refer to as a "medium shift." They also demonstrated the utility of auxotrophic strains for achieving high levels of ncAA incorporation into proteins. Furthermore, the Cohen laboratory recognized that, "these analogs [and surrogates] become powerful tools for the study of: (a) the specificity of the protein-forming mechanism and (b) the variation of enzyme activity and affinity that occurs with increasing numbers of incorporated analog molecules" (30). These

observations are still relevant today. We now know with some level of detail that incorporation of a particular amino acid into proteins depends on the ability of the protein translation apparatus of a particular cell or organism to tolerate noncanonical substrates. “Variation of enzyme activity” by the incorporation of ncAAs is still studied today and now forms part of a larger effort to perturb protein structure and function with ncAAs.

After the relatively sparse application of ncAAs to scientific problems in the 1970s and 1980s (31, 32), researchers began seriously revisiting the idea of incorporating ncAAs into proteins residue-specifically in 1990. Hendrickson and coworkers demonstrated the complete replacement of methionine by selenomethionine (**1**, scheme 1.1) in thioredoxin in 1990 (33). Their approach involved growing methionine-auxotrophic bacteria in medium containing a limited supply of methionine but ample amounts of selenomethionine. As the bacteria grow, they use up the supply of methionine, forcing incorporation of selenomethionine in place of methionine. Induction of protein expression after the depletion of methionine in the medium enabled a high level of selenomethionine incorporation into proteins. This method for incorporating selenomethionine into proteins has been adapted for use with a number of other ncAAs by the Budisa laboratory (22). The Tirrell laboratory and others have utilized an alternative approach involving extensive washing of bacterial cells in between growth in a medium containing all canonical amino acids and protein expression in medium containing ncAAs (34). With these protocols, scientists have developed methods for the incorporation of a large set of ncAAs into proteins residue-specifically, primarily by engineering *E. coli* expression hosts and aminoacyl-tRNA synthetases. These contemporary residue-specific incorporation techniques will be described in the following subsection.

Experimental approaches. In most cases, efficient aminoacylation of a tRNA with a noncanonical L-amino acid is sufficient to enable the residue-specific replacement of a canonical amino acid with the ncAA of interest. Thus, the genetic code can be intentionally "reinterpreted" to code for one or more ncAAs by controlling what substrates get attached to specific tRNAs. The conceptual approach is depicted in figure 1.1a, and specific strategies for intentional tRNA misaminoacylation are shown in figure 1.2.

The simplest approach to ncAA incorporation involves replacement of a canonical amino acid with a close structural analog in *E. coli*. With an appropriate analog, a cell strain auxotrophic in the amino acid to be replaced, and a technique for depleting the canonical amino acid from the medium prior to expression of proteins of interest, incorporation of a ncAA at genetically encoded positions can be essentially quantitative (if the ncAA is recognized efficiently by the protein translational machinery of the host). Figure 1.2a illustrates a typical procedure for the incorporation of homopropargylglycine (Hpg, **2**, scheme 1.1) in place of methionine in proteins produced by *E. coli*. Methionine-auxotrophic *E. coli* cells are first grown in minimal media containing all twenty canonical amino acids. Upon reaching a sufficient optical density, cells are washed to remove methionine from the medium and resuspended in the expression medium, which usually contains high concentrations of ncAA (in this case, Hpg). Alternatively, cells can be grown in medium that contains both methionine in small quantities and a ncAA in large quantities. By the time cells have reached an optical density sufficient for protein expression, the concentration of canonical amino acid has been depleted, resulting in the same effect as removal of the canonical amino acid through washing (this is sometimes called selective pressure incorporation). After the medium is depleted of the canonical amino acid to be

replaced and supplemented with large amounts of the ncAA, all proteins synthesized within the cells will contain ncAAs. If a particular protein is to be studied, standard expression, purification, and characterization techniques can be employed to isolate the desired protein. On the other hand, if a proteome-wide response to a particular stress or signal is to be studied, the set of newly synthesized proteins can be isolated and studied using approaches to be described in the "Applications: Proteomics" subsection.

Researchers have employed several variations of medium shifts and selective pressure incorporation in recent years. The techniques have been extended to mammalian cell lines for monitoring the production of newly synthesized proteins (35), and attempts have also been made to extend the procedure to yeast expression systems (36, 37). Recently, reports have indicated that two or three canonical amino acids can be replaced by ncAAs simultaneously in *E. coli*, allowing for more drastic reinterpretations of the genetic code (38, 39).

While the medium shift procedure can enable the incorporation of a number of ncAAs into proteins, further expansion of the number of translationally active amino acids available requires additional engineering of expression hosts. To date, most work in the field has focused on altering aminoacyl-tRNA synthetase (aaRS) activity in *E. coli* to enable additional ncAAs to serve as protein building blocks. These approaches, which are performed in combination with medium shifts, are summarized in figure 1.2b–d. The kinetics of amino acid activation appear to dictate the translational activity of most ncAAs (40). When amino acid activation kinetics are slow, ncAAs cannot be joined to tRNA molecules at a high enough rate to support protein synthesis. Increasing the concentrations of an aaRS by outfitting an *E. coli* expression strain with a plasmid-borne copy of an

endogenous aaRS raises the aminoacylation activity of the host strain, enabling some rather poor aaRS substrates to support protein synthesis in the host. Figure 1.2b illustrates this strategy for the leucine analog hexafluoroleucine (Hfl, **3**) (41). LeuRS overexpression in *E. coli* cells auxotrophic in leucine production enables quantitative replacement of Leu by Hfl.

Sometimes simple overexpression of an aaRS is not enough to overcome poor aaRS activity toward noncanonical substrates. However, the judicious mutation of an aaRS can greatly improve the activation kinetics of nonstandard substrates and thereby lead to more ncAAs that can be quantitatively incorporated into proteins. Computational and screening approaches have proven useful in this regard (42-44). For example, as shown in figure 1.2c, the bulky methionine analog azidonorleucine (Anl, **4**) has only very weak translational activity in an *E. coli* host overexpressing wild-type methionyl-tRNA synthetase (MetRS) (45). However, when *E. coli* is outfitted with one of several MetRS active site mutants identified in high-throughput screens for translational activity in the presence of Anl, the bulky azide amino acid can quantitatively replace Met (46, 47). Kinetic characterizations of the mutant MetRSs revealed vastly improved activation of Anl.

Occasionally, efficient activation of an amino acid substrate is not sufficient to enable the incorporation of a particular ncAA into proteins because of proofreading mechanisms in aaRSs. Several aaRSs have a second active site responsible for ensuring that tRNAs are aminoacylated with their cognate amino acids; these editing active sites can discriminate between amino acids that differ by as little as a methyl group (e.g., valine versus isoleucine) and cleave incorrectly aminoacylated substrates (21). However, attenuation of aaRS editing activity can substantially increase the promiscuity of the enzyme, enabling aminoacylation of a greater pool of substrates. Figure 1.2d portrays such

an approach with leucyl-tRNA synthetase (LeuRS) and the substrate norleucine (Nrl, **5**). Mutation of a critical threonine to a much bulkier tyrosine in the editing active site of LeuRS greatly impairs the editing function of the enzyme, leading to high translational activity of amino acids known to be activated by LeuRS (48-51).

In some ways, manipulating the editing and aminoacylation activities (through overexpression and/or mutation) of aaRSs represent complementary approaches to ncAA incorporation. These techniques for manipulating aaRSs for residue-specific incorporation were all developed in *E. coli*, but work from the Hang laboratory suggests that mutant aaRSs from *E. coli* can be employed in other microbes such as *Salmonella typhimurium* (52). Site-specific incorporation approaches using mutated aaRS (discussed below) have been very successful in mammalian cells, suggesting that residue-specific ncAA incorporation approaches in mammalian cells with mutated aaRSs are also worth investigating. While most aaRS manipulations are performed via the introduction of additional copies of aaRS genes on plasmids, genomic manipulations of aaRSs also hold some promise, as evidenced by the recent report of Abdeljabbar et al. (53). Manipulations of aaRS activities through aaRS overexpression, mutation, and genomic manipulation and the application of these techniques to additional organisms will continue to increase the power of methods for the residue-specific incorporation of ncAAs into proteins.

Combining residue-specific incorporation and in vitro protein synthesis. The key requirement for achieving the global replacement of a canonical amino acid in a protein is usually the efficient aminoacylation of the appropriate tRNA. Several groups have approached this aminoacylation problem in conjunction with cell-free protein synthesis,

forming a powerful combination because of the precise control over the components of the translational machinery possible in cell-free environments. Cell-free protein synthesis with chemically acylated tRNAs (54), aaRS-catalyzed tRNA aminoacylation (55), and ribozyme-catalyzed tRNA aminoacylation (56, 57) have all been demonstrated with residue-specific ncAA incorporation. Several groups have also shown that sense codons can be reassigned to residues containing noncanonical backbones such as α-hydroxy acids, *N*-methyl amino acids, and *N*-substituted glycines (poly *N*-substituted glycines are also called peptoids) (58-63), and the Hecht group has made some progress in engineering *E. coli* ribosomes to accept D-amino acids as translationally active substrates (64, 65). Furthermore, powerful nonribosomal methods to synthesize genetically encoded small molecules and polymers are also emerging (11, 66). All of these approaches enable the production of genetically encoded polymers with compositions that are substantially different from those of naturally occurring proteins, and these polymers may have properties that are vastly different from naturally occurring biopolymers. Future applications of these genetically encoded polymers could prove to be very powerful.

Site-specific incorporation approaches. ***Conceptual advances.*** The first examples of site-specific ncAA incorporations into proteins combined the use of chemical aminoacylation techniques, stop codon suppression, and species-specific recognition of tRNA molecules using cell-free protein synthesis (67, 68). In 1989, the groups of Chamberlain and Schultz each reported strategies for the in vitro incorporation of a single ncAA into polypeptide chains in response to amber codons contained within genes coding for proteins of interest. A key component in each of these systems was the combination of an in vitro translation

system from one species and a suppressor tRNA molecule from another species that was not recognized by the aaRSs of the in vitro translation system. The inability of the aaRSs of the in vitro translation system to recognize the tRNA (chemically acylated with the ncAA of interest) makes the tRNA orthogonal to the translation system, a recurring concept in site-specific incorporation of ncAAs into proteins.

Although powerful and quite general, chemical aminoacylation techniques used with in vitro suppression are limited by the amount of the acylated suppressor tRNA that can be generated in a somewhat technically demanding process. For this reason, researchers in several laboratories initiated research aimed at the development of stop codon (or nonsense) suppression techniques for the incorporation of ncAAs into proteins inside living cells or organisms. The move from in vitro to in vivo suppression systems required the development of additional orthogonal components to ensure both the fidelity of the genetic code and the fidelity of ncAA incorporation at specified locations (69-71). Figure 1.3 illustrates the three key criteria that must be met in order to establish the orthogonality of additional translational components for nonsense suppression in living cells. First, the suppressor tRNA to be added to the cell (figure 1.3a) must not be a substrate for any of the wild-type aaRSs already present in the cell in order to ensure that only the ncAA of interest is used to decode nonsense codons. Next, the aaRS to be added (figure 1.3b) must specifically recognize both its suppressor tRNA and ncAA substrates. Finally, the ncAA (figure 1.3c) to be added to the genetic code must not be a substrate for any of the wild-type aaRSs in the cell. Failure to meet these strict criteria may result in ncAA incorporation in response to sense codons, or canonical amino acid incorporation in

response to stop codons; both of these situations reduce the fidelity of protein translation. A properly functioning orthogonal pair is illustrated in figure 1.3d.

This challenging orthogonality problem was solved partially by a number of groups before an integrated solution was reported. First, the RajBhandary group reported a system in which an orthogonal tRNA-aaRS pair was required for in vivo synthesis of full-length genes containing an amber codon (69). In this work, *E. coli* $tRNA^{Gln}$ and GlnRS were adapted for use in mammalian cell lines as an orthogonal tRNA-aaRS suppressor pair. Differences in tRNA recognition between *E. coli* and mammalian cells enabled *E. coli* GlnRS to selectively aminoacylate *E. coli* $tRNA^{Gln}$ with glutamine and suppress an amber codon in a CAT reporter gene in multiple mammalian cell lines. In 1997, the Schultz group reported the first attempt to evolve an orthogonal tRNA-aaRS pair in *E. coli* (71). They described the rational design of an *E. coli* $tRNA^{Gln}$ variant and directed evolution of a GlnRS mutant with improved selectivity for the $tRNA^{Gln}$ variant compared to wild-type $tRNA^{Gln}$ in *E. coli*. Though they did not find a GlnRS variant that could recognize the mutant $tRNA^{Gln}$ better than the wild-type $tRNA^{Gln}$, their work demonstrated the feasibility of engineering recognition patterns in an aaRS-tRNA pair. Efforts to use an *S. cerevisiae* tRNA-aaRS pair in *E. coli* were also undertaken during this time period (72). While the tRNA-aaRS pair was found to be orthogonal in *E. coli*, no suitable GlnRS mutant capable of selectively charging a noncanonical glutamine analog was identified. Furter demonstrated the first functioning suppression system able to incorporate a ncAA in response to an amber codon in *E. coli* (70). This system utilized a yeast phenylalanine tRNA-aaRS pair to incorporate *p*-fluorophenylalanine (**6**) into proteins site-specifically in response to an amber codon. Although this system was site-specific with respect to the

incorporation of the ncAA, the system fell short of being fully orthogonal because the yeast PheRS still recognized phenylalanine (Phe) and incorporated Phe in place of **6** in approximately thirty percent of the model proteins expressed, and small amounts of **6** were found at other Phe positions within the protein, also.

In the late 1990s and early 2000s, researchers developed additional translationally active tRNA-aaRS pairs by exploiting kingdom-specific tRNA-aaRS recognition elements. While early efforts to establish almost completely orthogonal pairs in mammalian cells were achieved with rational design alone (73, 74), most work in *E. coli* employed directed evolution techniques to improve the orthogonality of existing tRNA-aaRS pairs imported from other organisms (74, 75). The efforts of Schultz and coworkers were particularly important in this regard. Wang and Schultz established a general selection system for isolating tRNA-aaRS pairs orthogonal to the translational machinery of *E. coli* (75). This selection system was used with a nearly orthogonal *Methanococcus jannaschii* $tRNA^{Tyr}$-TyrRS pair to further improve its suppression of amber stop codons in *E. coli*. The Schultz group was also the first to establish high-throughput selection and screening methods for isolating aaRSs with altered amino acid specificities (76). Researchers have performed site-specific incorporation of several dozen ncAAs into proteins produce in *E. coli* via combination of these methods for the generation of orthogonal tRNA-aaRS pairs and aaRS-amino acid pairs. Early work by the group of Yokoyama demonstrated that the "orthogonal pair" strategy could be applied in mammalian cells (77) without the use of evolutionary methods. Additional engineering work has enabled the development of site-specific incorporation methods in yeast (78) and improvement of strategies applicable to mammalian cells (79, 80). These techniques will be described in more detail below.

Experimental approaches. Site-specific incorporation of ncAAs into proteins requires the use of specialized translational components in conjunction with codon suppression in order to "add" an amino acid to the genetic code. Figure 1.4 outlines various ways to achieve this goal; most of these approaches involve a suppressor tRNA that is aminoacylated with a ncAA of choice. This tRNA, which uses a "nonstandard" codon, is appropriately decoded during protein synthesis in either an in vitro or an in vivo translation system.

Chemical acylation provides researchers with the most general strategy for incorporating ncAAs site-specifically into proteins. If an appropriately designed, orthogonal suppressor tRNA is available for a given translation system, it can be chemically acylated with a very broad range of ncAAs. When coupled with a gene containing the codon to be suppressed, nearly any protein can be synthesized containing the ncAA of choice at a specific site. Figure 1.4b depicts this process in an in vitro translation system. The in vitro translation system, gene expression, and protein purification are essentially the same as in systems lacking the chemically acylated suppressor tRNA. The protein yields achievable by using chemically acylated tRNAs in an in vitro translation system are usually in the microgram range. However, a number of scientific questions can be thoroughly studied even with a small amount of protein containing noncanonical amino acids (81). Extensions of chemical acylation to cellular systems have also proven to be quite fruitful. As an example of one elegant approach, the Dougherty and Lester groups have studied ion channels and other membrane proteins in *Xenopus* oocytes at length using chemical acylation techniques. In their system, depicted in figure 1.4c, an mRNA containing a nonsense codon at the site specified for ncAA incorporation and an orthogonal suppressor tRNA acylated with a ncAA of choice are injected into the oocyte.

Electrophysiology and other techniques can be used to interrogate channel function and will be discussed below in the "Applications: Membrane proteins" subsection.

The enzymatic, site-specific incorporation of ncAAs into proteins in living cells or organisms is technically simpler than chemical acylation-based methods if appropriate genetic components are available. Enzymatic approaches in living systems require an orthogonal aaRS-tRNA pair that can be expressed in the host cells or organism of interest. This pair, and the ncAA to be incorporated in response to the "nonsense" codon (or other nonstandard codon), should meet the strict requirements for orthogonality discussed above. The ncAA should also be able to access the cytoplasm of the expression host by means of passive or active cellular transport mechanisms. If these conditions are met, site-specific methodologies become quite powerful. Figure 1.4d depicts the incorporation process in *E. coli*. A plasmid-based system encoding constitutively expressed orthogonal tRNA and aaRS genes is transformed into an appropriate strain along with an inducible gene that codes for the appropriate "nonsense" or other specialized codon. For strategies related to the development of orthogonal tRNA-aaRS pairs the reader is referred to several recent reviews (82-87). The resulting system can then be treated essentially like any bacterial expression system. After cell growth in minimal (76) or rich medium (88), the ncAA is added to the medium and protein expression is induced. Cells are harvested after induction, and the protein of interest can then be isolated using standard purification methods. One of the primary advantages of this method over residue-specific methodologies is that no medium shift is required before expression; the genetic code is manipulated by adding translational machinery to cells rather than by manipulation of the extracellular environment. However, cellular suppression-based incorporation techniques can sometimes

suffer from low protein yields. Incomplete suppression in these systems is very common and leads to a reduced amount of full-length protein compared to similarly expressed genes encoded entirely by sense codons. Typical yields reported in the literature are in the 1–10 mg/L range, although improved expression systems have recently been described (88-93).

Extensions to the basic strategy of employing an orthogonal tRNA-aaRS pair have expanded approaches to incorporating ncAAs into proteins site-specifically. Methods for evolving orthogonal tRNA-aaRS pairs in *Saccharomyces cerevisiae* have been developed and applied to site-specific incorporation in this eukaryotic model organism (78). Subsequent systems have improved on early work by using genes coding for orthogonal tRNAs that incorporate the A and B box elements required for high-level transcription in yeast (94, 95). Recent reports have demonstrated that orthogonal tRNA-aaRS pairs can also be introduced into the organisms *Pichia pastoris* and *Mycobacterium tuberculosis* and used to incorporate ncAAs site-specifically into proteins (96, 97). Similar methodological extensions have been made to mammalian expression systems. While orthogonal tRNA-aaRS pairs have not been directly evolved in mammalian systems, they have been imported and used for nonsense suppression successfully (77, 98, 99). Again, improvement of tRNA expression has facilitated the development of more tRNA-aaRS pairs for incorporation of noncanonical amino acids into proteins in mammalian cells (79, 80).

Researchers have investigated many additional strategies for improving or expanding the scope of site-specific ncAA incorporation into proteins produced in vitro or in living cells or organisms. Extensive work to improve the types of ncAAs that can be incorporated in response to nonsense codons have focused on the development of new aaRS variants, including those based on the recently discovered pyrrolysyl-tRNA

synthetase (100-103) and by employing aaRSs possessing expanded or altered editing capabilities (104-106). A systematic approach to the creation of mutually orthogonal aaRSs has been reported (107), and this methodology may prove especially useful if the genetic code is expanded to include large numbers of ncAAs in addition to the twenty canonical amino acids. Both in vitro and in vivo attempts to use four-base codons to incorporate ncAAs into proteins have proven somewhat successful (108-111), as have attempts to incorporate either two of the same or multiple, chemically distinct ncAAs into a single protein (109, 112-114). However, the limitations of organism survival place constraints on the manipulations that can be made to the protein biosynthesis machinery. The Chin laboratory has recently developed a system in which a large portion of the protein biosynthesis machinery has been relieved of its requirements to support cellular viability (115). Using the same orthogonality concept described above, a set of mRNA-ribosome pairs that are separate from the wild-type ribosome and mRNA pools were isolated. By freeing the ribosome from its usual responsibilities of supporting all cellular protein synthesis, researchers have successfully evolved ribosome variants that enable more facile expansion of the genetic code. The most notable of these variants is a ribosome capable of efficiently decoding quadruplet codons (116). Building on a previously successful improvement of suppression efficiency in orthogonal ribosomes (117), this quadruplet-decoding ribosome is capable of incorporating multiple ncAAs into a single protein with very high efficiency: one amino acid is introduced in response to a four-base codon, and a second amino acid is incorporated in response to an amber codon. Future work in engineering the ribosome appears to be a very promising approach for improving efforts to site-specifically incorporate ncAAs into proteins.

Applications

Scientists have been extremely successful in inventing and refining techniques for genetically encoding ncAAs and incorporating them into proteins. However, developing incorporation strategies does not always answer scientific questions, except possibly in areas directly related to the study of aminoacyl-tRNA synthetases and the protein translation apparatus. In this section of the review, we will describe approaches to protein science and engineering that have either been significantly impacted or that could be significantly impacted in the future by the employment of ncAAs. We will focus on select recent areas of study and refer the reader to earlier reviews (29, 31, 32) and other recent reviews (22, 84-87, 118-121) for additional uses of ncAAs. Scientists have developed applications of genetically encoded ncAAs covering an extremely broad range of topics.

Protein crystallography. The determination of protein crystal structures using X-ray crystallography is an extremely important part of molecular biology and biochemistry. While reflection data can be collected from any high-quality protein crystal, the lack of heavy atoms in naturally occurring proteins prevents researchers from learning phase information from these samples without additional information. This dilemma, known to crystallographers as the phase problem, has been solved using a number of approaches over the years (122, 123). In the past twenty years, one of the methods of choice for resolving the phase problem has become the direct incorporation of heavy atoms into the structure of proteins and the application of multiwavelength anomalous dispersion (MAD). This technique can be used on proteins by residue-specifically replacing of one of the canonical amino acids in the molecule with an amino acid containing a heavy atom. In 1990,

Hendrickson and coworkers rediscovered Cowie and Cohen's 1957 work in which quantitative replacement of methionine by selenomethionine (SeMet, **1**) was reported (30). SeMet is nearly perfect for X-ray crystallography applications because it is almost identical in shape to Met and possesses a heavy atom appropriate for MAD phasing techniques. Using selective pressure incorporation, Hendrickson et al. were able to demonstrate the near-quantitative replacement of Met by SeMet in T4 thioredoxin. Furthermore, reflection data acquired with the SeMet-containing proteins indicated, "…that MAD phasing of prospective selenomethionyl proteins should be readily feasible since diffraction ratios are in excess of those that have proven adequate for related problems" (33). Indeed, the application of MAD to SeMet-containing proteins was used successfully twice in 1990 to solve the structures of previously uncharacterized proteins (124, 125). Extensions to proteins expressed in mammalian (126, 127) and baculovirus expression systems (128, 129), and other improvements to expression procedures over the years have led to the establishment of MAD phasing on SeMet-containing proteins as a method of choice in protein crystallography (130-133). MAD has become an indispensible technique for modern protein crystallography.

Several additional ncAAs containing heavy atoms have been introduced into proteins with the aim of improving crystallographic techniques. Residues **7–10** have been incorporated residue-specifically into proteins with high replacement of the corresponding canonical amino acids. Telluromethionine (**7**) has been shown to enable the multiple isomorphous replacement (MIR) phasing strategy using laboratory radiation sources (134-136), while selenocysteine (**8**) can be used in place of (137) or in combination with **1** to enable MAD phasing (138). The feasibility of using residues **9** and **10** for structure

determination has also been demonstrated (139, 140). Iodinated ncAAs also appear to be promising for aiding structure determination efforts. Proteins containing site-specifically incorporated *p*-iodophenylalanine (**11**) and 3-iodotyrosine (**12**) have been crystallized and used to produce reflection data on laboratory radiation sources (141, 142). Iodine-generated single-wavelength anomalous dispersion (SAD) enabled the data to be properly phased, and high-quality structures have been determined from these crystals. Although these emerging approaches to incorporating heavy atoms into proteins have yet to impact X-ray crystallography in the same way as selenomethionine, they may aid future structure determination efforts, especially those that can be accomplished in the laboratory rather than at a synchrotron.

Nuclear magnetic resonance spectroscopy. Like X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy is a crucial tool for the structural and biophysical characterization of proteins. Using isotopic labeling techniques enabling the incorporation of ^{13}C, ^{15}N, and/or ^{2}H, into all or part of proteins of interest enables a vast range of studies to be performed on proteins of increasing sizes (143-145). These experiments, however, can be quite complex due to the number and similarity of chemical shifts present in a single experiment. Judicious incorporation of ncAAs residue- and site-specifically into protein samples can eliminate some of the complexities inherent in NMR studies of proteins.

The lack of fluorine in most biological systems and the ready availability of fluorinated amino acid analogs makes ^{19}F NMR an attractive option for simplifying some protein NMR experiments (146, 147). Methods for the residue-specific biosynthetic incorporation of fluorinated ncAAs into proteins have been used extensively in the

generation of ^{19}F NMR samples (147). Fluorinated aromatic residues **6**, **13–20** and difluoromethionine (**21**) are examples of some of the probes that have been used in experiments (147-150). Aided by the extreme environmental sensitivity of fluorine, researchers have used proteins containing these residues to explore several protein properties including structure, folding, and ligand binding. Rule, Pratt, Ho, and coworkers performed a number of structural studies on D-lactate dehydrogenase (LDH) protein samples labeled with fluorinated aromatic amino acids in the late 1980s and 1990s (151-153). Changes in the ^{19}F NMR spectra of the protein in the presence of ligand or spin-labeled lipid molecules gave the researchers a better idea of which residues of the protein were sensitive to ligand and which residues were involved in membrane contacts.

Fluorine NMR has been used in conjunction with folding studies on a number of proteins. Ropson and Frieden were the first to report the use of ^{19}F NMR to study protein folding (154). They conducted equilibrium folding experiments on the intestinal fatty acid–binding protein labeled with 6-fluorotryptophan (**15**). Interestingly, their spectra showed the appearance of some peaks only at intermediate urea concentrations, indicative of a previously unrecognized folding intermediate in this biomolecule. Since 1992, this general technique has been used with numerous proteins in the context of protein folding and ligand binding, suggesting that the method is useful for studying proteins with a variety of structures (150, 155-162). An important extension of ^{19}F NMR on proteins has been the combination of this sensitive spectroscopic method with stopped-flow experimental techniques. Hoeltzli and Frieden studied the kinetics of *E. coli* dihydrofolate reductase (DHFR) in the first report of stopped-flow NMR experiments with fluorinated proteins (163). In combination with complementary fluorescence and circular dichroism refolding

experiments, the researchers determined that DHFR appears to unfold through a pathway involving an intermediate state in which individual amino acid side chains gain a large amount of mobility while the majority of secondary structural elements remain intact.

Although residue-specifically fluorinated proteins enable a wide range of experiments, they can sometimes be complicated by sample heterogeneity (fluorination at a particular residue is usually no higher than 95%), difficulties in the peak assignments of individual fluorine atoms, and perturbations of protein structure and/or function (147, 164). Site-specific incorporation of fluorinated amino acids provides an alternative approach to the generation of isotopically labeled NMR samples. Furter was the first to report one such technique using *p*-fluorophenylalanine (**6**), which was substituted site-specifically in dihydrofolate reductase (DHFR) expressed in *E. coli* (70). The Frieden laboratory has employed Furter's method as a complementary approach to global fluorination. Bann and Frieden used a combination of site- and residue-specific incorporation of **6** into the bacterial periplasmic chaperone PapD and used these proteins to identify at least three distinct steps in the folding landscape (165).

Two groups have recently developed very efficient expression systems that improve on Furter's method for protein fluorination at aromatic residues, enabling essentially quantitative site-specific incorporation of fluorinated amino acids **22** and **23** into proteins for NMR studies (91-93). Hammill et al. demonstrated that, like previous ^{19}F NMR work, **22** serves as an extremely sensitive monitor of the environment surrounding the amino acid. Incorporation of the noncanonical residue into nitroreductase and histidinol dehydrogenase enabled both short- and long-range detection of ligand binding events. Cellitti et al. have employed ^{19}F NMR on site-specifically labeled proteins in combination

with other NMR techniques, and in this work the environmental sensitivity of fluorinated amino acids is again evident in ligand binding experiments. Li et al. have demonstrated that the exquisite sensitivity of ^{19}F NMR also enables site-specifically fluorinated proteins to be probed in intact *E. coli* cells (166).

While fluorinated protein samples can help answer many scientific questions, site-specifically labeled, native proteins would provide scientists with the opportunity to investigate a much wider scope of problems. A consortium of scientists at The Scripps Research Institute and The Novartis Research Foundation led by Peter Schultz and Bernard Geierstanger has developed some very promising approaches to introducing site-specific ^{13}C, ^{15}N, or ^{2}H labels into proteins (also discussed by Jones et al. (167)). Deiters et al. first used ^{15}N-labeled **24** to study sperm whale myoglobin, and isotopically labeled derivatives of **24** have since been used to study a thioesterase domain of human fatty acid synthase (FAS-TE) (91) and a cytochrome P450 enzyme (168). Cellitti et al. demonstrated the incorporation of isotopically labeled, photocaged tyrosine **25** into FAS-TE and showed that UV irradiation allowed for essentially complete decaging at multiple amino acid positions (91). The group also went on to label FAS-TE at eleven different amino acid positions with ^{19}F-labeled **23**, ^{15}N- and ^{13}C-labeled **24**, and ^{15}N-labeled, decaged **25** in order to exhaustively study protein-tool ligand interactions. Figure 1.5 depicts a summary of the chemical shift data acquired in these studies, which indicates that a number of residues are affected by the binding of a tool compound, including amino acids located in disordered loops that may have been difficult to identify using other methods. This impressive, detailed work indicates a promising future for biosynthetically produced, site-specifically labeled proteins in NMR studies. The precedent for studying protein folding and ligand

binding has been firmly established using fluorinated amino acids. The advent of improved site-specific ncAA incorporation techniques for insertion of fluorinated and native, isotopically labeled amino acids should allow for many more detailed protein characterizations via NMR in the future.

Protein folding and stability. Proper protein folding and robust maintenance of a specific three-dimensional structure oftentimes dictate a protein's ability to perform additional functions such as catalysis or binding. Understanding and quantifying the phenomena responsible for maintaining proteins in a specified conformation can provide insight into how to stabilize folded structures and how to design and/or engineer proteins with various functions. Because folded proteins are oftentimes only thermodynamically stable by only a few kilocalories per mol, individual noncovalent interactions such as hydrogen bonds and van der Waals contacts can greatly influence whether or not a protein is able to assume a folded conformation and the kinetics that dictate the speed at which this conformation is reached. Conventional mutagenesis allows for the study of many of the noncovalent interactions that govern protein folding and stability, but the limited side chain structures of the canonical amino acids do not always allow for a full exploration of these phenomena. Incorporating ncAAs into proteins allows for more subtle perturbations than many canonical amino acid mutations allow. Researchers have perturbed proteins both locally and globally with ncAAs in order to study factors governing protein folding and stability.

Local perturbations. The subtle mutations possible with ncAAs allow investigators to perturb a single noncovalent interaction within a large protein structure and study the

perturbation's effects on the overall protein characteristics. In the 1990s, the Schultz laboratory published a series of studies demonstrating the utility of in vitro unnatural amino acid mutagenesis for studying local aspects of protein stability. Mendel et al. incorporated structural variants of leucine, including ncAAs **26–31**, at position 133 of T4 phage lysozyme (T4L) in order to systematically study several effects including enlargement and shrinking of the cavity into which leucine 133 points (169). The researchers discovered that sequential removal of methyl groups from position 133 reduced the overall stability of the protein in a nonlinear fashion, a phenomenon that could also be reproduced in computational studies. These and other findings led the authors to suggest that many noncovalent interactions are important in determining protein stability and that a combination of modeling and unnatural amino acid mutagenesis could be used in the future to shed light on the factors leading to stable, folded proteins. Ellman et al. reported a related study examining the effects of backbone mutations at position 82 of T4L (170). In this work, the thermal stabilities of T4L variants containing ncAAs **32–42** at position 82 were measured used CD spectroscopy and heat inactivation assays. The results indicated that both protein backbone conformational restriction and the angle of the restriction are important factors in determining protein stability. Cornish et al. investigated the effects of β-branched amino acids on the stabilities of two α-helices in T4L using ncAAs and computational models (171). Incorporation of β-branched amino acids at two positions in T4L resulted in protein stabilization in one case and substantial destabilization in another. These findings emphasize the fact that context appears to play a role in the effect a β-branched amino acid has on protein stability.

Substantial efforts have also been devoted to the study of hydrogen bonds in proteins using systematic mutations. Thorson et al. used ncAAs **6** and **37–39** in order to study mutations to particular hydrogen bonds in Staphyloccocal nuclease (SNase) (172). Mutations of two glutamates to the weak hydrogen bond acceptor **37** and two tyrosines to weaker hydrogen bond donor **38** and repulsive lone pair interactor **39** confirmed the existence of two glutamate-tyrosine hydrogen bonds. Stability measurements on protein variants confirmed that the hydrogen bonds in question were responsible for one to two kilocalories per mol stabilization of the folded protein state, confirming previous estimates of hydrogen bond strengths. Studies on hydrogen bonds in α-helices, β-sheets, and β-turns with α-hydroxy acid analogs of leucine and isoleucine (**40** and **41**, respectively) demonstrated that hydrogen bonds in key secondary structural features can be investigated by changing the hydrogen bonding character of the backbone (173-175). These studies all demonstrate the utility of backbone variations in assessing the thermodynamic contributions from hydrogen bonds involving polypeptide main chain atoms. Thorson et al. used noncanonical amino acids to perform a linear free energy analysis on a hydrogen bond in SNase (176). The authors focused on the importance of the hydrogen bond mediated by the hydroxyl group of tyrosine 27. Substitution of tyrosine by amino acids **16**, **42**, and **43**, which have increasingly acidic pK_as, confirmed that the side chain of tyrosine 27 serves as a hydrogen bond donor in SNase. The use of a series of tyrosine analogs eliminated context-dependent, confounding factors that oftentimes complicate the direct assessment of the importance of a particular hydrogen bond in a protein.

The above studies illustrate the power of atomic-level mutations in the analysis of protein stability. However, systematic local perturbation of soluble proteins using ncAAs

has become less common in recent years, perhaps because of difficulties in achieving high protein yields using in vitro protein production methods, or the availability of alternative protein production techniques such as protein semisynthetic methods. The recent site-specific incorporation of α-hydroxy acids into proteins produced in *E. coli* may lead to new possibilities for protein stability studies in the future (177). Regardless of the future of local perturbations of globular proteins, the work outlined above was crucial in setting the stage for the study of membrane proteins using extensive unnatural amino acid mutagenesis, a topic to be discussed below.

Global perturbations. Global changes in the amino acid composition of a protein allows for the study of aggregate effects that may be too small to study individually. The fluorination effect has been a heavily studied topic for quite some time (178), including in the context of biological systems (179-181). Coiled coils have been a particularly useful protein-based system for studying the fluorination effect because the residues mediating protein-protein interactions in this system are well defined. Tang et al. have studied variants of one such multimeric α-helical protein, A1, containing different degrees of fluorination within its hydrophobic core. The investigators found that increasing the amount of fluorine-fluorine contacts within the hydrophobic core by substituting leucine with trifluoroleucine (**44**) or hexafluoroleucine (**3**) led to peptides that were increasingly resistant to thermal and chemical denaturation (41, 182). Son et al. found similar trends in a model coiled coil system in which isoleucine residues were replaced with trifluoroisoleucine (**45**) residues or valine residues were replaced with trifluorovaline (**46**) residues, although the degree of stabilization differed depending on whether isoleucine or

valine was converted into its trifluorinated form (183). Montclare et al. also recently reported that stabilization of the model protein A1 is relatively insensitive to the stereochemistry of fluorinated amino acids incorporated into the protein (184). Upon fluorination, amino acid side chain volumes increase by a significant amount, and size increases may be responsible in part for the improved stabilities of fluorinated coiled coils. In order to investigate this possibility in more detail, Van Deventer et al. utilized the amino acid homoisoleucine (Hil, **29**) (185). Hil has a nearly identical side chain molecular surface area to trifluoroleucine but retains the aliphatic character of canonical amino acids. Interestingly, replacing the leucines in A1 with **29** results in stabilization that is equal to or better than the stabilization of A1 that results from replacement of Leu with trifluoroleucine. These results suggest that side chain size does play a role in dictating protein stability in the context of hydrophobic interactions. However, femtosecond timescale experiments performed on proteins containing solvent-exposed fluorinated and aliphatic amino acids indicate that solvation dynamics near a protein surface change drastically depending on the absence or presence of nearby fluorinated groups (186). Future experiments using size-matched aliphatic and fluorinated amino acids may continue to enable the elucidation of the unique properties of fluorinated biomolecules.

Perturbing the amino acid composition of large proteins can alter protein folding and stability significantly, especially when the proteins in question assume more complicated structures than coiled coils. Although sometimes global replacement of a canonical amino acid within a protein can yield biomolecules with reduced thermal stabilities (187, 188) or an increased propensity for aggregation (189), these effects are quite dependent on the identity of the protein and the ncAA substitutions made. Oftentimes,

such substitutions are tolerated quite well in proteins. For example, Wang et al. report that fluorination of murine interleukin-2 (IL-2) at isoleucine residues results in IL-2 molecules that are nearly as effective as their nonfluorinated counterparts in mammalian cell proliferation assays (190). Budisa and coworkers have proposed using ncAAs as general probes of protein folding (191). In one report out of the Budisa laboratory, researchers investigated the effects of incorporating (4*R*)- and (4*S*)-fluoroproline (**47**, **48**) in place of proline in enhanced green fluorescent protein (eGFP) (192). Incorporation of **47** into eGFP resulted in proteins located in inclusion bodies of *E. coli* that could not be refolded. On the other hand, incorporation of **48** into eGFP resulted in a protein with significantly faster refolding kinetics and overall refolding yields than nonfluorinated eGFP, which the authors attributed to **48**'s higher C^{γ}-*endo* puckering and *cis* isomerization preferences compared to proline. This work provides one of the first examples of a protein that has improved folding properties when a ncAA is incorporated throughout its structure. The properties of single-chain Fv fragments of antibody fragments have also been subjected to proline fluorination with **47** and **48** (193). Interestingly, the stability of the scFv was found to be improved when proline was replaced with **47**, the stereoisomer that was found to be detrimental to eGFP folding. Budisa and coworkers recently employed methionine analogs **5** and **49** in the study of protein misfolding leading to prion disease (194). In this work, they used **5** and **49** because of their increased hydrophobicity and hydrophilicity, respectively, relative to methionine. Global replacement of methionine by **5** in recombinant human prion protein ($rhPrP^{C}$) resulted in less protein aggregation than in $rhPrP^{C}$ containing methionine. On the other hand, replacement of methionine by **49** yielded a much more aggregation-prone version $rhPrP^{C}$ than the methionine version of the protein. CD spectroscopy also revealed

significant differences in secondary structure and unfolding behavior as a function of temperature. These results can be seen in figure 1.6. These interesting experiments suggest that the hydrophobicity of the moieties present at positions in the rhPrP^{C} normally occupied by methionine can significantly impact how well the protein maintains its folded state, and that hydrophilic moieties appear to substantially impact the proper folding of the protein in question. Because methionine oxidation results in the formation of more hydrophilic functionalities at methionine positions, this work suggests that oxidative stresses may play a role in the development of prion disease and other diseases caused by protein misfolding. These techniques may also be applicable to studying a number of other cellular proteins that can undergo oxidation at methionine residues (195-197). Studying protein stability using global perturbations introduced with ncAAs appears to be useful in a variety of settings. Both model proteins and more complex proteins can be perturbed in ways such that protein stability is either negatively or positively affected. Observing these changes in stability provides fundamental information regarding how changing molecular properties results in changes to protein properties as a whole and also provides ideas for engineering proteins with altered stabilities. Thus, global perturbations provide a complementary approach to local, site-specific perturbations. These two approaches add substantial capability to researchers' toolkit for assessing factors contributing to protein stability and engineering more stable, faster folding proteins.

Membrane proteins. Membrane proteins play extremely important roles in organisms from all classes of life and may comprise upwards of one quarter of all open reading frames in the genomes of fully sequenced organisms (198). These proteins also form a large

portion of "druggable" protein targets (199-201). However, structural characterization of this important set of proteins has continued to be plagued by problems with overexpression and crystallization, although there have been some recent advances (for example, see (202-204)). The dearth of detailed molecular information has necessitated the adaptation of other experimental techniques for studying membrane proteins, including the very powerful combination of conventional mutagenesis and the patch clamp technique (205). Several laboratories have found that employing ncAAs in studies of receptors and ion channels further augments the power of more traditional biochemical characterization methods. Investigations of topics including channel architecture, the functional significance of individual amino acids and noncovalent interactions, and mechanisms of ligand binding have been aided significantly with the use of ncAAs.

Exploiting unique functionalities of noncanonical amino acids. Several studies of membrane proteins have employed ncAAs to learn about structural and mechanistic features of particular channels and receptors. Surprisingly, many of these investigations have utilized techniques that take advantage of "highly unnatural" features of noncanonical amino acids, an observation that Dougherty made in a recent review (206). Gallivan et al. employed biocytin (**50**) in order to detect surface-exposed residues in muscle-type nicotinic acetylcholine receptor (nAChR) (207). This work used information on protein expression and the binding of streptavidin probes to proteins present in intact *Xenopus* oocytes to reveal the orientations of several amino acids within the main immunogenic region of the nAChR. England et al. employed ultraviolet light-cleavable amino acid **51** and base-cleavable α-hydroxy acids **32**, **40**, and **52** in order to investigate characteristics of a portion

of the muscle-type nAChR known as the Cys-loop (208, 209). These experiments confirmed the functional significance of the Cys-loop in signaling and helped to establish the disulfide bond connectivity within the loop.

Spectroscopically active ncAAs have proven very useful for studying structural features of G protein-coupled receptors (GPCRs). Fluorescent amino acid **53** has been employed in FRET studies of the GPCR tachykinin neurokinin-2 (NK2) in order to study the structure of the protein (210, 211). In vitro preparations of NK2 containing the fluorescent amino acid were exposed to a ligand labeled with a FRET partner. Distance constraints within the receptor were determined from measured FRET effects, leading to a more accurate picture of the general structure of the receptor. Local mechanistic features of the GPCR rhodopsin were recently probed using *p*-azidophenylalanine **54** as a vibrational spectroscopic probe (212). The unique vibrational signature and sensitivity of the azide to changes in the local electrostatic environment enabled Ye et al. to investigate structural changes in the receptor upon light activation. Placement of **54** at different sites within rhodopsin enabled the elucidation of the order of several sequential helix movements as the protein assumes its active conformation, leading to a general model for GPCR activation (213). This work appears to be extremely promising and should be applicable to any number of membrane proteins due to the exquisite sensitivity of azides to local environment.

Noncanonical amino acids of various sizes have also been employed to examine the properties of voltage-gated potassium channels. The mechanism of channel inactivation in potassium channel Kv1.4 has been studied in mammalian cells using amino acids bulkier than tyrosine (80). The authors observe slower channel inactivation times when a naturally

occurring tyrosine residue is replaced by **24** or **55**. They argue that these observations provide support for a mechanism in which the N-terminus of the channel moves through a side pore in the channel to inhibit the flow of ions through the inner pore. A series of phenylalanine analogs of various sizes (**56–59**) was used to examine the importance of a particular Phe residue at position 233 within the *Shaker* voltage-gated potassium channel (214). Interestingly, channels containing the large aliphatic amino acid cyclohexylalanine **60** in place of Phe still retain function, indicating that the size of Phe is the most important functional characteristic of the residue in question. Phe analogs of various sizes also affect channel behavior in a systematic way, supporting the hypothesis that a Phe to Trp mutation at position 233 alters channel function primarily on the basis of its increased bulk. The above examples exploit properties of ncAAs that make these residues distinct from canonical amino acids. The use of these unique functionalities has enabled the study of a wide variety of structural and mechanistic aspects of membrane proteins, and several of these techniques should continue to provide insights into the functions of numerous receptors and channels.

Atomic-level perturbations. In the absence of three-dimensional structural information, meaningful characterization of the role of a particular amino acid or noncovalent interaction within a protein oftentimes requires atom-by-atom perturbations. Since the Dougherty and Lester laboratories first described the incorporation of noncanonical amino acids into membrane proteins in *Xenopus* oocytes (215), much work involving ncAAs in membrane proteins has exploited very subtle perturbations to study aspects of channel and receptor function (206, 216).

Atomic-level mutations have revealed important information about the roles that individual amino acids play in receptor gating. Hanek et al. reported studies on the nAChR investigating the role of a particular valine in the gating of the receptor (217). These studies uncovered a "pin-into-socket" binding mechanism within the channel by examining dose-response curves of channels substituted with amino acids **61–63**. Another investigation authored by Lummis et al. investigated the role of a particular proline in the gating of a 5-hydroxytryptamine type 3 (5-HT_3) receptor (218). The authors hypothesized that a proline residue located in between two transmembrane helices was involved in the gating mechanism of the receptor. Although conventional mutagenesis of the channel resulted in a loss of receptor gating activity, introduction of amino acids **33**, **47**, **48**, and **64–66** in place of proline resulted in functional channels. Interestingly, the effector concentration for half maximal response (EC_{50}) of the resulting channels correlate very strongly with the propensity for each ncAA replacing proline to assume a *cis* protein backbone conformation. Figure 1.7 depicts the linear relationship observed between *cis*-*trans* isomerization and receptor activation. Based on this striking correlation and additional studies, the authors conclude that the relatively facile *cis*-*trans* isomerization of proline, and not any other properties of this cyclic canonical amino acid, is responsible for the gating characteristics of the 5-HT_3 receptor. The data from their studies also led to a proposal of a new model for the gating mechanism of the receptor (figure 1.7c).

Atomic-level mutations have also provided insight into the importance of particular noncovalent interactions within receptors and channels. Hydrogen bonds involving main chain and side chain interactions have been particularly well studied. In 1995, Nowak et al. made mutations to three tyrosine residues in muscle-type nAChR that included analogs of

tyrosine with varying hydroxyl side chain pK_as and other amino acids lacking hydroxyl groups altogether. Based on dose-response curves of the mutant channels, the authors concluded that only one of the residues examined was involved in a hydrogen bonding interaction through the hydroxyl group (215). Beene et al. used similar atomic perturbations to study four different tyrosine residues present in the 5-HT_3 receptor binding site (219), finding that each residue plays a unique functional role within the channel. More recent work with the 5-HT_3 receptor and ncAAs has determined the roles of two highly conserved residues within loop A of the ligand binding site, one of which appears to form a critical hydrogen bond with the ligand (220). Last, the nicotinic pharmacophore of the $\alpha4\beta2$ neuronal nAChR was investigated with the help of **40** (221).

Hydrogen bonds with main-chain atoms can be perturbed effectively with the aid of α-hydroxy amino acids and other ncAAs. Studies of a conserved proline residue within the M1 region of ligand-gated ion channel receptors using amino acids **32**, **33**, **40**, **52**, and **67** have suggested that proline is favored for its hydrogen bonding characteristics (222, 223). Gleitsman et al. used double-mutant cycles with ncAAs to study hydrogen bonding networks in the muscle-type nAChR (224). Their detailed investigations were instrumental in correcting homology models that had incorrectly predicted the hydrogen bonding patterns based on structures of the acetylcholine-binding protein. Similar studies on an aspartate residue participating in several hydrogen bonds (225) and the hydrogen bonds in a beta sheet structure (226) of the muscle nAChR have also provided information regarding the functional significance of particular hydrogen bonding patterns.

Additional ncAAs have been used to examine other important noncovalent interactions that play important roles in agonist and antagonist binding events (227, 228).

Numerous receptors and channels have a very large number of aromatic residues that directly line the binding site, and as early as 1990, aromatic residue-mediated cation-π interactions were proposed in acetylcholine-binding proteins (reviewed in (229, 230)). The use of ncAAs has contributed greatly to the investigation of potential cation-π interactions between ligands and the ligand binding sites of membrane proteins. Zhong et al. first described the use of a series of aromatic ncAAs to study ligand binding in the muscle nAChR in 1998 (231). Experimental measurements of EC_{50} values of channels containing a number of tryptophan analogs at three positions of the ligand binding site showed very little dependence on the tryptophan analog incorporated. However, at the α149 position, the EC_{50} value varied widely depending on the analogs used. Remarkably, comparison of experimentally determined EC_{50} values with fluorinated tryptophan analogs **13–15**, **68–70** and quantum mechanical calculations of cation-π binding strengths of each of these analogs revealed a very strong correlation. This strong relationship between ligand binding and cation-π binding strength at a single amino acid, depicted in figure 1.8, clearly established the presence of a single cation-π interaction in the nAChR. Subsequent studies have continued to reveal important characteristics of numerous cation-π interactions in nAChR binding events, including differences in the binding of nicotine, acetylcholine, and epibatidine to the receptor (232). Cashin et al. compared nicotine and acetylcholine binding to the nAChR with the drug epibatidine and established that epibatidine has a cation-π interaction similar to acetylcholine (233). Xiu et al. investigated differences between the muscle nAChR and the A2B3 form of the α4β2 neuronal nAChR that likely plays a role in nicotine addiction (234). One key finding that they reported was the existence of a strong cation-π interaction between nicotine and a tryptophan residue in α4β2 homologous to

tryptophan α149 in muscle nAChR, which they confirmed with the same series of fluorinated tryptophan analogs described above. This cation-π interaction between nicotine and tryptophan is absent in the muscle nAChR, explaining a longstanding mystery regarding the selectivity of nicotine for neuronal receptors over the muscle nAChR.

In addition to the extensive studies on cation-π interactions in nAChRs, researchers have employed ncAAs to study potential cation-π interactions in other integral membrane proteins. Studies on the Cys-loop family of receptors have revealed that many proteins within the Cys-loop family bind ligands using cation-π interactions, but utilizing aromatic residues at different structural locations (232, 235-238). Studies on the voltage-gated sodium channel $Na_v1.4$ involving ncAAS have also proven fruitful, identifying cation-π interaction binding sites for tetrodotoxin, calcium binding, and local anesthetics (239-241). Use of ncAAs in the voltage-gated potassium channel *Shaker* yielded insights into the location of a cation-π interaction and the structural conformation of the channel, and these findings helped to resolve discrepancies between computational structural predictions and previous experimental work (242). Last, McMenimen et al. have demonstrated that the tryptophan thought to be responsible for cation-π-mediated magnesium ion blockade in the *N*-methyl-D-aspartate receptor is instead favored for its large, hydrophobic, and flat characteristics (243).

Research efforts highlighted above demonstrate that investigations of membrane proteins with ncAAs can provide a wide variety of structural and mechanistic information ranging from atomic to whole-protein distance scales. To date, the majority of investigations in this area have focused on ligand- and voltage-gated ion channels, but recent reports employing ncAAs in the study of GPCRs (212, 244-246) suggest that efforts

to broaden the applicability of ncAAs to a wider range of membrane proteins are underway. As long as three-dimensional structural information about membrane proteins remains scarce, ncAAs should continue to play an important role in the study of these complex biomolecules.

Posttranslational modification mimicry. Although only twenty amino acids are usually genetically encoded in proteins, nature has devised a huge number of posttranslational modifications that modulate the structure, activity, and localization of proteins (2). The precise effects that these modifications have on proteins can be extremely difficult to study due to the dynamic nature of many of these modifications in living cells and the challenges in purifying or preparing uniformly modified proteins to study in vitro. Proteins bearing authentic posttranslational modifications or close structural mimics of these modifications can be prepared in conjunction with ncAAs. Both chemical routes and direct incorporation of amino acids mimicking posttranslational modifications have been explored and applied to the study of posttranslationally modified proteins.

Chemical approaches. Performing chemistry on proteins forms one major approach to generating molecules containing posttranslational modifications found in naturally occurring proteins. Many protein modification techniques have been developed for the chemical transformation of residues found in naturally occurring proteins (247), and some of these methods have been exploited in the generation of structures mimicking posttranslational modifications. For example, the Shokat laboratory has developed an elegant method for modifying cysteine residues in recombinantly produced proteins in such

a way that they resemble methyl-lysines (248). Semisynthetic methods such as native chemical ligation and expressed protein ligation also provide another efficient chemical route to proteins containing posttranslational modifications (5-9). NcAAs provide a complementary approach to these methods because of the improved yields that biosynthetic protein production offers compared to other preparations of posttranslationally modified proteins. Approaches to employing bioorthogonal chemistry for the selective modification of biomolecules have recently been reviewed from the perspectives of the range of bioorthogonal chemistries explored (249), the particular utility of azide-alkyne click chemistry (250), and chemistries that can be performed on ncAAs in proteins (251).

Several groups have utilized chemistries introduced by the incorporation of ncAAs in order to create mimics of posttranslationally modified proteins. Three separate groups have found chemical transformations that lead to dehydroalanine-containing proteins, two of which (Seebeck et al. and Wang et al.) involve the use of ncAAs (252-254). The Schultz laboratory has demonstrated that analogs of methyl- and acetyl-lysine (**71**–**73**) can be chemically installed in proteins by performing chemistry on dehydroalanine-containing proteins (255), and using dehydroalanine (**74**) as a chemical handle has been proposed as a general starting point for protein modification (256). The Chin laboratory has demonstrated the utility of protecting groups in the development of a method for preparing site-specifically and quantitatively methylated histones. They installed dimethyllysine (**75**) by site-specific incorporation of the protected lysine *N*-tert-butyloxycarbonyl-L-lysine (**76**) and an ingenious protection-deprotection scheme (257, 258). Using the same ncAA (**76**), Virdee and coworkers combined genetic code expansion, intein chemistry, and chemoselective ligation to create a method termed GOPAL (genetically encoded

orthogonal protection and activated ligation) to synthesize diubiquitin chains that have not been chemically accessible by previous means of synthesis (259). The resulting proteins were used to solve the structure of a previously uncharacterized diubiquitin conjugate and to study the specificities of human deubiquitinases. Eger et al. have also demonstrated the utility of a combination of site- and residue-specifically incorporated azide and alkyne amino acids **77** and **78** to chemically synthesize diubiquitin molecules using copper-catalyzed azide-alkyne cycloadditions (CuAAC) (260). This method might be simpler and more versatile than the methods described by the Chin laboratory, but the resulting structures contain unnatural triazole ring linkages rather than authentic linkages found in nature.

The complexities of protein glycosylation have inspired a number of approaches to the chemical and enzymatic synthesis of sugar-decorated proteins (recently reviewed by Gamblin et al. (261)). Part of the appeal of chemical approaches to adding glycans to proteins is the ability to precisely define the structure of a glycan prior to attaching it to a protein without having to worry about the enzymatic efficiencies of glycosyltransferases or potential constraints within the protein translation machinery. Davis and coworkers have recently reported two alkene-based routes to glycosylated proteins using chemistry described above. Each of these methods leads to the construction of quantitatively glycosylated model proteins, although they require the use of proteins containing only one cysteine and one methionine each, respectively (254, 262). The Schultz laboratory demonstrated the modification of ketone-containing amino acid **79** in the Z-domain of staphylococcal protein A with an aminooxy analog of *N*-acetylglucosamine (263). After nearly quantitative reaction with the first sugar, the protein could be enzymatically

modified at the glycan with subsequent sugars to produce more complex carbohydrate structures. Chemical attachment of multisugar structures to ketone-containing proteins was also achieved, although not in quantitative yields. Van Kasteren et al. have used azidohomoalanine (**77**) and homopropargylglycine (**2**) to "click" sugar structures to model proteins using CuAAC (264, 265). Unlike the ketone-mediated couplings, CuAAC afforded nearly quantitative yields when complex carbohydrate structures were ligated to proteins. CuAAC-mediated glycosylation was also combined with cysteine modification to generate doubly glycosylated proteins, although this required the use of proteins containing only one methionine and cysteine residue each. Despite these drawbacks, these chemically produced glycan mimics were used to study glycan recognition and as probes of glycan-binding activity in vivo. One especially interesting observation was that neuronal cells in rat brain tissue sections appear to selectively bind proteins containing GlcNAc modifications, while nearby glial cells do not appear to possess the same binding capabilities. Chemical approaches to the generation of posttranslationally modified proteins appear to be especially useful in glycoprotein synthesis, and the generality of this approach should be very useful for generating quantitatively modified structures for studying several outstanding questions in glycobiology.

Direct incorporation approaches. Ribosomal synthesis of proteins containing appropriately designed ncAAs allows for the creation of some authentic, genetically encoded posttranslationally modified proteins and metabolically stable mimics of other modifications. Significant effort has been dedicated to the genetic incorporation of subtle posttranslational modifications into proteins. The Schultz laboratory has biosynthetically

incorporated sulfotyrosine (**80**) into proteins expressed in *E. coli* and used it in a number of contexts. In an initial report, they demonstrated the improved inhibition of thrombin by sulfo-hirudin versus desulfo-hirudin (266). They also solved the X-ray crystal structure of the sulfo-hirudin-thrombin complex using biosynthetically produced sulfo-hirudin (267). In this case, the genetic encoding of the modification enabled the generation of large protein samples quantitatively sulfated at a specific position, a feat difficult to achieve by any other protein production method. More recently, Schultz and coworkers have explored the utility of sulfated antibodies in the context of directed evolution (268, 269); these results will be discussed below in the "Evolution" subsection.

Lysine modifications are particularly important posttranslational modifications. For example, lysine acetylation, which results in the amino acid N^{ε}-acetyllysine (**81**), impacts the function of many proteins. A report has described the biosynthesis of proteins containing this modification, including the naturally lysine-acetylated protein rat mitochondrial manganese superoxide dismutase (100). Chin and coworkers have recently applied this strategy to the study of histone acetylation (270). The preparation of histone H3 quantitatively acetylated at lysine 56 enabled the mechanistic study of a number of previously proposed effects of histone acetylation on histone properties. For example, FRET experiments indicated that reconstituted nucleosomes acetylated at lysine 56 of H3 have DNA breathing increased by seven times over nonacetylated nucleosomes, explaining previously observed changes in gene expression from H3 mutants lacking lysine at position 56. Suga and coworkers have also investigated the effects of lysine acetylation and methylation on heterochromatin protein 1 recognition of histone H3 N-terminal peptides by reprogramming the genetic code to combinatorially incorporate acetylated and methylated

lysine residues into genetically encoded peptides (271). Aside from histone acetylation, cyclophillin A (CypA), a protein with important roles in immunosuppression and viral infection, is also acetylated in human cells. Using an orthogonal acetyllysyl-tRNA synthetase/$tRNA_{CUA}$ pair and amber suppression, Lammers et al. produced homogeneously and site-specifically acetylated recombinant CypA in *E. coli* (272). This approach enabled structural and biophysical analyses on acetylated CypA for the first time and revealed important roles of acetylation on CypA function. These roles include suppression of the protein's catalytic activity, recognition of the HIV-1 capsid, cyclosporine binding, and calcineurin inhibition.

Xie et al. have genetically encoded a phosphotyrosine mimic in proteins (273). While phosphotyrosine is subject to hydrolytic and enzymatic cleavage, amino acid **82** is not due to the more stable linkages involved. The single negative charge on the amino acid is a reasonable substitute for the doubly charged phosphate group present in phosphorylated amino acids. DNA binding studies with a fragment of human signal transducer and activator of translation-1 (STAT1) containing amino acid **82** showed that it had an intermediate apparent affinity for DNA in between analogous phosphorylated STAT1 and unphosphorylated STAT1. These results indicate that the phosphotyrosine mimic may be used to study the effects of phosphorylation at specific protein sites, although the mimic is not a perfect analog of authentic phosphorylation. The amino acid 3-nitrotyrosine (**83**), a residue commonly associated with disease and oxidative damage, has also been incorporated into proteins produced in *E. coli* (274). Studies on the catalytic activity of manganese superoxide dismutase (MnSOD) indicate that nitration of a particular tyrosine in the protein greatly impacts the catalytic activity of MnSOD, suggesting that

future studies of tyrosine nitration in proteins may shed additional light on the molecular and cellular effects of this modification. As stated above, protein modification by glycosylation is an area of widespread interest. However, biosynthetic incorporation of glycans into proteins has been a rather difficult task, and early reports on the subject (275, 276) have since been retracted (277, 278). One report of the successful incorporation of a glycosylated amino acid into proteins using an in vitro translation system has appeared in the literature (279), suggesting that future efforts to directly incorporate glycosylated amino acids into proteins may yet prove fruitful.

More than half a dozen posttranslational modifications or mimics have been incorporated into proteins using biosynthetic approaches. These approaches to incorporating small posttranslational modifications and chemical approaches to incorporating modifications such as glycan structures into proteins appear to be relatively straightforward. The successful creation of uniformly modified proteins and elucidation of many of significantly altered functional properties of these proteins suggests that ncAAs will continue to provide researchers with powerful, specific tools for studying the effects of posttranslational modifications on biological processes.

Immune modulation. Collaborative efforts at The Scripps Research Institute have recently raised the possibility of using ncAAs as aids in the development of vaccines. Many approaches to developing vaccines have previously been described that include the use of viruses, T-helper epitopes, and sophisticated adjuvants (280-282), but difficulties remain in developing robust immune responses against self-proteins. Grünewald et al. first presented the idea of incorporating immunogenic amino acids into self-proteins in 2008 (283). In this

work, researchers incorporated *p*-nitrophenylalanine (pNO_2Phe, **84**) into murine tumor necrosis factor-alpha (mTNF-α) based on previous reports of the highly immunogenic nature of nitroaryl groups. Immunization of mice with purified mTNF-α containing **84** resulted in a very high level of mTNF-α-specific antibodies in the serum of the treated animals. In contrast, immunizations using wild-type mTNF-α or a variant containing a tyrosine to phenylalanine mutation did not result in significant antibody titers. The utility of the immunization procedure was tested using a mouse model of severe endotoxemia. Figure 1.9 depicts the results of challenging mice with bacterial lipopolysaccharides. While the presence of mTNF-α in the serum usually contributes to septic shock and death of the challenged mice, these data indicate that the presence of antibodies against mTNF-α can significantly improve the survival rate of the challenge. A follow-up report on this work expanded upon the above studies (284). In this work, the investigators found that the site of pNO_2Phe within mTNF-α affected the strength of the immune response elicited in the mouse. They also note, somewhat surprisingly, that antibodies isolated from the serum of immunized mice do not necessarily recognize peptide epitopes containing the immunogenic amino acid, although further investigation will be necessary to examine this finding in more detail. Finally, the researchers incorporated pNO_2Phe into murine retinol-binding protein (mRBP4) and immunized mice with the resulting protein, again finding robust immune responses after administration of the proteins containing **84**. In each of these papers, the authors stress that they have produced immunogenic materials based on precise molecular manipulations, an approach not possible with other immunization techniques. These results raise the tantalizing possibility that immunogenic ncAAs may lead to a general approach to developing therapies involving recognition of a self-protein or

other weakly immunogenic protein. While application to treatments for cancer or other diseases may still be quite distant, the initial experiments described here show a great deal of promise.

Evolution. As far as scientific inquiries have been able to establish, the genetic code has remained more or less constant since its establishment in the so-called frozen accident (285). Therefore, evolution has operated with a single genetic code for more than one billion years (with some minor variations arising along the way (286)). However, there is no fundamental barrier to performing evolution experiments with alternative genetic codes. Using ncAAs, researchers have recently combined ncAAs and directed evolution to investigate new approaches to genetic code expansion and protein engineering.

Evolution of organisms with altered genetic codes. Little is known about how the standard genetic code was established. Before becoming fixed, changes to the genetic code likely involved either introduction of a new amino acid into proteins or the substitution of one amino acid for another within proteins (286-288). These significant changes likely required substantial adaptations by the organism in order to accommodate the alterations. Two groups have investigated this adaptation process using laboratory evolution techniques. In 1983, Wong grew tryptophan-auxotrophic *Bacillus subtilis* on solid media supplemented with 4-fluorotryptophan (**13**) in place of tryptophan (Trp) (289). After just two rounds of selection and two more of mutagenesis and selection, Wong isolated a mutant that grew logarithmically in a liquid culture supplemented with **13**, but only linearly when the medium was supplemented with tryptophan. These impressive results demonstrated that it

is possible to change an organism's amino acid preference from a canonical one contained in the genetic code to a noncanonical one. Bacher and Ellington performed similar selection experiments using *E. coli* auxotrophic in Trp production and then identified genetic mutations in the evolved strains (290). Three thousand hours of serial dilutions in liquid cultures containing increasing percentages of **13** in place of Trp resulted in strains showing improved growth rates in medium containing 99.97% **12**, although these strains still showed a growth preference for Trp. Identification and characterization of mutations revealed that one mutation in the gene encoding tryptophanyl-tRNA synthetase enables improved discrimination against **13**, but mutations in other genes conferred improved tolerance of the noncanonical substrate to the organism. Bacher et al. also subjected the phage Qβ to selection in media containing 95% 6-fluorotryptophan (**15**) (291). Two independent phage lines were subjected to twenty-five rounds of selection. Seven mutations became fixed in each of these lines, and, surprisingly, these mutations did not involve the alteration of any tryptophan codons. In the case of both *E. coli* and phage, the evolved organisms retained their abilities to grow on tryptophan while acquiring new growth capabilities. These results suggest that organisms can adapt to tolerate ambiguity within their genetic codes, implying that the "ambiguous intermediate" theory of genetic code expansion is plausible (288). Future studies of the ncAA accommodation process may shed additional light on the mechanisms by which changes to the genetic code are accepted in living creatures.

Directed protein evolution. In the past two decades, scientists have developed a number of high-throughput selections and screens enabling the directed evolution of proteins (1, 292-

295) and higher order protein-based systems (296). These efforts have enabled researchers to study the evolution of individual proteins and engineer proteins with new functions (297). Most work in this area has focused on exploring the sequence space defined by polypeptide chains containing the twenty canonical amino acids. However, two groups have employed ncAAs to explore alternative protein sequence spaces using directed evolution. The Tirrell laboratory has focused on the development of functional proteins fluorinated at leucine positions by substituting trifluoroleucine (**44**) for leucine. Global replacement of leucine in chloramphenicol acetyl transferase (CAT) resulted in a protein with greatly reduced half-life of inactivation at 60 °C (188). Two rounds of error-prone PCR and screening for mutants with increased activity after prolonged incubation at 60 °C yielded a mutant with substantially improved thermostability properties. Interestingly, the nonfluorinated mutant protein retained the thermostability of the parent CAT, indicating that mutations introduced during the course of evolution allowed the protein to improve its properties when fluorinated while retaining its original function. This trend is similar to the trend observed when an amino acid in the genetic code of an organism is replaced globally; both organisms and proteins appear to first adapt to accept an ambiguous genetic code before gaining a preference for the altered genetic code. In another set of evolution experiments, Yoo et al. evolved a GFP variant containing **44** in place of leucine using eleven rounds of random mutagenesis and screening using fluorescence-activated cell sorting (FACS) (189). A summary of the progress observed during the course of the directed evolution experiment is depicted in figure 1.10. The fluorescence of cells expressing various mutants in the presence of leucine or **44** indicates that nonfluorinated GFP mutants retain (and perhaps even slightly improve) their function, similar to the trend

observed in CAT evolution. The authors also found a marked improvement in the folding kinetics of the evolved GFPs in the absence of fluorination. Interestingly, several of the mutations observed in the final isolated mutant were also observed after screening for "folding-enhanced" GFP variants (298), suggesting that evolving proteins containing ncAAs may be a general method for improving the folding properties of proteins.

The Schultz laboratory has recently applied site-specific incorporation methodologies to directed evolution problems. Their work to date has focused on using ncAAs that are known to have affinities for particular molecules to engineer antibody fragments. In an initial report, they described the development of a phage display system that can be used in conjunction with an orthogonal aaRS-tRNA pair to encode for antibody fragments containing twenty-one amino acids (269). In this paper, they demonstrated the utility of encoding sulfotyrosine (**80**) in a saturation mutagenesis library. Despite an expression bias against antibody fragments containing **80**, they were able to isolate variants from a naïve germline library containing sulfotyrosine after panning for binders against the HIV protein gp120. Perhaps these results were not too surprising since naturally occurring high-affinity antibodies against gp120 are known to be sulfated (299, 300), but the demonstration does prove that ncAAs can be used effectively in phage display applications. A follow-up report has shown that functional antibody variants can be isolated from saturation mutagenesis libraries of the tyrosine-sulfated 412d antibody fragment (268), although affinity maturation of the fragment proved challenging given the high affinity of the starting protein. Finally, this technique has also been applied to acyclic sugar binding (301). The amino acid *p*-boronophenylalanine (**85**) was encoded in a saturation mutagenesis library of phage-displayed antibody fragments, and the library was panned

against a glucamine resin. Many of the clones isolated after three rounds of panning were found to contain amino acid **85**. Again, these results might have been expected based on the well-known propensity of boronates and diols to form boronate esters. However, these proof-of-principle experiments suggest that future libraries of antibody fragments containing ncAAs may yield high-affinity binders to antigens that are normally difficult to target. Taken as a whole, directed evolution experiments with proteins containing ncAAs appear to be promising for evolving proteins with properties that are difficult to introduce using only canonical amino acids, including particular molecular recognition and catalysis events. Furthermore, both protein and organism evolution experiments with ncAAs may provide insights into the kinds of adaptations necessary to maintain evolutionary fitness as amino acid compositions are changed or expanded.

Proteomic studies. Labeling and identifying the proteins expressed in biological systems can provide great insights into the inner workings of these systems, including spatial and temporal information about the proteome (302, 303). Recent work with ncAAs has demonstrated that reactive amino acid analogs can function as effective tags for labeling and identifying newly synthesized proteins in a range of biological systems. The versatility of bioorthogonal chemistry enables the use of the same set of ncAAs for visualization and identification of newly synthesized proteins.

Fluorescent labeling of newly synthesized proteins. Although a number of effective genetic, enzymatic, and chemical strategies for fluorescently labeling proteins have been developed (recently reviewed by Sletten and Bertozzi (249)), most of these approaches

have two major shortcomings: the identity of the protein to be labeled must be known in advance of the experiment, and the DNA encoding the protein of interest must be genetically modified in order to enable labeling. Metabolic incorporation of amino acids that can be tagged or visualized in some way provides an approach to labeling newly synthesized proteins without genetically modifying the system in question and without knowing the identities of the proteins to be labeled in advance. A longstanding strategy for labeling and visualization of newly synthesized proteins in living cells and organisms has been the use of [^{35}S]methionine in conjunction with autoradiography (304). This method has enabled the study of a number of systems without genetic manipulations, but experiments involving [^{35}S]methionine must be performed with care due to the inherent dangers of working with radioactivity. Recently described approaches to residue-specifically incorporating reactive amino acids into proteins (for examples, see (46, 47, 305, 306)) have enabled alternative chemical approaches to visualizing newly synthesized proteins in cellular systems. In 2005, Beatty et al. described the chemical modification of proteins produced in *E. coli* containing amino acids **2** (Homopropargylglycine, Hpg) or **86** using a fluorogenic coumarin dye and copper-catalyzed azide-alkyne 1,3 dipolar cycloaddition (CuAAC) to generate fluorescently labeled proteins (307). The researchers found high, specific labeling of *E. coli* only after the cells had been incubated with ncAAs. Gel electrophoresis revealed that both an overexpressed recombinant protein and endogenously expressed *E. coli* proteins were labeled, suggesting that incorporation of ncAAs occurred in all newly synthesized proteins in the bacteria.

The concept of visualizing new protein synthesis proteome-wide using bioorthogonal chemistry has been extended to mammalian cells. Beatty, Liu, and

coworkers demonstrated that amino acid **2** could be used in mammalian cells as a chemical handle for labeling proteins expressed during a specified pulse time of the alkyne-containing amino acid (35). Aided by the methionine auxotrophy of mammalian cells, the technique was applied in several different cell types and studied by microscopy and flow cytometry. Fluorescence quantification by flow cytometry revealed that labeling was selective for cells that had been exposed to the alkyne-containing amino acid by approximately one order of magnitude and that the extent of labeling could be varied by introducing small amounts of methionine along with **2**. Visualization of the proteome has also been extended to labeling multiple populations of newly synthesized proteins using multiple pulses of ncAAs. In this work, proteins synthesized during specific time windows were distinguished by applying sequential pulses of **2** and **77** (Azidohomoalanine, Aha) to mammalian cells in culture. The researchers showed that after the completion of both pulses, proteins containing **2** could be labeled with an azide-containing fluorophore while proteins containing **77** could be labeled with an alkyne-containing fluorophore and visualized simultaneously within the same spatial area (308).

Monitoring the production of newly synthesized proteins within cells and organisms may shed light on a number of problems, including the localized synthesis of new protein populations. One area that has already been investigated using the visualization approaches described above is protein synthesis in neuronal dendrites, a poorly understood and somewhat controversial subject (309). Dieterich et al. designed fluorescent tags for visualization of azide or alkyne-containing newly synthesized neuronal proteins in situ (310). Their CuAAC tagging enabled detection of newly synthesized proteins after as little as 10 minutes of sample exposure to **2** or **77**, and sequential pulses enabled tagging of

multiple time-defined populations of newly synthesized proteins. Furthermore, using strain-promoted azide-alkyne ligation (249), the authors were able to study the diffusion of newly synthesized proteins by appending quantum dots to azide-containing neuronal proteins. The Flanagan laboratory has used ncAAs to aid their studies of the spatial regulation of protein synthesis in neuronal axons and dendrites. Use of **77** and an alkyne fluorophore helped confirm that the transmembrane receptor DCC, which regulates axon growth and guidance, colocalizes with newly synthesized proteins, revealing its previously undiscovered role in translation regulation (311).

Studies of protein *S*-acylation dynamics have also been aided by the visualization of newly synthesized proteins (312). Simultaneous monitoring of *S*-acylation and protein turnover in the protein H-Ras^{G12V} revealed a palmitate half-life of approximately fifty minutes on the protein. The authors note that these results are consistent with palmitate half-life values determined using radioactive compounds and suggest that their nonradioactive approach to monitoring turnover of this posttranslational modification could be applied generally to any cellular *S*-acylated protein. The successful application of ncAAs to study temporal and spatial aspects of protein synthesis and turnover in complex biological systems is a significant accomplishment. These promising results suggest that further employment of ncAAs to visualize the dynamics of the proteome will continue to yield information about where proteins are synthesized within cells and organisms and how these proteins are transported and degraded within these systems.

Future applications of proteome visualization may benefit greatly from recent methodological developments in chemical labeling methods and cell-selective incorporation of ncAAs into proteins. CuAAC ligations for the detection of proteins have

been quite successful in fixed cells, but their application to live cells may not be possible due to the toxic concentrations of copper needed to promote the chemistries involved. Strain-promoted azide-alkyne ligations provide a nontoxic alternative to labeling newly synthesized proteins, as the reaction of strained alkynes with azides does not require copper to promote the labeling chemistry (249, 313). Beatty et al. demonstrated the utility of such an approach by designing a set of membrane-permeant cyclooctynes functionalized with fluorophores and labeling of azide-tagged proteins in live mammalian cells (314). Multiple recent reports of CuAAC using short labeling times and newly developed ligand catalysts have demonstrated that glycans on the surface of live cells can be labeled without affecting cellular viability (315, 316). These results should also be applicable to the labeling of newly synthesized proteins appearing on the exteriors of live cells, but it is unclear whether these findings will enable labeling of proteins appearing in the interiors of cells without affecting cellular viability. The Lin laboratory has recently demonstrated an alternative chemistry for visualization of newly synthesized proteins (317). Use of the methionine analog homoallylglycine (**87**) and an ultraviolet light-promoted reaction with a fluorescent tag enables detection of proteins containing **87** within a precisely defined area. The method appears to be promising, although further application of the approach will likely require improved specificity, and the use of ultraviolet light may limit the method's use to fixed cells.

The use of translationally active methionine analogs in the visualization of newly synthesized proteins enables researchers to study all newly synthesized proteins within a system simultaneously. However, in nature, interactions between different types of cells within the same organism or between the cells of multiple organisms may result in very

different cellular responses in the components of the interacting systems, leading to the problem of determining which cells express particular proteins at a given time. Ngo et al. have recently demonstrated an approach to selectively visualize the newly synthesized proteins from one cell type in a mixed cell population (318). The key component of this method is the use of a mutant methionyl-tRNA synthetase (MetRS) capable of efficiently and selectively charging **4** (Azidonorleucine, Anl) onto $tRNA^{Met}$. When the mutant MetRS (termed NLL-MetRS) is expressed in a particular cell type, proteins expressed in the presence of **4** incorporate this azide-containing amino acid. In the absence of either the mutant synthetase or **4**, cells remain unlabeled. Thus, only proteins synthesized in cells expressing the mutant synthetase will be tagged with azides upon introduction of **4** into the medium. Ngo and coworkers illustrated this concept using a mixed population of *E. coli* cells expressing two different recombinant proteins. Only proteins expressed in the strain containing NLL-MetRS were labeled during CuAAC as determined by Western blotting and microscopy. This technique was also found to be applicable in a mixed population of *E. coli* and mouse alveolar macrophages. Figure 1.11 depicts the results of labeling experiments following the infection of a macrophage culture with various strains of *E. coli* cells. Only *E. coli* cells expressing the NLL-MetRS were labeled with a TAMRA-alkyne probe after fixing the coculture. Macrophages in the same culture were unlabeled, and positive and negative controls further proved that both cellular populations were synthesizing proteins during the course of the infection and that the labeling in the system was very specific. Hang and coworkers have demonstrated the use of cell-selective labeling in cocultures of *Salmonella typhimurium* and mammalian cells (52). Use of Anl or the long-chain alkyne analog 2-aminooctynoic acid (AOA, **88**) enabled selective labeling of *S.*

typhimurium cells with specificities similar to those observed by Ngo et al. These visualization experiments validate the idea of selectively labeling protein populations expressed in a cell type involved in interactions with other cells. The combined selectivity of ncAA incorporation and CuAAC-mediated fluorescence visualization ensures that a minority protein population can be examined in a vast background of other protein populations and should be applicable to a wide range of systems involving cell-cell interactions.

Protein identification. The same concept of metabolically incorporating chemically reactive ncAAs into newly synthesized proteins or more specific protein populations can also be extended to protein identification techniques. Tagging newly synthesized proteins with affinity purification reagents enables selective separation and enrichment of proteins that have been synthesized in a defined temporal or spatial window or within a particular cellular population. The resulting samples can then be analyzed using mass spectrometry identification techniques. Dietrich et al. first described this technique in 2006, naming the resulting approach bioorthogonal noncanonical amino acid tagging (BONCAT) (319, 320). In this work, the researchers demonstrated several important principles in the development of this technique for use in mammalian cell culture. First, the azide-containing noncanonical amino acid azidohomoalanine (**77**) did not perturb cultured mammalian cells significantly. The visible phenotypes of cells remained the same whether or not **77** was added to cultures, and the protein populations present in cells incubated with methionine or **77** were indistinguishable when examined with autoradiograms of one-dimensional gels. Second, as was found in fluorescence visualization studies, the tagging chemistry was very

specific for proteins that had been synthesized in the presence of **77**. Finally, proteins tagged with an alkyne-FLAG tag could be enriched from the much larger overall protein population with a streptavidin column, digested into smaller peptides on the column, and identified using tandem mass spectrometry. Experiments performed in HEK293 cells identified proteins from a wide range of gene ontological categories, showing that a highly diverse set of proteins can be isolated using BONCAT. Numerous improvements to enriching azide- or alkyne-containing proteins have been reported in support of the BONCAT method. All of these approaches have aimed to replace the on-column trypsinization step of the original BONCAT procedure with separate column purification and digestion steps. For example, Kramer et al. reported selective cleavage of newly synthesized proteins containing **77** at the azide side chains of the proteins (321). Other approaches have involved the synthesis of purification tags designed to enable the use of click chemistry to conjugate affinity reagents to azide- or alkyne-containing proteins, and enrich tagged proteins under mild conditions using appropriate columns. Nessen et al. demonstrated the use of strain-promoted click chemistry for the selective enrichment of newly synthesized proteins from *E. coli* (322). In this work, the use of **77** and a cleavable cyclooctyne resin enabled the enrichment and identification of newly synthesized proteins from whole-cell lysates. Szychowski et al. designed and synthesized five biotin-azide probes that can be cleaved over a wide range of conditions including reducing, mild acidic conditions, and ultraviolet irradiation (323). Using a GFP model system, the authors showed that an acid-cleavable tag enables highly selective conjugation with alkyne-containing proteins and leaves a small mass tag on labeled proteins after cleavage, an important consideration for proteomic studies. Hang and coworkers have developed several

tags that can be used for proteomic studies (52, 324). Application of these tags to labeling and mass spectrometric identification of newly synthesized proteins produced in *S. typhimurium* harboring MetRS-NLL in the presence of AOA (**88**) enabled identification of a large number of proteins from complex samples. Furthermore, samples from cocultures were enriched for proteins expressed in *S. typhimurium*, suggesting that the protein purification procedure successfully separated alkyne-containing proteins from unlabeled proteins.

BONCAT represents a potentially significant advance in the field of proteomics. The ability to selectively enrich a set of newly synthesized proteins from a larger proteomic sample may result in more sensitive detection method for protein subsets of interest. Furthermore, as demonstrated in the case of visualizing new protein synthesis, dynamic aspects of the proteome may be studied by varying the pulse time and duration of ncAAs or by pulsing sequentially with multiple ncAAs. The BONCAT technique may also be used to compare proteomes from two or more different cell samples by applying existing techniques such as SILAC (303) or by combining ncAA incorporation techniques with the use of ICAT (325) or iTRAQ (326) reagents. Finally, protein identification techniques appear to be compatible with the selective metabolic labeling strategy of Ngo et al. and Grammel et al. (52, 318). Proteomic studies with ncAAs take advantage of several aspects of ncAAs including global canonical amino acid replacement, bioorthogonal side chain chemistries, and varied aminoacyl-tRNA synthetase selectivities toward these noncanonical substrates. These several features enable the application of noncanonical amino acids in a broad range of proteomic applications, which should lead to new insights into a number of complex biological systems and processes.

Outlook

The future looks bright for applying noncanonical amino acids to problems in protein science and engineering. The question with ncAAs is no longer, "what kinds of noncanonical amino acids can be incorporated into proteins?" The range of chemical structures employed in experiments with ncAAs is quite impressive, ranging from single-atom changes of canonical residues to ncAAs containing functional groups rarely, if ever, seen in nature. Instead, the question is now, "what types of problems can best be solved by incorporating noncanonical amino acids into proteins?" Recent research has hinted at the varied applications possible with ncAAs. Researchers have employed noncanonical amino acids to investigate problems ranging from atomic-level protein structural questions to organism-wide responses to proteins containing noncanonical amino acids, with time scales of individual experiments ranging over several orders of magnitude. The broad range of experiments performed using noncanonical amino acids suggests that these residues should be thought of as possible tools for studying many problems, and not necessarily as essential tools for studying any particular problem or class of problem. Generally speaking, good candidate problems are ones in which ncAAs offer a route to a substantial increase in the quality of information and/or desired protein properties that can be obtained from a particular set of experiments without drastically raising the level of difficulty of performing these experiments. The several application areas discussed in this review have already started to benefit from the use of noncanonical amino acids according to the criteria outlined above. These benefits are perhaps most apparent in the areas of X-ray crystallography and membrane protein studies, in which a very large amount of data has

been acquired using relatively simple approaches that are not possible without the use of noncanonical amino acids.

A few more questions come to mind regarding research with proteins containing noncanonical amino acids. First, aside from the areas discussed above, what additional problems might benefit from the use of ncAAs in the near future? Biophysical characterizations appear to be among the most accessible problems, especially in light of the number of successful uses of ncAAs in X-ray crystallography and NMR studies. Fluorescence spectroscopy can benefit from both the direct incorporation of fluorescent amino acids and the chemical conjugation of fluorescent labels to reactive chemical functionalities in proteins, and proof-of-principle experiments of both of these approaches already exist in the literature (for examples, see (104, 327-329); see Merkel et al. for a review on the subject of intrinsically blue fluorescent amino acids as tools for protein science (330)). Vibrational spectroscopy of proteins performed after the incorporation of bonds with unique infrared or Raman signatures should continue to grow in importance for protein characterizations. The work of Ye et al. on the membrane protein rhodopsin (212, 213) is an early example of the effective exploitation of IR-active amino acids (discussed in the "Applications" section), and some work with IR-active amino acids in globular proteins also shows promise (331-333). Continued application of ncAAs to study the biological functions of proteins in living cells and organisms will also continue to grow in importance. Along with the visualization and identification of proteomic responses to biological stimuli, the precise control of protein function in biological settings is highly desirable. Proteins with activities or locations controlled by light have proven to be very powerful over the years for studying biological systems (334, 335), and some work suggests that using

ncAAs to achieve such photocontrolled systems may yield a number of new tools in this area (336-340). Given the large number of selective chemical conjugation strategies accessible with ncAAs (249, 251), protein therapeutics may also benefit from applications of ncAAs. Several simple strategies for modifying the pharmacochemical properties of proteins with ncAAs and selective chemistries have been reported in the scientific literature, including examples of protein PEGylation (313, 341-344) and viral surface modifications (345, 346). Commercial applications of these conjugation strategies may yield more drug-like bioconjugates for use as therapeutics. Furthermore, many more subtle changes to protein properties possible with ncAAs, such as those described in conjunction with protein stability, membrane proteins, and immune modulation, may facilitate further improvement of protein-based therapeutics.

What will be the role of in vitro protein synthesis in future work with noncanonical amino acids? The ribosomal production of genetically encoded polymers containing multiple ncAAs, ester linkages, *N*-methyl amino acids, and *N*-substituted glycine residues have all been reported in vitro (58-63). Cyclic peptides, which have many advantages over linear peptides as therapeutic entities, can be formed with the use of noncanonical amino acids in in vitro settings (347-349). These are intriguing molecules, but if the yields of such in vitro productions remain low, applications should be considered carefully and should provide a definite advantage over chemical peptide synthesis techniques. One proposed use of this production approach is as a platform for generating genetically encoded libraries of therapeutically relevant peptide drug candidates. These methodologies may complement or improve upon existing approaches to generating DNA-encoded chemical libraries (350). Genetically templated, highly unnatural molecules may also give researchers tools to

examine fundamental aspects of macromolecular folding and evolution. What kinds of three-dimensional structures can nonbiological polymers attain, how do they evolve, and what do these findings tell us about the functional and evolutionary properties of proteins?

Finally, will engineering additional components of the translational machinery make a large contribution to the ribosomal production of polymers containing noncanonical functionalities? Recent work on the ribosome (64, 65, 116, 117) and elongation factor-TU (351) raises the exciting possibility of expanding the translational capabilities of organisms or in vitro translation systems far beyond the capabilities of existing protein translation machinery. However, even the most promising recent experimental results suggest that engineering the translational machinery will require large numbers of incremental improvements to existing biological machinery. The limits to how far the ribosome and other translational components can be evolved away from their natural functions are still unclear, especially when this evolution is performing in living organisms. Performing evolution in vitro may be more appropriate for the creation of translational machinery with highly unnatural functions, but new in vitro protein expression systems and evolution techniques will likely have to be developed in order to enable this challenging type of translational apparatus engineering. Regardless of how the protein synthesis machinery is engineered, any resulting system should ideally be simple enough to use that it could be widely adopted by the scientific community and versatile enough to be employed for the generation of a wide variety of genetically encoded polymers.

In this review, we have highlighted the use of noncanonical amino acids as tools to study scientific problems. With careful experimental design, ncAAs can greatly augment scientists' abilities to address an extremely broad range of questions; the ncAA toolkit is

very large and very effective. Research with noncanonical amino acids is poised to move beyond the methodological development efforts of a select few researchers to the collaborative exploitation of these methods across numerous scientific disciplines.

Acknowledgements

Work at Caltech on noncanonical amino acids is supported by NIH grant GM62523 and by the Institute for Collaborative Biotechnologies under contract W911NF-09-D-0001 with the Army Research Office.

References

1. Tracewell CA & Arnold FH (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* **13,** 3-9.
2. Walsh CT, Garneau-Tsodikova S, & Gatto GJ (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44,** 7342-7372.
3. Chan WC & White PD (2000) *Fmoc Solid Phase Peptide Synthesis: A Practical Approach* (Oxford University Press, New York).
4. Benoiten NL (2005) *Chemistry of Peptide Synthesis* (CRC, Boca Raton, FL, USA) 1st Ed.
5. Dawson PE & Kent SBH (2000) Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **69,** 923-960.
6. Imperiali B & Taylor EV (2009) Native Chemical Ligation: Semisynthesis of Post-translationally Modified Proteins and Biological Probes. *Protein Eng.,* Nucleic

Acids and Molecular Biology, eds Kohrer C & RajBhandary UL (Springer-Verlag, Berlin), pp 65-96.

7. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72,** 249-289.

8. Muir TW (2008) Studying protein structure and function using semisynthesis. *Biopolymers* **90,** 743-750.

9. Flavell RR & Muir TW (2009) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **42,** 107-116.

10. Cherfas J (1982) *Man-Made Life: An Overview of the Science, Technology, and Commerce of Genetic Engineering* (Pantheon Books, New York) First American Ed.

11. Brudno Y & Liu DR (2009) Recent progress toward the templated synthesis and directed evolution of sequence-defined synthetic polymers. *Chem. Biol.* **16,** 265-276.

12. Chapeville F*, et al.* (1962) On role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **48,** 1086-1092.

13. Crick FHC (1958) On protein synthesis. *Symp. Soc. Exp. Biol.* **12,** 138-163.

14. Johnson AE, Woodward WR, Herbert E, & Menninger JR (1976) N-epsilon acetyllysine transfer ribonucleic acid - biologically active analog of aminoacyl transfer ribonucleic acids. *Biochemistry* **15,** 569-575.

15. Hecht SM, Alford BL, Kuroda Y, & Kitano S (1978) Chemical aminoacylation of transfer-RNAs. *J. Biol. Chem.* **253,** 4517-4520.

16. Heckler TG*, et al.* (1984) T4 RNA ligase mediated preparation of novel chemically misacylated phenylalanine transfer-RNAs. *Biochemistry* **23,** 1468-1473.

17. Baldini G, Martoglio B, Schachenmann A, Zugliani C, & Brunner J (1988) Mischarging escherichia coli phenylalanine transfer RNA with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analog of phenylalanine. *Biochemistry* **27,** 7951-7959.

18. Happ E, Scalfihapp C, & Chladek S (1987) Aminoacyl derivatives of nucleosides nucleotides and polynucleotides 43. New approach to the synthesis of 2'(3')-O-aminoacyl oligoribonucleotides. *J. Org. Chem.* **52,** 5387-5391.

19. Berg P, Bergmann FH, Ofengand EJ, & Dieckmann M (1961) Enzymic synthesis of amino acyl derivatives of ribonucleic acid. *J. Biol. Chem.* **236,** 1726-1734.

20. Berg P & Ofengand EJ (1958) An enzymatic mechanism for linking amino acids to RNA. *Proc. Natl. Acad. Sci. U. S. A.* **44,** 78-86.

21. Ibba M, Francklyn CS, & Cusack S (2005) *The Aminoacyl-tRNA Synthetases* (Landes Bioscience, Austin, TX).

22. Budisa N (2006) *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins* (Wiley-VCH, Weinheim, Germany).

23. Doctor BP & Mudd JA (1963) Species specificity of amino acid acceptor ribonucleic acid and aminoacyl soluble ribonucleic acid synthetases. *J. Biol. Chem.* **238,** 3677-3681.

24. Goodman HM, Abelson J, Landy A, Brenner S, & Smith JD (1968) Amber suppression - a nucleotide change in an anticodon of a tyrosine transfer RNA. *Nature* **217,** 1019-1024.

25. Ibba M & Soll D (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69,** 617-650.

26. Capecchi MR & Gussin GN (1965) Suppression in vitro - identification of a serine-sRNA as a nonsense suppressor. *Science* **149,** 417-422.

27. Engelhar Dl, Webster RE, Wilhelm RC, & Zinder ND (1965) In vitro studies on mechanism of suppression of a nonsense suppressor mutation. *Proc. Natl. Acad. Sci. U. S. A.* **54,** 1791-1797.

28. Smith JD, Abelson JN, Clark BFC, Goodman HM, & Brenner S (1966) Studies on amber suppressor tRNA. *Cold Spring Harbor Symp. Quant. Biol.* **31,** 479-485.

29. Richmond MH (1962) Effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **26,** 398-420.

30. Cowie DB & Cohen GN (1957) Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* **26,** 252-261.

31. Hortin G & Boime I (1983) Applications of amino acid analogs for studying co-translational and posttranslational modifications of proteins. *Methods Enzymol.* **96,** 777-784.

32. Wilson MJ & Hatfield DL (1984) Incorporation of modified amino acids into proteins in vivo. *Biochim. Biophys. Acta* **781,** 205-215.

33. Hendrickson WA, Horton JR, & Lemaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD) - a vehicle for direct determination of 3-dimensional structure. *EMBO J.* **9,** 1665-1672.

34. Link AJ & Tirrell DA (2005) Reassignment of sense codons in vivo. *Methods* **36,** 291-298.

35. Beatty KE*, et al.* (2006) Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Ed.* **45,** 7364-7367.

36. Wiltschi B, Wenger W, Nehring S, & Budisa N (2008) Expanding the genetic code of Saccharomyces cerevisiae with methionine analogues. *Yeast* **25,** 775-786.

37. Budisa N, Wenger W, & Wiltschi B (2010) Residue-specific global fluorination of Candida antarctica lipase B in Pichia pastoris. *Mol. Biosyst.* **6,** 1630-1639.

38. Lepthien S, Merkel L, & Budisa N (2010) In vivo double and triple labeling of proteins using synthetic amino acids. *Angew. Chem. Int. Ed.* **49,** 5446-5450.

39. Merkel L, Schauer M, Antranikian G, & Budisa N (2010) Parallel incorporation of different fluorinated amino acids: on the way to "teflon" proteins. *ChemBioChem* **11,** 1505-1507.

40. Kiick KL, Weberskirch R, & Tirrell DA (2001) Identification of an expanded set of translationally active methionine analogues in Escherichia coli. *FEBS Lett.* **502,** 25-30.

41. 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.

42. Wang P, Vaidehi N, Tirrell DA, & Goddard WA (2002) Virtual screening for binding of phenylalanine analogues to phenylalanyl-tRNA synthetase. *J. Am. Chem. Soc.* **124,** 14442-14449.

43. Kast P & Hennecke H (1991) Amino acid substrate specificity of Escherichia coli phenylalanyl-transfer RNA-synthetase altered by distinct mutations. *J. Mol. Biol.* **222,** 99-124.

44. 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.

45. 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.

46. 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.

47. Tanrikulu IC, Schmitt E, Mechulam Y, Goddard WA, & Tirrell DA (2009) Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15285-15290.

48. Mursinna RS, Lee KW, Briggs JM, & Martinis SA (2004) Molecular dissection of a critical specificity determinant within the amino acid editing domain of leucyl-tRNA synthetase. *Biochemistry* **43,** 155-165.

49. Mursinna RS & Martinis SA (2002) Rational design to block amino acid editing of a tRNA synthetase. *J. Am. Chem. Soc.* **124,** 7286-7287.

50. Tang Y & Tirrell DA (2002) Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **41,** 10635-10645.

51. Tang Y, Wang P, Van Deventer JA, Link AJ, & Tirrell DA (2009) Introduction of an aliphatic ketone into recombinant proteins in a bacterial strain that overexpresses an editing-impaired leucyl-tRNA synthetase. *ChemBioChem* **10,** 2188-2190.

52. Grammel M, Zhang MZM, & Hang HC (2010) Orthogonal alkynyl amino acid reporter for selective labeling of bacterial proteomes during infection. *Angew. Chem. Int. Ed.* **49,** 5970-5974.

53. Abdeljabbar DM, Klein TJ, Zhang SY, & Link AJ (2009) A single genomic copy of an engineered methionyl-tRNA synthetase enables robust incorporation of azidonorleucine into recombinant proteins in E-coli. *J. Am. Chem. Soc.* **131,** 17078-17079.

54. Forster AC*, et al.* (2003) Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6353-6357.

55. Josephson K, Hartman MCT, & Szostak JW (2005) Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* **127,** 11727-11735.

56. Ohta A, Yamagishi Y, & Suga H (2008) Synthesis of biopolymers using genetic code reprogramming. *Curr. Opin. Chem. Biol.* **12,** 159-167.

57. Ohuchi M, Murakami H, & Suga H (2007) The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr. Opin. Chem. Biol.* **11,** 537-542.

58. Subtelny AO, Hartman MCT, & Szostak JW (2008) Ribosomal synthesis of n-methyl peptides. *J. Am. Chem. Soc.* **130,** 6131-6136.

59. Tan ZP, Forster AC, Blacklow SC, & Cornish VW (2004) Amino acid backbone specificity of the Escherichia coli translation machinery. *J. Am. Chem. Soc.* **126,** 12752-12753.

60. Zhang BL*, et al.* (2007) Specificity of translation for N-alkyl amino acids. *J. Am. Chem. Soc.* **129,** 11316-11317.

61. Kawakami T, Murakami H, & Suga H (2008) Ribosomal synthesis of polypeptoids and peptoid-peptide hybrids. *J. Am. Chem. Soc.* **130,** 16861-16863.

62. Kawakami T, Murakami H, & Suga H (2008) Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* **15,** 32-42.

63. Ohta A, Murakami H, Higashimura E, & Suga H (2007) Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14,** 1315-1322.

64. Dedkova LM, Fahmi NE, Golovine SY, & Hecht SM (2003) Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.* **125,** 6616-6617.

65. Dedkova LM, Fahmi NE, Golovine SY, & Hecht SM (2006) Construction of modified ribosomes for incorporation of D-amino acids into proteins. *Biochemistry* **45,** 15541-15551.

66. Brudno Y, Birnbaum ME, Kleiner RE, & Liu DR (2010) An in vitro translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Biol.* **6,** 148-155.

67. Bain JD, Glabe CG, Dix TA, Chamberlin AR, & Diala ES (1989) Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **111,** 8013-8014.

68. Noren CJ, Anthonycahill SJ, Griffith MC, & Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244,** 182-188.

69. Drabkin HJ, Park HJ, & RajBhandary UL (1996) Amber suppression in mammalian cells dependent upon expression of an Escherichia coli aminoacyl-tRNA synthetase gene. *Mol. Cell. Biol.* **16,** 907-913.

70. Furter R (1998) Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Sci.* **7,** 419-426.

71. Liu DR, Magliery TJ, Pasternak M, & Schultz PG (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 10092-10097.

72. Liu DR & Schultz PG (1999) Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 4780-4785.

73. Drabkin HJ, Estrella M, & Rajbhandary UL (1998) Initiator-elongator discrimination in vertebrate tRNAs for protein synthesis. *Mol. Cell. Biol.* **18,** 1459-1466.

74. Kowal AK, Kohrer C, & RajBhandary UL (2001) Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2268-2273.

75. Wang L & Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8,** 883-890.

76. Wang L, Brock A, Herberich B, & Schultz PG (2001) Expanding the genetic code of Escherichia coli. *Science* **292,** 498-500.

77. Sakamoto K*, et al.* (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30,** 4692-4699.

78. Chin JW*, et al.* (2003) An expanded eukaryotic genetic code. *Science* **301,** 964-967.

79. Liu WS, Brock A, Chen S, Chen SB, & Schultz PG (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4,** 239-244.

80. Wang WY*, et al.* (2007) Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* **10,** 1063-1072.

81. Cornish VW, Mendel D, & Schultz PG (1995) Probeing protein structure and function with an expanded genetic code. *Angew. Chem. Int. Ed.* **34,** 621-633.

82. Wang L & Schultz PG (2005) Expanding the genetic code. *Angew. Chem. Int. Ed.* **44,** 34-66.

83. Wang L, Xie J, & Schultz PG (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 225-249.

84. Wang Q, Parrish AR, & Wang L (2009) Expanding the genetic code for biological studies. *Chem. Biol.* **16,** 323-336.

85. 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.

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

87. Young TS & Schultz PG (2010) Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J. Biol. Chem.* **285,** 11039-11044.

88. Ryu YH & Schultz PG (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat. Methods* **3,** 263-265.

89. Young TS, Ahmad I, Yin JA, & Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361-374.

90. Guo JT, Melancon CE, Lee HS, Groff D, & Schultz PG (2009) Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed.* **48,** 9148-9151.

91. Cellitti SE*, et al.* (2008) In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **130,** 9268-9281.

92. Hammill JT, Miyake-Stoner S, Hazen JL, Jackson JC, & Mehl RA (2007) Preparation of site-specifically labeled fluorinated proteins for F-19-NMR structural characterization. *Nat. Protoc.* **2,** 2601-2607.

93. Jackson JC, Hammill JT, & Mehl RA (2007) Site-specific incorporation of a F-19-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **129,** 1160-1166.

94. Chen S, Schultz PG, & Brock A (2007) An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in Saccharomyces cerevisiae. *J. Mol. Biol.* **371,** 112-122.

95. Wang Q & Wang L (2008) New methods enabling efficient incorporation of unnatural amino acids in yeast. *J. Am. Chem. Soc.* **130,** 6066-6067.

96. Young TS, Ahmad I, Brock A, & Schultz PG (2009) Expanding the genetic repertoire of the methylotrophic yeast Pichia pastoris. *Biochemistry* **48,** 2643-2653.

97. Wang F, Robbins S, Guo JT, Shen WJ, & Schultz PG (2010) Genetic incorporation of unnatural amino acids into proteins in Mycobacterium tuberculosis. *PLoS ONE* **5,** e9354.

98. Hino N*, et al.* (2005) Protein photo-crosslinking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* **2,** 201-206.

99. Zhang ZW*, et al.* (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 8882-8887.

100. Neumann H, Peak-Chew SY, & Chin JW (2008) Genetically encoding N-epsilon-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4,** 232-234.

101. Nguyen DP*, et al.* (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* **131,** 8720-8721.

102. Fekner T, Li X, Lee MM, & Chan MK (2009) A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed.* **48,** 1633-1635.

103. Chen PR*, et al.* (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed.* **48,** 4052-4055.

104. Summerer D*, et al.* (2006) A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9785-9789.

105. Oki K, Sakamoto K, Kobayashi T, Sasaki HM, & Yokoyama S (2008) Transplantation of a tyrosine editing domain into a tyrosyl-tRNA synthetase variant enhances its specificity for a tyrosine analog. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 13298-13303.

106. Ai HW, Shen WJ, Brustad E, & Schultz PG (2010) Genetically encoded alkenes in yeast. *Angew. Chem. Int. Ed.* **49,** 935-937.

107. Neumann H, Slusarczyk AL, & Chin JW (2010) De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc.* **132,** 2142-2144.

108. Hohsaka T, Ashizuka Y, Taira H, Murakami H, & Sisido M (2001) Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* **40,** 11060-11064.

109. Rodriguez EA, Lester HA, & Dougherty DA (2006) In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8650-8655.

110. Anderson JC*, et al.* (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7566-7571.

111. Taki M, Matsushita J, & Sisido M (2006) Expanding the genetic code in a mammalian cell line by the introduction of four-base codon/anticodon pairs. *ChemBioChem* **7,** 425-428.

112. Hohsaka T, Ashizuka Y, Sasaki H, Murakami H, & Sisido M (1999) Incorporation of two different nonnatural amino acids independently into a single protein through extension of the genetic code. *J. Am. Chem. Soc.* **121,** 12194-12195.

113. Huang Y*, et al.* (2010) A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli. *Mol. Biosyst.* **6,** 683-686.

114. Wan W*, et al.* (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chem. Int. Ed.* **49,** 3211-3214.

115. Rackham O & Chin JW (2005) A network of orthogonal ribosome-mRNA pairs. *Nat. Chem. Biol.* **1,** 159-166.

116. Neumann H, Wang KH, Davis L, Garcia-Alai M, & Chin JW (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464,** 441-444.

117. Wang KH, Neumann H, Peak-Chew SY, & Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25,** 770-777.

118. Tirrell DA (2009) Reinterpreting the Genetic Code: Implications for Macromolecular Design, Evolution and Analysis. *Physical Biology: From Atoms to Medicine*, ed Zewail AH (Imperial College Press, London), pp 165-188.

119. Connor RE & Tirrell DA (2007) Non-canonical amino acids in protein polymer design. *Polymer Reviews* **47,** 9-28.

120. Beatty KE & Tirrell DA (2009) Noncanonical amino acids in protein science and engineering. *Protein Eng.,* Nucleic Acids and Molecular Biology, eds Kohrer C & RajBhandary UL (Springer, Berlin), pp 127-154.

121. Wu X & Schultz PG (2009) Synthesis at the interface of chemistry and biology. *J. Am. Chem. Soc.* **131,** 12497-12515.

122. McPherson A (2009) *Introduction to Macromolecular Crystallography* (Wiley-Blackwell, Hoboken, NJ) 2nd Ed.

123. Kasai N & Kakudo M (2005) *X-Ray Diffraction By Macromolecules* (Springer-Verlag, Berlin).

124. Graves BJ*, et al.* (1990) Structure of interleukin 1-alpha at 2.7-A resolution. *Biochemistry* **29,** 2679-2684.

125. Yang W, Hendrickson WA, Crouch RJ, & Satow Y (1990) Structure of ribonuclease-H phased at 2-A resolution by MAD analysis of the selenomethionyl protein. *Science* **249,** 1398-1405.

126. Wu H, Lustbader JW, Liu Y, Canfield RE, & Hendrickson WA (1994) Structure of human chorionic-gonadotropin at 2.6-angstrom resolution from MAD analysis of the selenomethionyl protein. *Structure* **2,** 545-558.

127. Lustbader JW*, et al.* (1995) The expression, characterization, and crystallization of wild-type and selenomethionyl human chorionic-gonadotropin. *Endocrinology* **136,** 640-650.

128. Bellizzi JJ, Widom J, Kemp CW, & Clardy J (1999) Producing selenomethionine-labeled proteins with a baculovirus expression vector system. *Structure Fold. Des.* **7,** R263-R267.

129. McWhirter SM*, et al.* (1999) Crystallographic analysis of CD40 recognition and signaling by human TRAF2. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8408-8413.

130. Rice LM, Earnest TN, & Brunger AT (2000) Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **56,** 1413-1420.

131. Walsh MA, Evans G, Sanishvili R, Dementieva I, & Joachimiak A (1999) MAD data collection - current trends. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **55,** 1726-1732.

132. Hendrickson WA (1999) Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron Rad.* **6,** 845-851.

133. Hendrickson WA & Ogata CM (1997) Phase determination from multiwavelength anomalous diffraction measurements. *Macromolecular Crystallography, Pt A,* Methods in Enzymology, ed Carter CW (Elsevier), Vol 276, pp 494-523.

134. Boles JO*, et al.* (1995) Expression, characterization and crystallographic analysis of telluromethionyl dihydrofolate reductase. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **51,** 731-739.

135. Boles JO*, et al.* (1994) Bio-incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nat. Struct. Biol.* **1,** 283-284.

136. Budisa N*, et al.* (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur. J. Biochem.* **230,** 788-796.

137. Sanchez J-F, Hoh F, Strub M-P, Aumelas A, & Dumas C (2002) Structure of the Cathelicidin Motif of Protegrin-3 Precursor: Structural Insights into the Activation Mechanism of an Antimicrobial Protein. *Structure* **10,** 1363-1370.

138. Strub MP*, et al.* (2003) Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* **11,** 1359-1367.

139. Bae JH*, et al.* (2001) Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **309,** 925-936.

140. Boles JO, Henderson J, Hatch D, & Silks LA (2002) Synthesis and incorporation of [6,7]-selenatryptophan into dihydrofolate reductase. *Biochem. Biophys. Res. Commun.* **298,** 257-261.

141. Xie JM*, et al.* (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22,** 1297-1301.

142. Sakamoto K*, et al.* (2009) Genetic encoding of 3-iodo-L-tyrosine in Escherichia coli for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* **17,** 335-344.

143. Hibler DW*, et al.* (1989) [4] Isotopic labeling with hydrogen-2 and carbon-13 to compare conformations of proteins and mutants generated by site-directed mutagenesis, I. *Methods Enzymol.*, (Academic Press), Vol 177, pp 74-86.

144. Muchmore DC*, et al.* (1989) [3] Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.*, (Academic Press), Vol 177, pp 44-73.

145. Tugarinov V, Hwang PM, & Kay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu. Rev. Biochem.* **73,** 107-146.

146. Danielson MA & Falke JJ (1996) Use of F-19 NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25,** 163-195.

147. Frieden C, Hoeltzli SD, & Bann JG (2004) The preparation of F-19-labeled proteins for NMR studies. *Energetics of Biological Macromolecules, Pt E,* Methods in Enzymology, eds Holt JM, Johnson ML, & Ackers GK), Vol 380, pp 400-415.

148. Bann JG, Pinkner J, Hultgren SJ, & Frieden C (2002) Real-time and equilibrium F-19-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 709-714.

149. Cleve P, Robinson V, Duewel HS, & Honek JF (1999) Difluoromethionine as a Novel 19F NMR Structural Probe for Internal Amino Acid Packing in Proteins. *J. Am. Chem. Soc.* **121,** 8475-8478.

150. Eichler JF, Cramer JC, Kirk KL, & Bann JG (2005) Biosynthetic incorporation of fluorohistidine into proteins in E-coli: A new probe of macromolecular structure. *ChemBioChem* **6,** 2170-2173.

151. Rule GS, Pratt EA, Simplaceanu V, & Ho C (1987) Nuclear magnetic resonance and molecular genetic studies of the membrane-bound D-lactate dehydrogenase of Escherichia coli. *Biochemistry* **26,** 549-556.

152. Peersen OB, Pratt EA, Truong HTN, Ho C, & Rule GS (1990) Site-specific incorporation of 5-fluorotryptophan as a probe of the structure and function of the membrane-bound D-lactate dehydrogenase of Escherichia coli - a F-19 nuclear magnetic resonance study. *Biochemistry* **29,** 3256-3262.

153. Sun ZY*, et al.* (1993) A F-19-NMR study of the membrane-binding region of D-lactate dehydrogenase of Escherichia coli. *Protein Sci.* **2,** 1938-1947.

154. Ropson IJ & Frieden C (1992) Dynamic NMR spectral analysis and protein folding - identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by F-19 NMR. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7222-7226.

155. Sun ZY, Pratt EA, Simplaceanu V, & Ho C (1996) A F-19-NMR study of the equilibrium unfolding of membrane-associated d-lactate dehydrogenase of Escherichia coli. *Biochemistry* **35,** 16502-16509.

156. Hoeltzli SD & Frieden C (1994) F-19 NMR spectroscopy of [6-F-19]tryptophan-labeled Escherichia coli dihydrofolate reductase - equilibrium folding and ligand-binding studies. *Biochemistry* **33,** 5502-5509.

157. Shu Q & Frieden C (2004) Urea-dependent unfolding of murine adenosine deaminase: sequential destabilization as measured by F-19 NMR. *Biochemistry* **43,** 1432-1439.

158. Shu Q & Frieden C (2005) Relation of enzyme activity to local/global stability of murine adenosine deaminase: F-19 NMR studies. *J. Mol. Biol.* **345,** 599-610.

159. Li H & Frieden C (2005) Phenylalanine side chain behavior of the intestinal fatty acid-binding protein - The effect of urea on backbone and side chain stability. *J. Biol. Chem.* **280,** 38556-38561.

160. Li H & Frieden C (2006) Fluorine-19 NMR studies on the acid state of the intestinal fatty acid binding protein. *Biochemistry* **45,** 6272-6278.

161. Khan F, Kuprov I, Craggs TD, Hore PJ, & Jackson SE (2006) 19F NMR studies of the native and denatured states of green fluorescent protein. *J. Am. Chem. Soc.* **128,** 10729-10737.

162. Li H & Frieden C (2007) Comparison of C40/82A and P27A C40/82A barstar mutants using F-19 NMR. *Biochemistry* **46,** 4337-4347.

163. Hoeltzli SD & Frieden C (1995) Stopped-flow NMR spectroscopy - real-time unfolding studies of 6-F-19-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 9318-9322.

164. Minks C, Huber R, Moroder L, & Budisa N (1999) Atomic Mutations at the Single Tryptophan Residue of Human Recombinant Annexin V: Effects on Structure, Stability, and Activity. *Biochemistry* **38,** 10649-10659.

165. Bann JG & Frieden C (2004) Folding and domain-domain interactions of the chaperone PapD measured by F-19 NMR. *Biochemistry* **43,** 13775-13786.

166. Li CG*, et al.* (2010) Protein F-19 NMR in Escherichia coli. *J. Am. Chem. Soc.* **132,** 321-327.

167. Jones DH*, et al.* (2010) Site-specific labeling of proteins with NMR-active unnatural amino acids. *J. Biomol. NMR* **46,** 89-100.

168. Lampe JN*, et al.* (2008) Ligand-induced conformational heterogeneity of cytochrome P450 CYP119 identified by 2D NMR spectroscopy with the unnatural amino acid C-13-p-methoxyphenylalanine. *J. Am. Chem. Soc.* **130,** 16168-16169.

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

170. Ellman JA, Mendel D, & Schultz PG (1992) Site-specific incorporation of novel backbone structures into proteins. *Science* **255,** 197-200.

171. Cornish VW, Kaplan MI, Veenstra DL, Kollman PA, & Schultz PG (1994) Stabilizing and destabilizing effects of placing beta-branched amino acids in protein alpha-helices. *Biochemistry* **33,** 12022-12031.

172. Thorson JS, Chapman E, & Schultz PG (1995) Analysis of hydrogen bonding strengths in proteins using unnatural amino acids. *J. Am. Chem. Soc.* **117,** 9361-9362.

173. Shin IJ, Ting AY, & Schultz PG (1997) Analysis of backbone hydrogen bonding in a beta-turn of staphylococcal nuclease. *J. Am. Chem. Soc.* **119,** 12667-12668.

174. Chapman E, Thorson JS, & Schultz PG (1997) Mutational analysis of backbone hydrogen bonds in Staphylococcal nuclease. *J. Am. Chem. Soc.* **119,** 7151-7152.

175. Koh JT, Cornish VW, & Schultz PG (1997) An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **36,** 11314-11322.

176. Thorson JS, Chapman E, Murphy EC, Schultz PG, & Judice JK (1995) Linear free energy analysis of hydrogen bonding in proteins. *J. Am. Chem. Soc.* **117,** 1157-1158.

177. Guo JT, Wang JY, Anderson JC, & Schultz PG (2008) Addition of an alpha-hydroxy acid to the genetic code of bacteria. *Angew. Chem. Int. Ed.* **47,** 722-725.

178. Dunitz JD (2004) Organic fluorine: Odd man out. *ChemBioChem* **5,** 614-621.

179. Marsh ENG (2000) Towards the nonstick egg: designing fluorous proteins. *Chem. Biol.* **7,** R153-R157.

180. Yoder NC & Kumar K (2002) Fluorinated amino acids in protein design and engineering. *Chem. Soc. Rev.* **31,** 335-341.

181. Yoder NC, Yuksel D, Dafik L, & Kumar K (2006) Bioorthogonal noncovalent chemistry: fluorous phases in chemical biology. *Curr. Opin. Chem. Biol.* **10,** 576-583.

182. Tang Y*, et al.* (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. *Angew. Chem. Int. Ed.* **40,** 1494-1496.

183. Son S, Tanrikulu IC, & Tirrell DA (2006) Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues. *ChemBioChem* **7,** 1251-1257.

184. Montclare JK, Son S, Clark GA, Kumar K, & Tirrell DA (2009) Biosynthesis and stability of coiled-coil peptides containing (2S,4R)-5,5,5-trifluoroleucine and (2S,4S)-5,5,5-trifluoroleucine. *ChemBioChem* **10,** 84-86.

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

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

187. Cirino PC, Tang Y, Takahashi K, Tirrell DA, & Arnold FH (2003) Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnol. Bioeng.* **83,** 729-734.

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

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

190. Wang P, Tang Y, & Tirrell DA (2003) Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc.* **125,** 6900-6906.

191. Moroder L & Budisa N (2010) Synthetic biology of protein folding. *Chemphyschem* **11,** 1181-1188.

192. Steiner T*, et al.* (2008) Synthetic biology of proteins: tuning GFPs folding and stability with fluoroproline. *PLoS ONE* **3,** e1680.

193. 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.

194. 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.

195. Levine RL, Moskovitz J, & Stadtman ER (2000) Oxidation of methionine in proteins: Roles in antioxidant defense and cellular regulation. *Iubmb Life* **50,** 301-307.

196. Stadtman ER, Moskovitz J, Berlett BS, & Levine RL (2002) Cyclic oxidation and reduction of protein methionine residues is an important antioxidant mechanism. *Mol. Cell. Biochem.* **234,** 3-9.

197. Hoshi T & Heinemann SH (2001) Regulation of cell function by methionine oxidation and reduction. *J. Physiol. (Lond).* **531,** 1-11.

198. Elofsson A & von Heijne G (2007) Membrane protein structure: Prediction versus reality. *Annu. Rev. Biochem.* **76,** 125-140.

199. Hopkins AL & Groom CR (2002) The druggable genome. *Nat. Rev. Drug Discov.* **1,** 727-730.

200. Klabunde T & Hessler G (2002) Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **3,** 929-944.

201. Lagerstrom MC & Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7,** 339-357.

202. Wagner S, Bader ML, Drew D, & de Gier JW (2006) Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24,** 364-371.

203. Link AJ & Georgiou G (2007) Advances and challenges in membrane protein expression. *AICHE J.* **53,** 752-756.

204. Johansson LC, Wohri AB, Katona G, Engstrom S, & Neutze R (2009) Membrane protein crystallization from lipidic phases. *Curr. Opin. Struct. Biol.* **19,** 372-378.

205. Neher E & Sakmann B (1976) Single-channel currents recorded from membrane of denervated frog muscle fibers. *Nature* **260,** 799-802.

206. 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.

207. Gallivan JP, Lester HA, & Dougherty DA (1997) Site-specific incorporation of biotinylated amino acids to identify surface-exposed residues in integral membrane proteins. *Chem. Biol.* **4,** 739-749.

208. England PM, Lester HA, Davidson N, & Dougherty DA (1997) Site-specific, photochemical proteolysis applied to ion channels in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11025-11030.

209. England PM, Lester HA, & Dougherty DA (1999) Mapping disulfide connectivity using backbone ester hydrolysis. *Biochemistry* **38,** 14409-14415.

210. Turcatti G*, et al.* (1997) Fluorescent labeling of NK2 receptor at specific sites in vivo and fluorescence energy transfer analysis of NK2 ligand-receptor complexes. *Receptors Channels*, (Harwood Acad Publ Gmbh), pp 201-207.

211. Turcatti G*, et al.* (1996) Probing the structure and function of the tachykinin neurokinin-2 receptor through biosynthetic incorporation of fluorescent amino acids at specific sites. *J. Biol. Chem.* **271,** 19991-19998.

212. Ye SX, Huber T, Vogel R, & Sakmar TP (2009) FTIR analysis of GPCR activation using azido probes. *Nat. Chem. Biol.* **5,** 397-399.

213. Ye SX*, et al.* (2010) Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464,** 1386-1389.

214. Tao X, Lee A, Limapichat W, Dougherty DA, & MacKinnon R (2010) A gating charge transfer center in voltage sensors. *Science* **328,** 67-73.

215. Nowak MW*, et al.* (1995) Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science* **268,** 439-442.

216. Dougherty DA (2008) Physical organic chemistry on the brain. *J. Org. Chem.* **73,** 3667-3673.

217. Hanek AP, Lester HA, & Dougherty DA (2008) A stereochemical test of a proposed structural feature of the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **130,** 13216-13218.

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

219. Beene DL, Price KL, Lester HA, Dougherty DA, & Lummis SCR (2004) Tyrosine residues that control binding and gating in the 5-hydroxytryptamine(3) receptor revealed by unnatural amino acid mutagenesis. *J. Neurosci.* **24,** 9097-9104.

220. Price KL*, et al.* (2008) A hydrogen bond in loop a is critical for the binding and function of the 5-HT3 receptor. *Biochemistry* **47,** 6370-6377.

221. Blum AP, Lester HA, & Dougherty DA (2010) Nicotinic pharmacophore: The pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13206-13211.

222. England PM, Zhang YN, Dougherty DA, & Lester HA (1999) Backbone mutations in transmembrane domains of a ligand-gated ion channel: Implications for the mechanism of gating. *Cell* **96,** 89-98.

223. Dang H, England PM, Farivar SS, Dougherty DA, & Lester HA (2000) Probing the role of a conserved M1 proline residue in 5-hydroxytryptamine(3) receptor gating. *Mol. Pharmacol.* **57,** 1114-1122.

224. Gleitsman KR, Kedrowski SMA, Lester HA, & Dougherty DA (2008) An intersubunit hydrogen bond in the nicotinic acetylcholine receptor that contributes to channel gating. *J. Biol. Chem.* **283,** 35638-35643.

225. Cashin AL, Torrice MM, McMenimen KA, Lester HA, & Dougherty DA (2007) Chemical-scale studies on the role of a conserved aspartate in preorganizing the agonist binding site of the nicotinic acetylcholine receptor. *Biochemistry* **46,** 630-639.

226. Gleitsman KR, Lester HA, & Dougherty DA (2009) Probing the role of backbone hydrogen bonding in a critical beta sheet of the extracellular domain of a Cys-loop receptor. *ChemBioChem* **10,** 1385-1391.

227. Li LT*, et al.* (2001) The tethered agonist approach to mapping ion channel proteins toward a structural model for the agonist binding site of the nicotinic acetylcholine receptor. *Chem. Biol.* **8,** 47-58.

228. Petersson EJ, Choi A, Dahan DS, Lester HA, & Dougherty DA (2002) A perturbed pK(a) at the binding site of the nicotinic acetylcholine receptor: Implications for nicotine binding. *J. Am. Chem. Soc.* **124,** 12662-12663.

229. Dougherty DA (1996) Cation-pi interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **271,** 163-168.

230. Ma JC & Dougherty DA (1997) The cation-pi interaction. *Chem. Rev.* **97,** 1303-1324.

231. Zhong WG*, et al.* (1998) From ab initio quantum mechanics to molecular neurobiology: A cation-pi binding site in the nicotinic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 12088-12093.

232. Beene DL*, et al.* (2002) Cation-pi interactions in ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors: The anomalous binding properties of nicotine. *Biochemistry* **41,** 10262-10269.

233. Cashin AL, Petersson EJ, Lester HA, & Dougherty DA (2005) Using physical chemistry to differentiate nicotinic from cholinergic agonists at the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **127,** 350-356.

234. Xiu XA, Puskar NL, Shanata JAP, Lester HA, & Dougherty DA (2009) Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature* **458,** 534-537.

235. Mu TW, Lester HA, & Dougherty DA (2003) Different binding orientations for the same agonist at homologous receptors: A lock and key or a simple wedge? *J. Am. Chem. Soc.* **125,** 6850-6851.

236. Lummis SCR, Beene DL, Harrison NJ, Lester HA, & Dougherty DA (2005) A cation-pi binding interaction with a tyrosine in the binding site of the GABA(C) receptor. *Chem. Biol.* **12,** 993-997.

237. Padgett CL, Hanek AP, Lester HA, Dougherty DA, & Lummis SCR (2007) Unnatural amino acid mutagenesis of the GABA(A) receptor binding site residues reveals a novel cation-pi interaction between GABA and beta(2)Tyr97. *J. Neurosci.* **27,** 886-892.

238. Pless SA*, et al.* (2008) A cation-pi interaction in the binding site of the glycine receptor is mediated by a phenylalanine residue. *J. Neurosci.* **28,** 10937-10942.

239. Santarelli VP, Eastwood AL, Dougherty DA, Ahern CA, & Horn R (2007) Calcium block of single sodium channels: Role of a pore-lining aromatic residue. *Biophys. J.* **93,** 2341-2349.

240. Santarelli VP, Eastwood AL, Dougherty DA, Horn R, & Ahern CA (2007) A cation-pi interaction discriminates among sodium channels that are either sensitive or resistant to tetrodotoxin block. *J. Biol. Chem.* **282,** 8044-8051.

241. Ahern CA, Eastwood AL, Dougherty DA, & Horn R (2008) Electrostatic contributions of aromatic residues in the local anesthetic receptor of voltage-gated sodium channels. *Circul. Res.* **102,** 86-94.

242. Ahern CA, Eastwood AL, Lester HA, Dougherty DA, & Horn R (2006) A cation-pi interaction between extracellular TEA and an aromatic residue in potassium channels. *J. Gen. Physiol.* **128,** 649-657.

243. McMenimen KA, Petersson EJ, Lester HA, & Dougherty DA (2006) Probing the Mg2+ blockade site of an N-methyl-D-aspartate (NMDA) receptor with unnatural amino acid mutagenesis. *ACS Chem. Biol.* **1,** 227-234.

244. Torrice MM, Bower KS, Lester HA, & Dougherty DA (2009) Probing the role of the cation-pi interaction in the binding sites of GPCRs using unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11919-11924.

245. Ye SX*, et al.* (2008) Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *J. Biol. Chem.* **283,** 1525-1533.

246. Huang LY*, et al.* (2008) Unnatural amino acid replacement in a yeast G protein-coupled receptor in its native environment. *Biochemistry* **47,** 5638-5648.

247. Hermanson GT (2008) *Bioconjugate Techniques* (Academic Press, London) 2 Ed.

248. Simon MD*, et al.* (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128,** 1003-1012.

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

250. Best MD (2009) Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **48,** 6571-6584.

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

252. Seebeck FP & Szostak JW (2006) Ribosomal synthesis of dehydroalanine-containing peptides. *J. Am. Chem. Soc.* **128,** 7150-7151.

253. Wang J, Schiller SM, & Schultz PG (2007) A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed.* **46,** 6849-6851.

254. Bernardes GJL, Chalker JM, Errey JC, & Davis BG (2008) Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: Versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* **130,** 5052-5053.

255. Guo JT, Wang JY, Lee JS, & Schultz PG (2008) Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed.* **47,** 6399-6401.

256. Chalker JM & Davis BG (2010) Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **14,** 781-789.

257. Nguyen DP, Alai MMG, Kapadnis PB, Neumann H, & Chin JW (2009) Genetically encoding N-epsilon-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* **131,** 14194-14195.

258. Nguyen DP, Alai MMG, Virdee S, & Chin JW (2010) Genetically directing epsilon-N, N-dimethyl-L-lysine in recombinant histones. *Chem. Biol.* **17,** 1072-1076.

259. Virdee S, Ye Y, Nguyen DP, Komander D, & Chin JW (2010) Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat. Chem. Biol.* **6,** 750-757.

260. Eger S, Scheffner M, Marx A, & Rubini M (2010) Synthesis of defined ubiquitin dimers. *J. Am. Chem. Soc.* **132,** 16337-16339.

261. Gamblin DP, Scanlan EM, & Davis BG (2009) Glycoprotein synthesis: an update. *Chem. Rev.* **109,** 131-163.

262. Floyd N, Vijayakrishnan B, Koeppe JR, & Davis BG (2009) Thiyl glycosylation of Olefinic Proteins: S-Linked Glycoconjugate Synthesis13. *Angew. Chem. Int. Ed.* **48,** 7798-7802.

263. Liu HT, Wang L, Brock A, Wong CH, & Schultz PG (2003) A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* **125,** 1702-1703.

264. Van Kasteren SI, Kramer HB, Gamblin DP, & Davis BG (2007) Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat. Protoc.* **2,** 3185-3194.

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

266. Liu CC & Schultz PG (2006) Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24,** 1436-1440.

267. Liu CC, Brustad E, Liu WS, & Schultz PG (2007) Crystal structure of a biosynthetic sulfo-hirudin complexed to thrombin. *J. Am. Chem. Soc.* **129,** 10648-10649.

268. 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.

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

270. Neumann H*, et al.* (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* **36,** 153-163.

271. Kang TJ, Yuzawa S, & Suga H (2008) Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation of epigenetic markers. *Chem. Biol.* **15,** 1166-1174.

272. Lammers M, Neumann H, Chin JW, & James LC (2010) Acetylation regulates Cyclophilin A catalysis, immunosuppression and HIV isomerization. *Nat. Chem. Biol.* **6,** 331-337.

273. Xie JM, Supekova L, & Schultz PG (2007) A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* **2,** 474-478.

274. Neumann H, Hazen JL, Weinstein J, Mehl RA, & Chin JW (2008) Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* **130,** 4028-4033.

275. Xu R*, et al.* (2004) Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J. Am. Chem. Soc.* **126,** 15654-15655.

276. Zhang ZW*, et al.* (2004) A new strategy for the synthesis of glycoproteins. *Science* **303,** 371-373.

277. Xu R*, et al.* (2009) Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J. Am. Chem. Soc.* **131,** 13883-13883.

278. Zhang ZW*, et al.* (2009) Retraction. *Science* **326,** 1187.

279. Fahmi NE, Dedkova L, Wang BX, Golovine S, & Hecht SM (2007) Site-specific incorporation of glycosylated serine and tyrosine derivatives into proteins. *J. Am. Chem. Soc.* **129,** 3586-3597.

280. Restifo NP (1996) The new vaccines: building viruses that elicit antitumor immunity. *Curr. Opin. Immunol.* **8,** 658-663.

281. Dalum I*, et al.* (1999) Therapeutic antibodies elicited by immunization against TNF-alpha. *Nat. Biotechnol.* **17,** 666-669.

282. Hackett CJ & Harn DA (2006) *Vaccine Adjuvants: Immunological and Clinical Principles* (Humana Press, Totowa, NJ).

283. Grünewald J*, et al.* (2008) Immunochemical termination of self-tolerance. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11276-11280.

284. Grünewald J*, et al.* (2009) Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4337-4342.

285. Crick FHC (1968) Origin of genetic code. *J. Mol. Biol.* **38,** 367-&.

286. Knight RD, Freeland SJ, & Landweber LF (2001) Rewiring the keyboard evolvability of the genetic code. *Nat. Rev. Genet.* **2,** 49-58.

287. Wong JTF (2005) Coevolution theory of the genetic code at age thirty. *Bioessays* **27,** 416-425.

288. Bacher JM, Hughes RA, Wong JTF, & Ellington AD (2004) Evolving new genetic codes. *Trends Ecol. Evol.* **19,** 69-75.

289. Wong JTF (1983) Membership mutation of the genetic code - loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6303-6306.

290. Bacher JM & Ellington AD (2001) Selection and characterization of Escherichia coli variants capable of growth on an otherwise toxic tryptophan analogue. *J. Bacteriol.* **183,** 5414-5425.

291. Bacher JM, Bull JJ, & Ellington AD (2003) Evolution of phage with chemically ambiguous proteomes. *BMC Evol. Biol.* **3,** 5414-5425.

292. Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5,** 568-574.

293. Kazlauskas RJ & Bornscheuer UT (2009) Finding better protein engineering strategies. *Nat. Chem. Biol.* **5,** 526-529.

294. Shivange AV, Marienhagen J, Mundhada H, Schenk A, & Schwaneberg U (2009) Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* **13,** 19-25.

295. Jackel C, Kast P, & Hilvert D (2008) Protein design by directed evolution. *Ann. Rev. Biophys.* **37,** 153-173.

296. Dougherty M & Arnold F (2009) Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* **20,** 486-491.

297. 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.

298. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79-88.

299. Choe H*, et al.* (2003) Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120. *Cell* **114,** 161-170.

300. Huang CC*, et al.* (2007) Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317,** 1930-1934.

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

302. Beynon RJ & Pratt JM (2005) Metabolic labeling of proteins for proteomics. *Mol. Cell. Proteomics* **4,** 857-872.

303. Mann M (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7,** 952-958.

304. Caro LG, Kolb JA, & Vantubergen RP (1962) High-resolution autoradiography 1. Methods. *J. Cell Biol.* **15,** 173-188.

305. 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.

306. van Hest JCM, Kiick KL, & Tirrell DA (2000) Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J. Am. Chem. Soc.* **122,** 1282-1288.

307. Beatty KE, Xie F, Wang Q, & Tirrell DA (2005) Selective dye-labeling of newly synthesized proteins in bacterial cells. *J. Am. Chem. Soc.* **127,** 14150-14151.

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

309. Sutton MA & Schuman EM (2006) Dendritic protein synthesis, synaptic plasticity, and memory. *Cell* **127,** 49-58.

310. Dieterich DC*, et al.* (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* **13,** 897-905.

311. Tcherkezian J, Brittis PA, Thomas F, Roux PP, & Flanagan JG (2010) Transmembrane receptor DCC associates with protein synthesis machinery and regulates translation. *Cell* **141,** 632-644.

312. Zhang MM, Tsou LK, Charron G, Raghavan AS, & Hang HC (2010) Tandem fluorescence imaging of dynamic S-acylation and protein turnover. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8627-8632.

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

314. Beatty KE*, et al.* (2010) Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **11,** 2092-2095.

315. del Amo DS*, et al.* (2010) Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc.* **132,** 16893-16899.

316. Hong V, Steinmetz NF, Manchester M, & Finn MG (2010) Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconj. Chem.* **21,** 1912-1916.

317. Song WJ*, et al.* (2010) A metabolic alkene reporter for spatiotemporally controlled imaging of newly synthesized proteins in mammalian cells. *ACS Chem. Biol.* **5,** 875-885.

318. Ngo JT*, et al.* (2009) Cell-selective metabolic labeling of proteins. *Nat. Chem. Biol.* **5,** 715-717.

319. Dieterich DC, Link AJ, Graumann J, Tirrell DA, & Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9482-9487.

320. Dieterich DC*, et al.* (2007) Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc.* **2,** 532-540.

321. Kramer G*, et al.* (2009) Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Mol. Cell. Proteomics* **8,** 1599-1611.

322. Nessen MA*, et al.* (2009) Selective enrichment of azide-containing peptides from complex mixtures. *J. Proteome Res.* **8,** 3702-3711.

323. Szychowski J*, et al.* (2010) Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J. Am. Chem. Soc.* **132,** 18351-18360.

324. Yang Y-Y, Grammel M, Raghavan AS, Charron G, & Hang HC (2010) Comparative analysis of cleavable azobenzene-based affinity tags for bioorthogonal chemical proteomics. *Chem. Biol.* **17,** 1212-1222.

325. Gygi SP*, et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994-999.

326. Ross PL*, et al.* (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3,** 1154-1169.

327. Lepthien S, Hoesl MG, Merkel L, & Budisa N (2008) Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16095-16100.

328. Brustad EM, Lemke EA, Schultz PG, & Deniz AA (2008) A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* **130,** 17664-17665.

329. Zhang ZW*, et al.* (2003) A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **42,** 6735-6746.

330. Merkel L, Hoesl MG, Albrecht M, Schmidt A, & Budisa N (2010) Blue fluorescent amino acids as in vivo building blocks for proteins. *ChemBioChem* **11,** 305-314.

331. Taskent-Sezgin H*, et al.* (2010) Azidohomoalanine: a conformationally sensitive IR probe of protein folding, protein structure, and electrostatics. *Angew. Chem. Int. Ed.* **49,** 7473-7475.

332. Groff D, Thielges MC, Cellitti S, Schultz PG, & Romesberg FE (2009) Efforts toward the direct experimental characterization of enzyme microenvironments: tyrosine 100 in dihydrofolate reductase. *Angew. Chem. Int. Ed.* **48,** 3478-3481.

333. Schultz KC*, et al.* (2006) A genetically encoded infrared probe. *J. Am. Chem. Soc.* **128,** 13984-13985.

334. Lee HM, Larson DR, & Lawrence DS (2009) Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds. *ACS Chem. Biol.* **4,** 409-427.

335. Mayer G & Heckel A (2006) Biologically active molecules with a "light switch." *Angew. Chem. Int. Ed.* **45,** 4900-4921.

336. Chou CJ, Young DD, & Deiters A (2009) A light-activated DNA polymerase. *Angew. Chem. Int. Ed.* **48,** 5950-5953.

337. Edwards WF, Young DD, & Deiters A (2009) Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chem. Biol.* **4,** 441-445.

338. Wang JY*, et al.* (2010) A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* **132,** 14812-14818.

339. Gautier A*, et al.* (2010) Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132,** 4086-4088.

340. Groff D, Wang F, Jockusch S, Turro NJ, & Schultz PG (2010) A new strategy to photoactivate green fluorescent protein. *Angew. Chem. Int. Ed.* **49,** 7677-7679.

341. Cazalis CS, Haller CA, Sease-Cargo L, & Chaikof EL (2004) C-terminal site-specific PEGylation of a truncated thrombomodulin mutant with retention of full bioactivity. *Bioconj. Chem.* **15,** 1005-1009.

342. Deiters A, Cropp TA, Summerer D, Mukherji M, & Schultz PG (2004) Site-specific PEGylation of proteins containing unnatural amino acids. *Biorg. Med. Chem. Lett.* **14,** 5743-5745.

343. Schoffelen S, Lambermon MHL, van Eldijk MB, & van Hest JCM (2008) Site-specific modification of Candida antarctica lipase B via residue-specific incorporation of a non-canonical amino acid. *Bioconj. Chem.* **19,** 1127-1131.

344. Teeuwen RLM*, et al.* (2009) "Clickable" elastins: elastin-like polypeptides functionalized with azide or alkyne groups. *Chem. Commun.* 4022-4024.

345. Strable E*, et al.* (2008) Unnatural amino acid incorporation into virus-like particles. *Bioconj. Chem.* **19,** 866-875.

346. Prasuhn DE*, et al.* (2008) Plasma clearance of bacteriophage Q beta particles as a function of surface charge. *J. Am. Chem. Soc.* **130,** 1328-1334.

347. Goto Y*, et al.* (2008) Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **3,** 120-129.

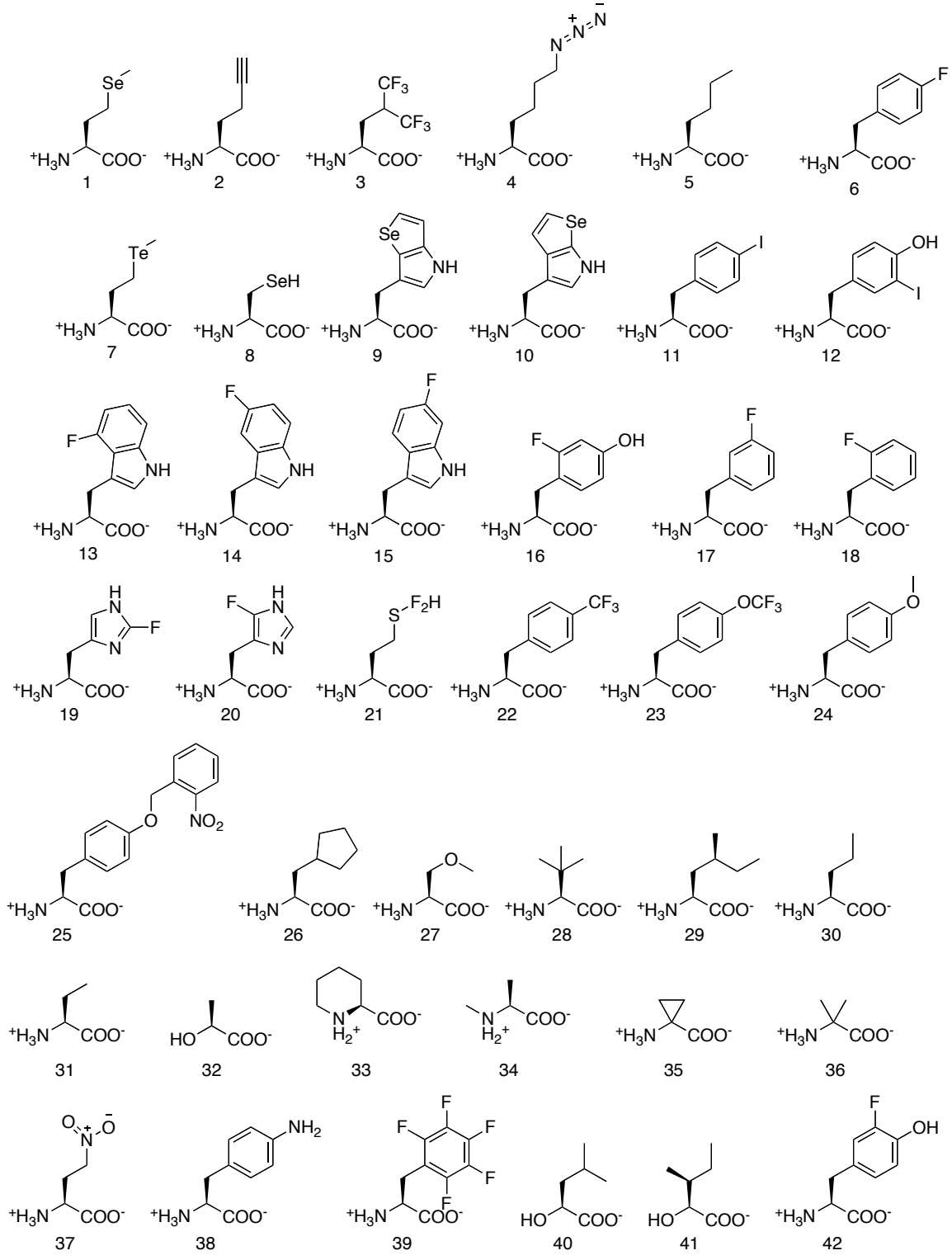
348. Sako Y, Morimoto J, Murakami H, & Suga H (2008) Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* **130,** 7232-7234.

349. Kawakami T*, et al.* (2009) Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **5,** 888-890.

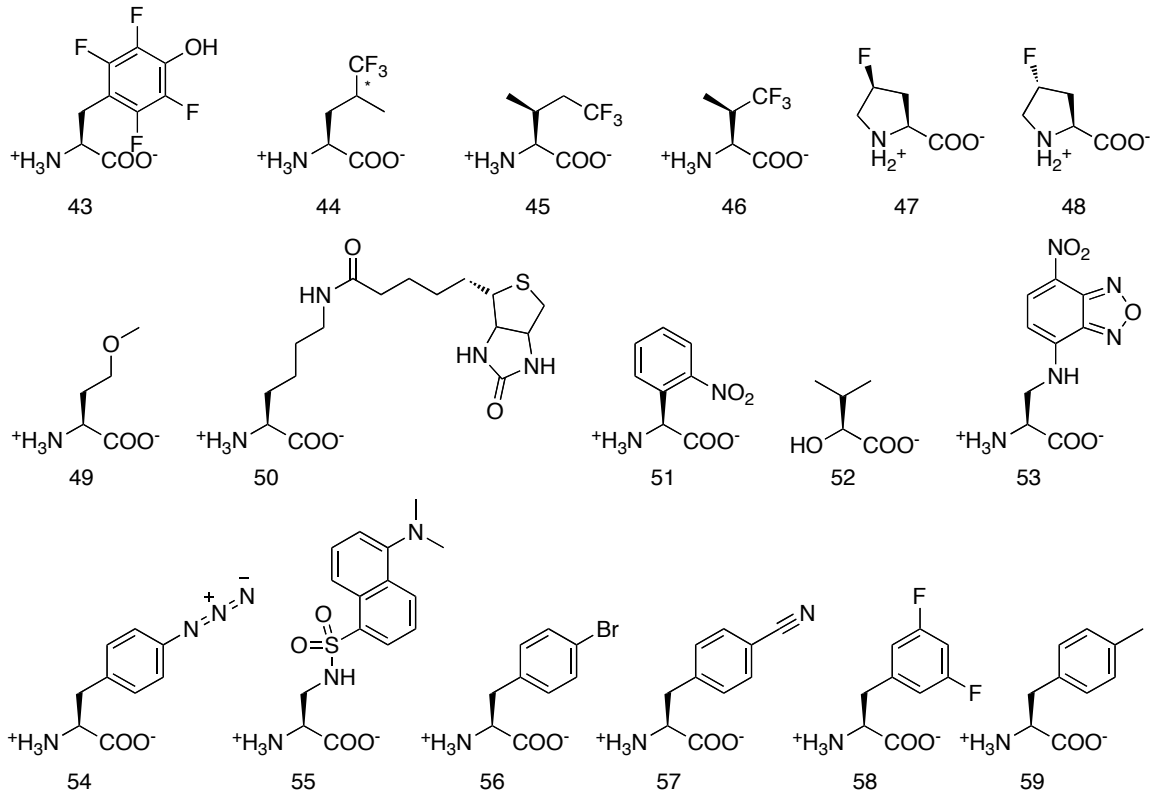
350. Clark MA (2010) Selecting chemicals: the emerging utility of DNA-encoded libraries. *Curr. Opin. Chem. Biol.* **14,** 396-403.

351. Doi Y, Ohtsuki T, Shimizu Y, Ueda T, & Sisido M (2007) Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc.* **129,** 14458-14462.

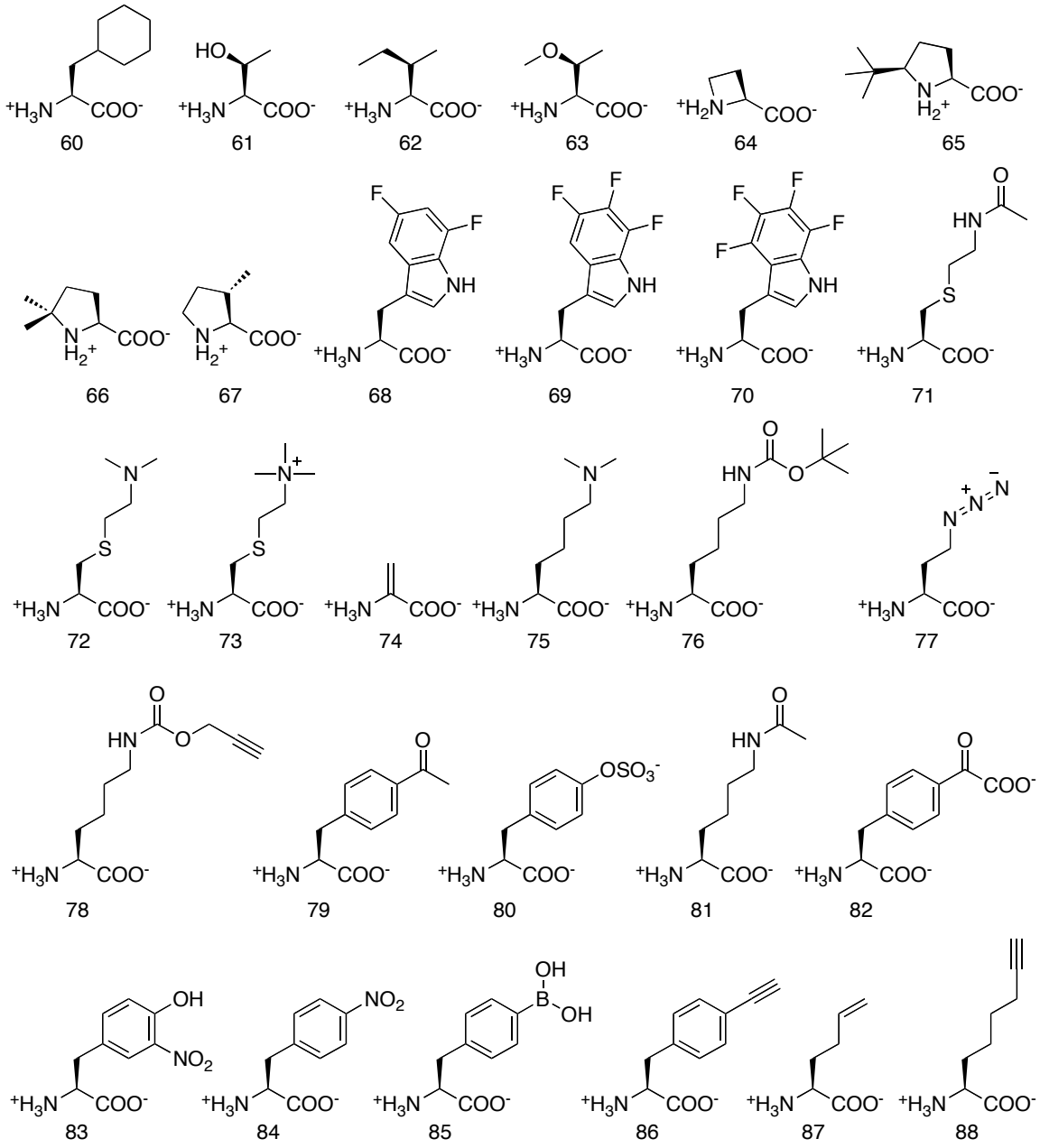
352. Pemble CW, Johnson LC, Kridel SJ, & Lowther WT (2007) Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by Orlistat. *Nat. Struct. Mol. Biol.* **14,** 704-709.



Scheme 1.1. Noncanonical amino acids. Caption to follow.



Scheme 1.1. Noncanonical amino acids, part 2. Caption to follow.



Scheme 1.1. Noncanonical amino acids, part 3. **1**, selenomethionine. **2**, homopropargylglycine. **3**, hexafluoroleucine. **4**, azidonorleucine. **5**, norleucine. **6**, *p*-fluorophenylalanine. **7**, telluromethionine. **8**, selenocysteine. **9**, β-selenolo[3,2-*b*]pyrolyl-alanine **10**, [6,7]selenatryptophan. **11**, *p*-iodophenylalanine. **12**, 3-iodotyrosine. **13**, 4-fluorotryptophan. **14**, 5-fluorotryptophan. **15**, 6-fluorotryptophan. **16**, *m*-fluorotyrosine. **17**, *m*-fluorophenylalanine. **18**, *o*-fluorophenylalanine. **19**, 2-fluorohistidine. **20**,

4-fluorohistidine. **21**, difluoromethionine. **22**, *p*-trifluoromethylphenylalanine. **23**, *p*-trifluoromethoxyphenylalanine. **24**, *p*-methoxyphenylalanine (also called *O*-methyltyrosine). **25**, *O*-nitrobenzyltyrosine. **26**, amino-3-cyclopentylpropanoic acid. **27**, *O*-methylserine. **28**, tert-leucine. **29**, 2-amino-4-methylhexanoic acid (homoisoleucine). **30**, norvaline. **31**, ethylglycine. **32**, α-hydroxyalanine. **33**, pipecolic acid. **34**, *N*-methylalanine. **35**, cyclopropylglycine. **36**, α-aminoisobutyric acid. **37**, γ-nitroglutamate. **38**, *p*-aminophenylalanine. **39**, pentafluorophenylalanine. **40**, α-hydroxyleucine. **41**, α-hydroxyisoleucine. **42**, *p*-fluorotyrosine. **43**, tetrafluorotyrosine. **44**, trifluoroleucine. **45**, trifluoroisoleucine. **46**, trifluorovaline. **47**, 4*R*-fluoroproline. **48**, 4*S*-fluoroproline. **49**, methoxinine. **50**, biocytin. **51**, (2-nitrophenyl)glycine. **52**, α-hydroxyvaline. **53**, 3-*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-2,3-diaminopropionic acid (NBD-Dap). **54**, *p*-azidophenylalanine. **55**, dansylalanine. **56**, *p*-bromophenylalanine. **57**, *p*-cyanophenylalanine. **58**, 3,5-difluorophenylalanine. **59**, *p*-methylphenylalanine. **60**, cyclohexylalanine, **61**, *allo*-threonine. **62**, *allo*-isoleucine. **63**, *allo*-*O*-methylthreonine. **64**, azetidine-2-carboxylic acid. **65**, 5-tert-butylproline. **66**, 5,5-dimethylproline. **67**, 3*S*-methylproline. **68**, 5,7-difluorotryptophan. **69**, 5,6,7-trifluorotryptophan. **70**, 4,5,6,7-tetrafluorotryptophan. **71**, sulfated acetyllysine. **72**, sulfated dimethyllysine. **73**, sulfated trimethyllysine. **74** dehydroalanine. **75**, dimethyllysine **76**, *N*-tertbutyloxycarbonyllysine. **77**, azidohomoalanine. **78**, alkyne analog of pyrrolysine. **79**, *p*-acetylphenylalanine. **80**, sulfotyrosine. **81**, N^{ϵ}-acetyllysine. **82**, *p*-carboxymethylphenylalanine. **83**, nitrotyrosine. **84**, *p*-nitrophenylalanine. **85**, *p*-boronophenylalanine. **86**, *p*-alkynylphenylalanine. **87**, homoallylglycine. **88**, 2-aminooctynoic acid.

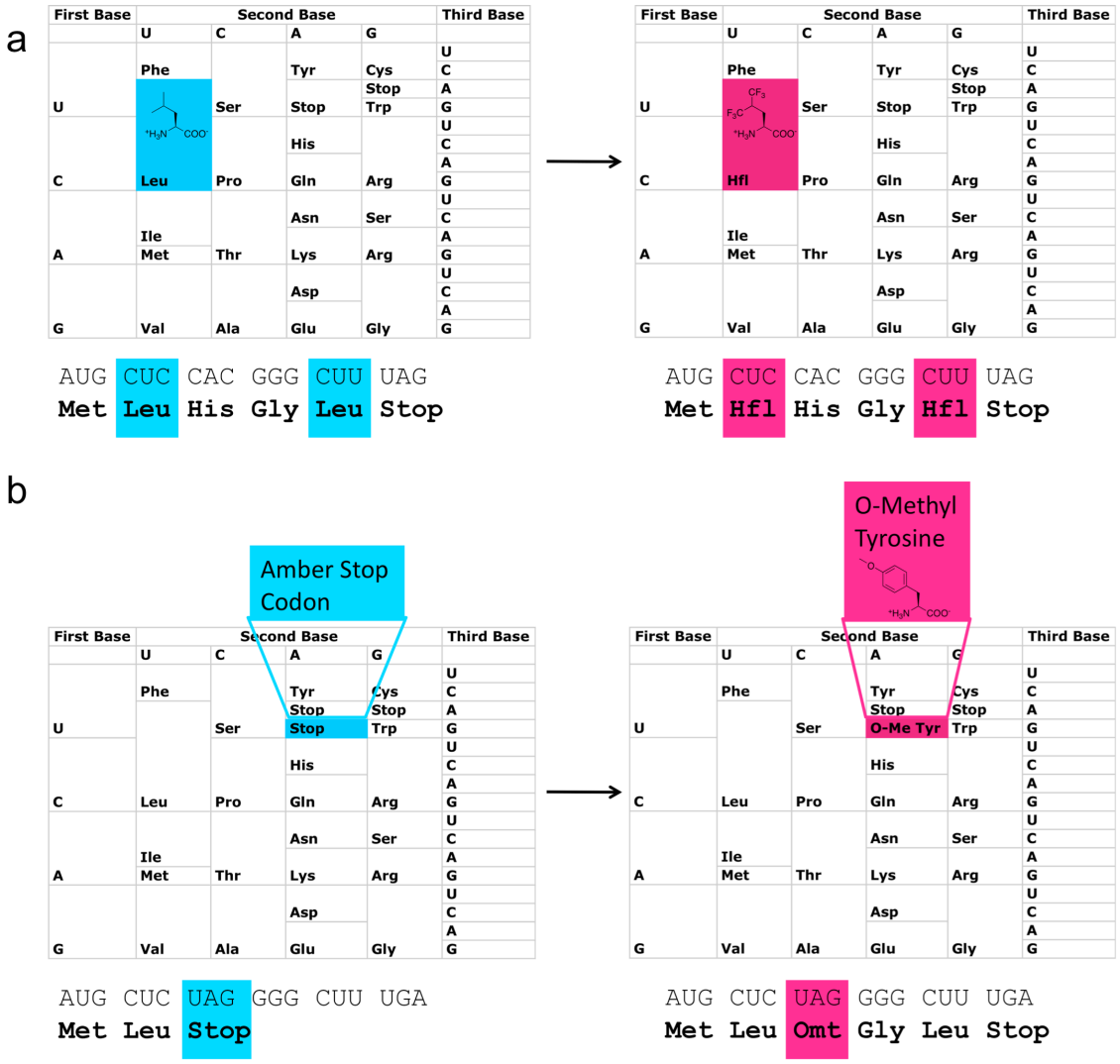


Figure 1.1. Overview of strategies for genetically incorporating noncanonical amino acids into proteins. (*a*) Residue-specific incorporation. A set of codons specifying one of the twenty canonical amino acids is "reprogrammed" to code for a ncAA. In this case, the six leucine (Leu) codons of the *E. coli* genetic code have been reassigned to the fluorinated leucine analog hexafluoroleucine (Hfl, **3**). (*b*) Site-specific incorporation. A stop codon is converted into a sense codon. The depicted example shows the amber stop codon being reprogrammed to code for *O*-methyltyrosine (Omt, **24**).

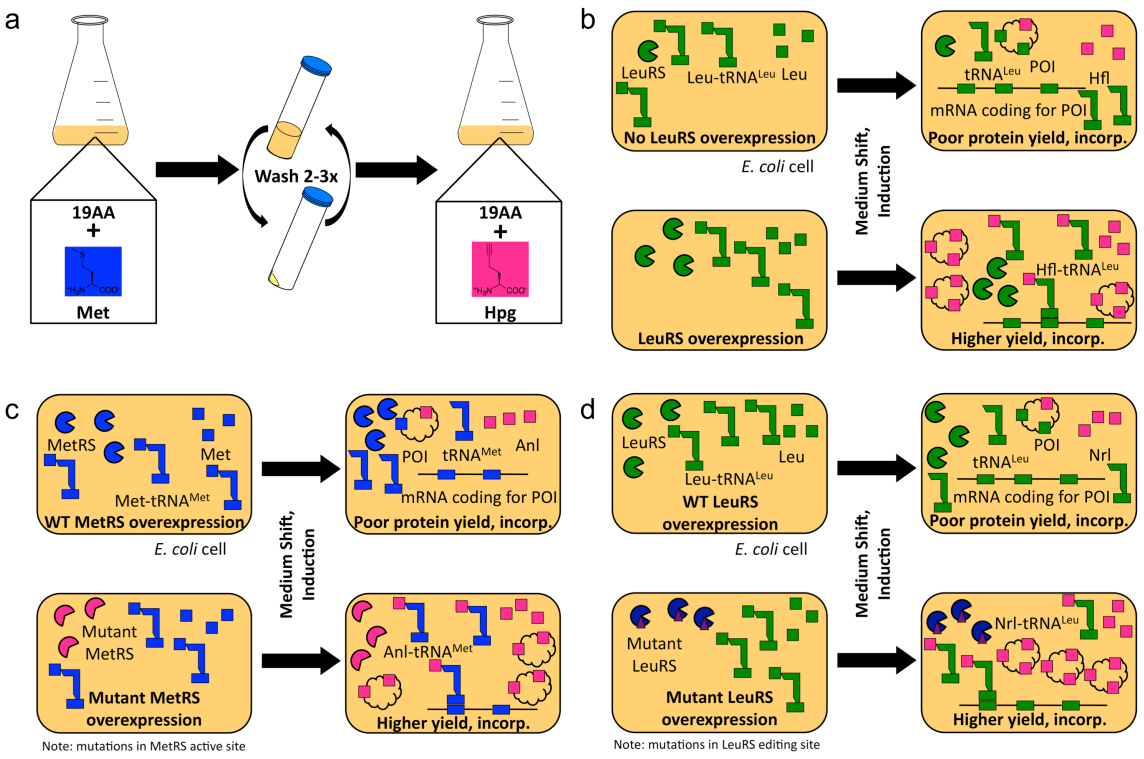


Figure 1.2. Methods for residue-specific incorporation of noncanonical amino acids into proteins. (*a*) Medium shift procedure to produce proteins containing the methionine (Met) analog homopropargylglycine (Hpg, **2**). Methionine-auxotrophic *E. coli* cells are first grown in medium containing twenty canonical amino acids. After reaching a particular optical density, the culture is pelleted and washed two to three times in an isotonic salt solution to remove methionine from the cells. After the final wash, cells are pelleted and resuspended in medium containing nineteen canonical amino acids plus Hpg. If a particular protein of interest is to be overproduced, protein expression can be initiated by the addition of an inducer. (*b–d*) Additional modifications to expression hosts. Engineering auxotrophic *E. coli* strains enables the incorporation of additional ncAAs into proteins. The examples that follow are depicted for the case of overexpressed proteins, but can also be applied in the absence of a particular protein of interest for proteomic applications. (*b*) Overexpression of aminoacyl-tRNA synthetase (aaRS). Noncanonical amino acids that

are poor aaRS substrates do not support protein synthesis unless the *E. coli* strain employed has augmented aaRS activity. In the case shown here, overexpression of leucyl-tRNA synthetase (LeuRS) enables efficient global replacement of hexafluoroleucine (Hfl, **3**) in place of leucine in a protein of interest (POI). (*c*) Overexpression of mutant aaRS. When no wild-type aaRS can activate a particular amino acid analog efficiently, mutant enzymes can be engineered and employed in *E. coli* expression hosts. Overexpression of a mutant methionyl-tRNA synthetase (MetRS) enables the global replacement of Met with azidonorleucine (Anl, **4**). (*d*) Overexpression of editing-deficient mutant aaRS. Some amino acid analogs are kinetically competent aaRS substrates but are subjected to editing after activation or aminoacylation. Altering the editing activity of an aaRS can enable incorporation of amino acid analogs that are proofread by the wild-type aaRS. For example, overexpression of an editing-deficient LeuRS enables norleucine (Nrl, **5**) to replace Leu in proteins.

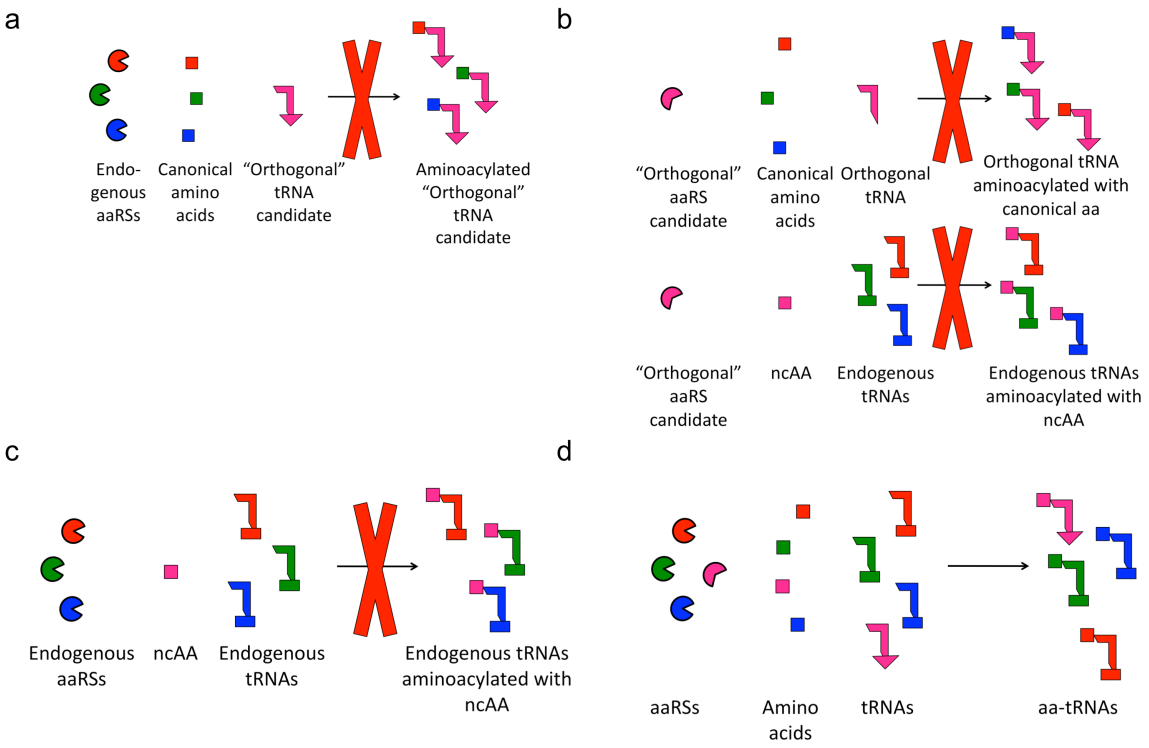


Figure 1.3. Orthogonality requirements for adding tRNAs and aminoacyl-tRNA synthetases (aaRSs) to a translation system. (*a*) Transfer RNA orthogonality. The heterologous tRNA to be introduced should not be a substrate for any of the endogenous aminoacyl-tRNA synthetases in order to avoid the aminoacylation of the tRNA with canonical amino acids. (*b*) AaRS orthogonality. The heterologous aaRS should not aminoacylate the heterologous tRNA with canonical amino acids or aminoacylate endogenous tRNAs with ncAAs. (*c*) Noncanonical amino acid orthogonality. The amino acid to be "added" to the genetic code should not be a substrate for any of the endogenous aaRSs. (*d*) Orthogonal pair. A properly functioning orthogonal aaRS-tRNA pair performs its aminoacylation task efficiently and specifically in the context of the endogenous translational machinery.

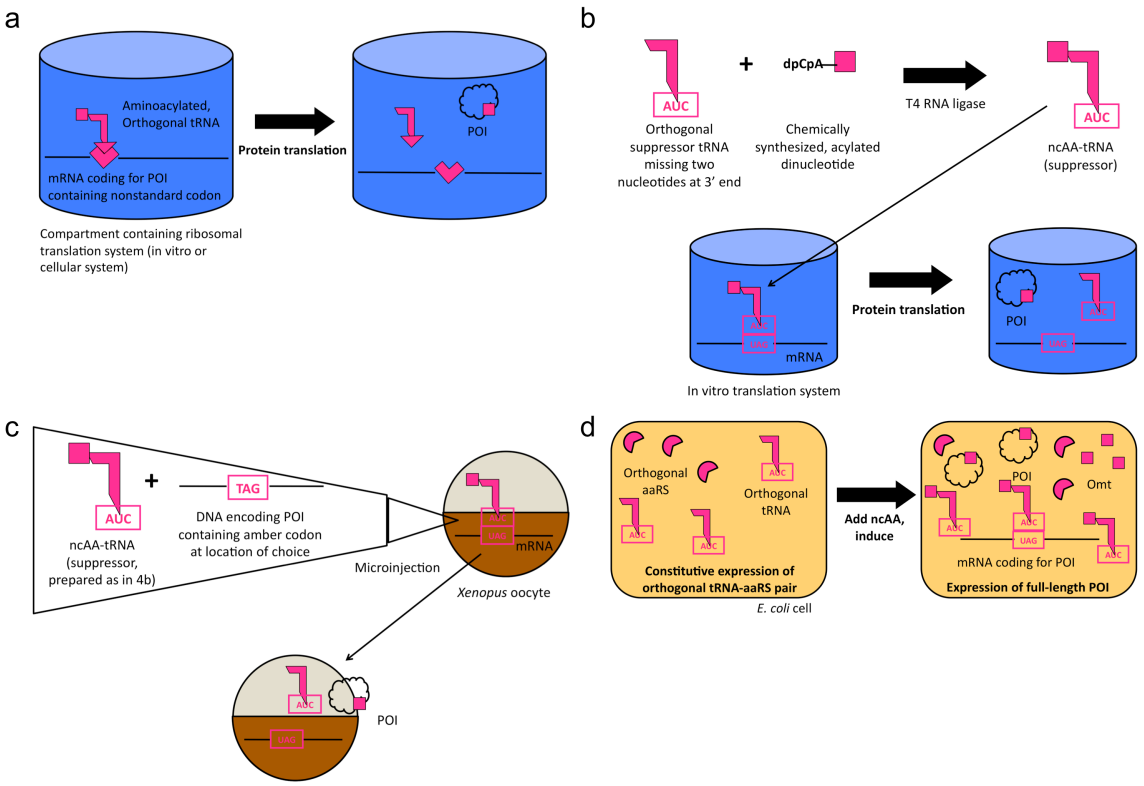


Figure 1.4. Methods for site-specific incorporation of noncanonical amino acids into proteins. (*a*) Suppression of nonstandard codon for site-specific incorporation. An aminoacylated, orthogonal tRNA is introduced into a translation system to decode a nonsense or other nonstandard codon, resulting in a full-length protein containing a noncanonical amino acid at one specified position in the protein. (*b–d*) Systems for employing site-specific incorporation techniques. (*b*) In vitro translation. Chemically synthesized, orthogonal aminoacyl-tRNAs can be employed in conjunction with an in vitro translation system to effect site-specific ncAA incorporation. (*c*) Microinjection into *Xenopus* oocytes. Simultaneous injection of a gene containing a nonstandard codon to be suppressed and an aminoacylated, orthogonal tRNA allows for the synthesis of proteins containing a site-specifically incorporated noncanonical amino acid inside a living cell. Membrane proteins are commonly studied using this technique. (*d*) Use of an orthogonal

pair. *E. coli* cells outfitted with an additional aaRS-tRNA pair can synthesize proteins containing a site-specifically incorporated twentyfirst amino acid. In this case, the aaRS enzymatically aminoacylates the tRNA during the course of the experiment, requiring no chemical synthesis of aminoacyl-tRNAs. Omt, *O*-methyltyrosine (**24**).

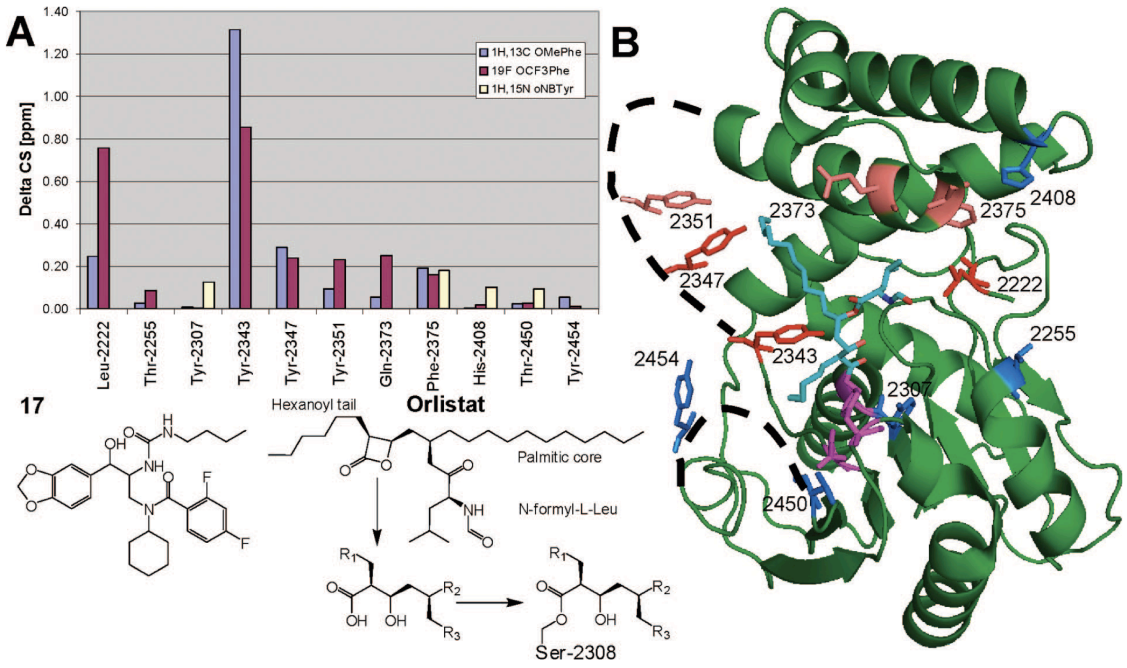


Figure 1.5. NMR studies of ligand binding employing noncanonical amino acids as isotopically labeled probes. (*a*) Combined chemical shift data of tool compound **17** binding to the thioesterase domain of human fatty acid synthase (FAS-TE). Chemicals shift changes ΔCS have been scaled as described by Cellitti et al. in order to enable comparison between ^{19}F, ^{13}C, and ^{15}N NMR experiments (91). Conformational exchange prevented acquisition of chemical shift data on some *o*-nitrobenzyl-tyrosine (oNBTyr, **25**) mutants. OMePhe, *p*-methoxyphenylalanine (**24**). OCF3Phe, *p*-trifluoromethoxylphenylalanine (**23**). (*b*) Structure of FAS-TE covalently modified with orlistat (2PX6.pdb) (352). The average chemical shift changes induced by the binding of **17** are calculated for each ncAA and color-coded for each position: ΔCS < 0.1 ppm, blue; 0.1-0.2 ppm, salmon; > 0.2 ppm, red. Disordered loops are indicated by dashed lines. The active site residues Ser-2308, Asp-2338, and His-2481 are shown in magenta. Adapted from (91) with permission. © 2008 American Chemical Society.

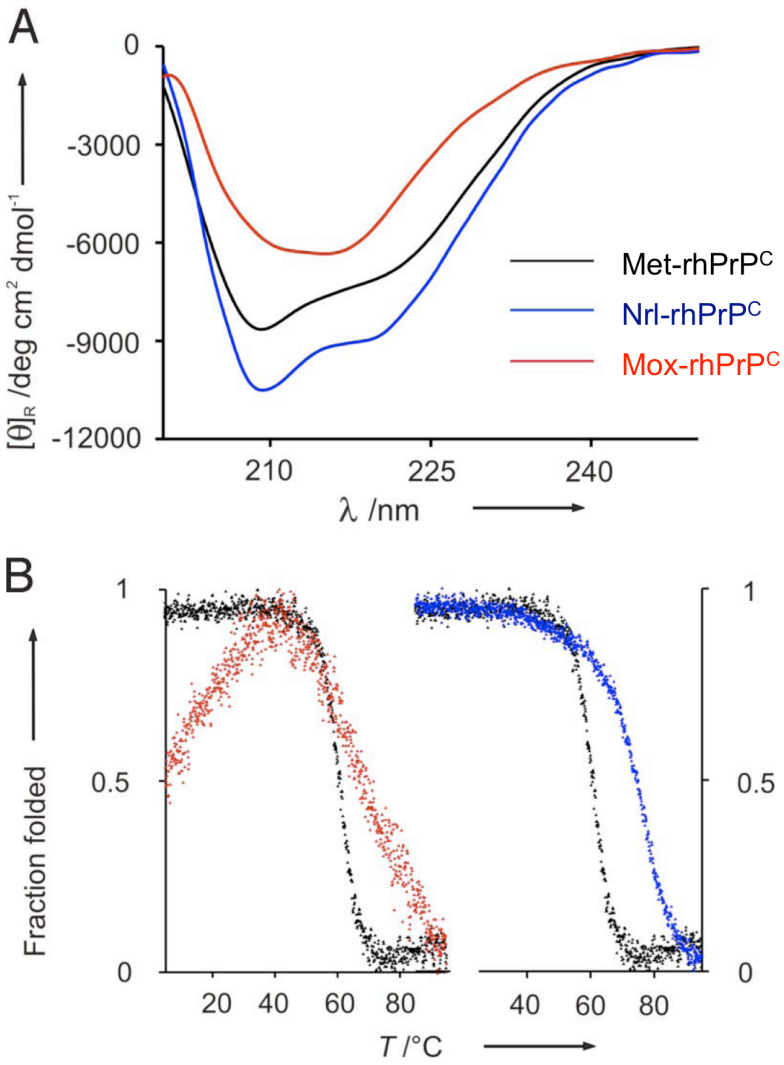


Figure 1.6. Secondary structure analysis and thermal denaturation of recombinant human prion protein (rhPrP^{C}) containing methionine analogs. (*a*) Circular dichroism spectra of Met-rhPrP^{C} and its norleucine (Nrl, **5**) and methoxinine (Mox, **49**) variants at 37 °C and 0.2 mg/mL in 10 mM Mes buffer at pH 6.0. (*b*) Thermal denaturation monitored by the changes of dichroic intensities at 222 nm as a function of temperature. Note that both the secondary structural characteristics and denaturation behaviors are greatly influenced by the hydrophobicities of the amino acids incorporated at the Met positions. Adapted from (194) with permission.

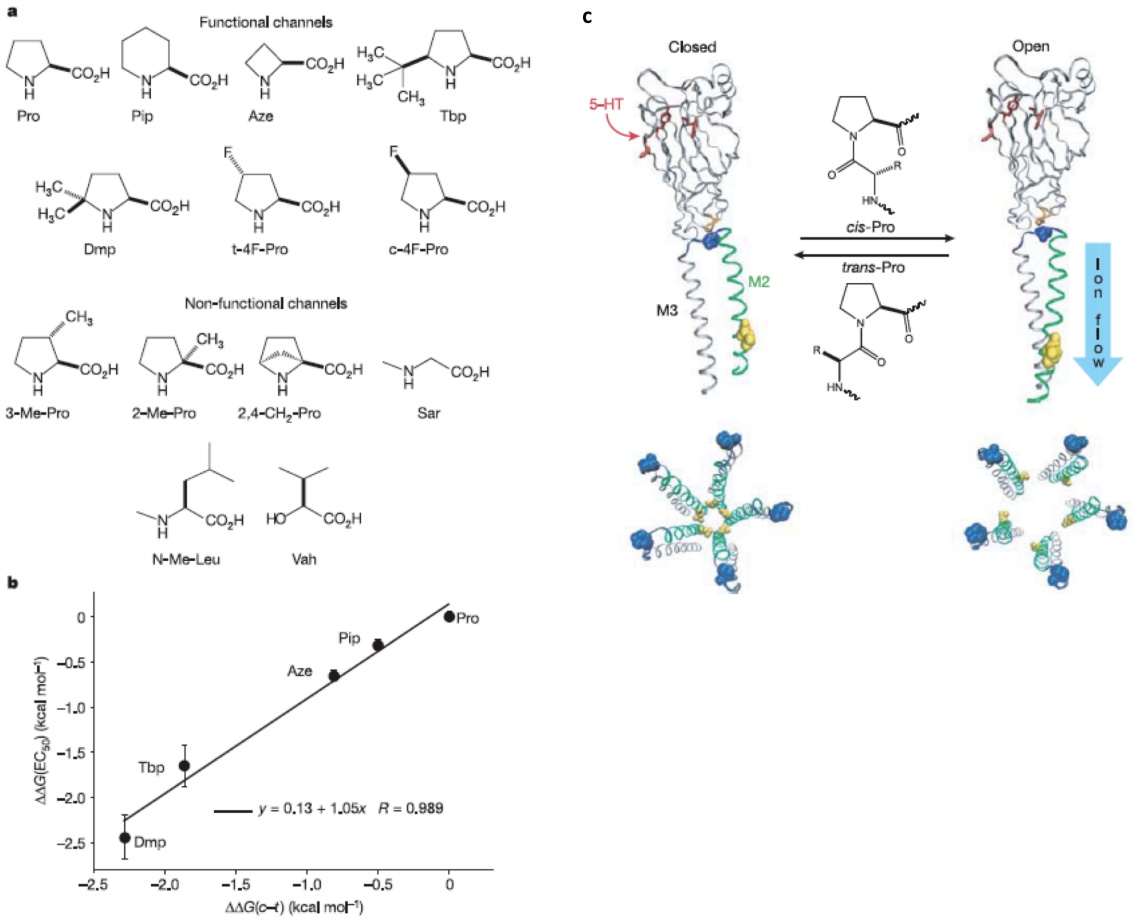


Figure 1.7. Unnatural amino acid mutagenesis on the 5-hydroxytryptamine type 3 receptor leads to a new model for receptor gating. (*a*) NcAAs used in study. These residues were incorporated in place of a key proline residue (Proline 8*) within the receptor. (*b*) The thermodynamics of the *cis–trans* isomerization propensities (ΔΔG(*c*-*t*)) of ncAAs and receptor activation by 5-hydroxytryptophan (ΔΔG(EC_{50})) are strongly correlated, suggesting that the isomerization properties of the residue at the position of interest are critical for forming functional channels. (*c*) Proposed model for receptor gating. Isomerization of the proline residue shown in blue (proline 8*) dictates M2-M3 loop conformation, which in turn controls ion flow through the channel. Adapted from (218) with permission. © 2005 Nature Publishing Group.

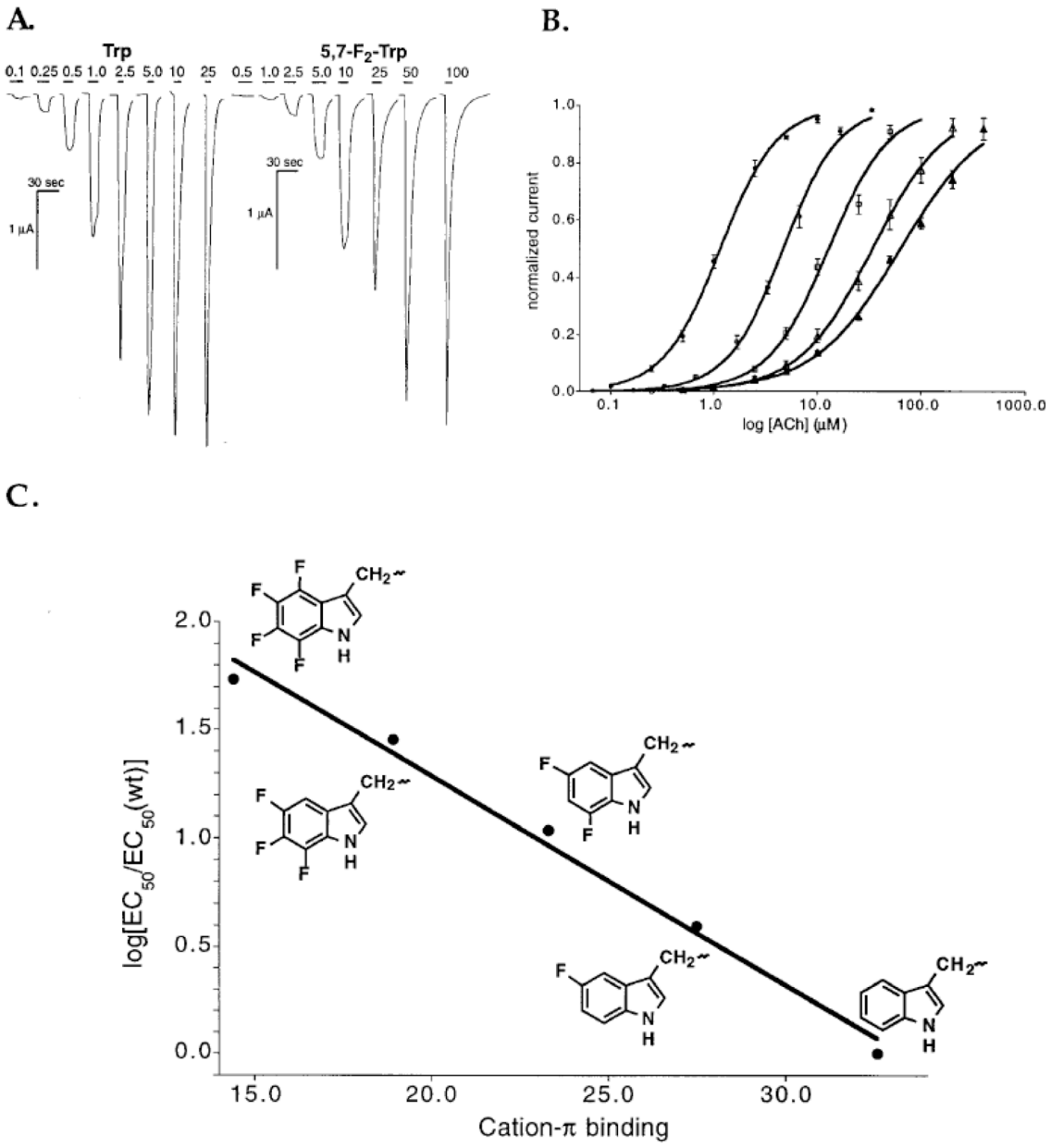


Figure 1.8. Use of noncanonical amino acids to investigate cation-π interactions in the muscle-type nicotinic acetylcholine receptor (nAChR) using patch-clamp experiments. (*a*) Receptor response to increasing doses of acetylcholine as measured by voltage-clamp current traces. The two experiments shown are from oocytes expressing tryptophan (Trp, Left) and 5,7-F_2-Trp (Right, **68**) at α149. Bars represent application of acetylcholine (μM). (*b*) Dose–response relations and fits to the Hill equation for (left to right): Trp; 5-F-Trp

(**14**); 5,7-F_2-Trp (**68**); 5,6,7-F_3-Trp (**69**); and 4,5,6,7-F_4-Trp (**70**). (*c*) Plot of log[EC_{50}/EC_{50}(wild type)] versus quantum mechanically calculated cation–π binding ability at α149 for the same residues as in (*b*). The strong correlation observed confirms that residue α149 binds to acetylcholine through a cation-π interaction. Adapted from (231) with permission.

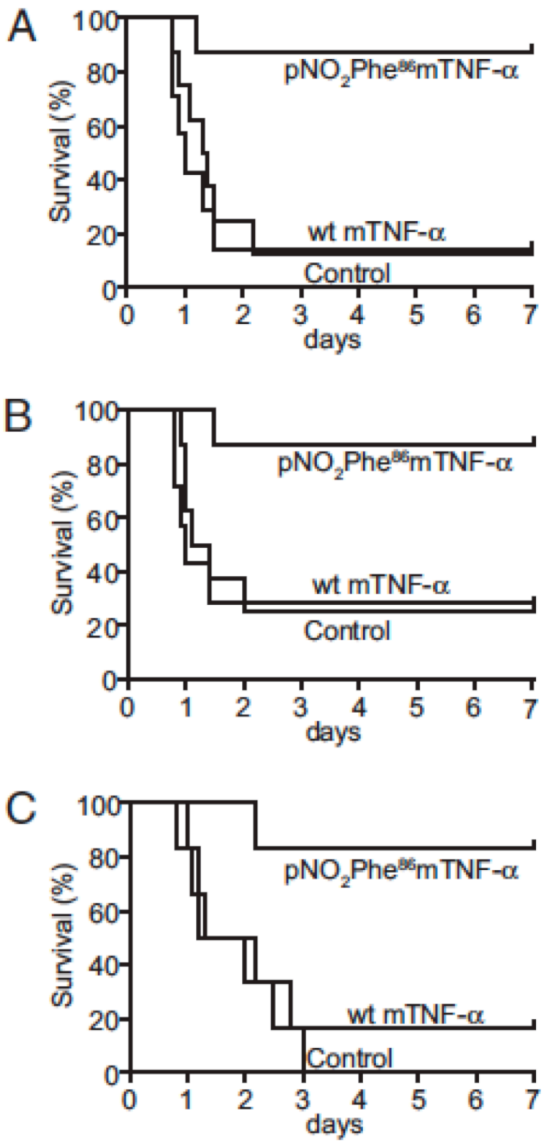


Figure 1.9. Use of ncAAs to breaking immunochemical self-tolerance. Immunization with *p*-nitrophenylalanine (**84**) at position 86 of murine tumor necrosis factor-α (pNO_2Phe^{86}mTNF-α) improves survival of mice in a TNF-α-dependent severe endotoxemia model. Kaplan–Meier survival plots of mice receiving active or passive immunizations are shown. (*a*) Mice (eight per group) immunized with pNO_2Phe86mTNF-α or WT mTNF-α are compared with seven mice receiving sham immunizations. The survival advantage of mice immunized with pNO_2Phe86mTNF-α ($P <$

0.01) versus WT is shown. (*b*–*c*) The survival advantage is preserved when antibodies (*b*) or pooled serum (*c*) from mice immunized with pNO_2Phe86mTNF-α is transferred to other mice. Mice (eight per group) injected with 100 μg of purified IgG from pNO_2Phe86mTNF-α or WT immunized mice were compared with controls receiving saline injection. Survival advantage of mice immunized with pNO_2Phe86mTNF-α ($P <$ 0.01) versus WT is shown. (*c*) Mice (six per group) received 100 μL of pooled serum from mice immunized with pNO_2Phe86mTNF-α or WT mTNF-α. Survival advantage of mice immunized with pNO_2Phe86mTNF-α ($P < 0.01$) versus WT is shown. Adapted from (283) with permission.

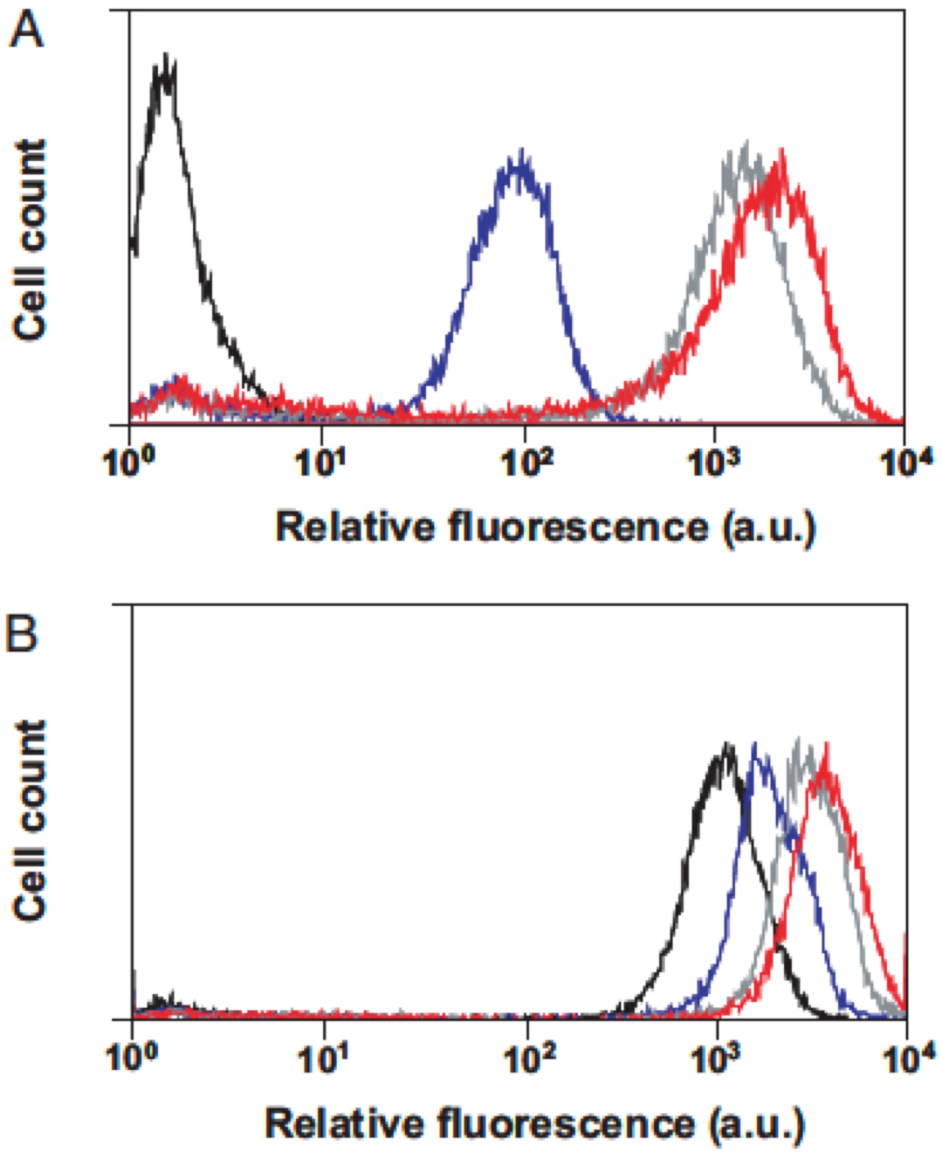


Figure 1.10. Flow cytometric analysis of cells expressing green fluorescent protein (GFP) and GFP variants during the course of evolving fluorinated GFPs. Proteins were expressed in media depleted of Leu and supplemented with trifluoroleucine (Tfl, **44**) (*a*) or in media containing all twenty canonical amino acids (*b*). Black line, GFPm (parent); blue line, 4.2.2 (mutant isolated after sequential construction and enrichment of four libraries); gray line, 8.3.3 (mutant isolated after enrichment of eight libraries); and red line, 11.3.3 (mutant isolated after enrichment of eleven libraries). The evolved, fluorinated variants have

regained fluorescence, likely through improved folding properties. Adapted from (189) with permission.

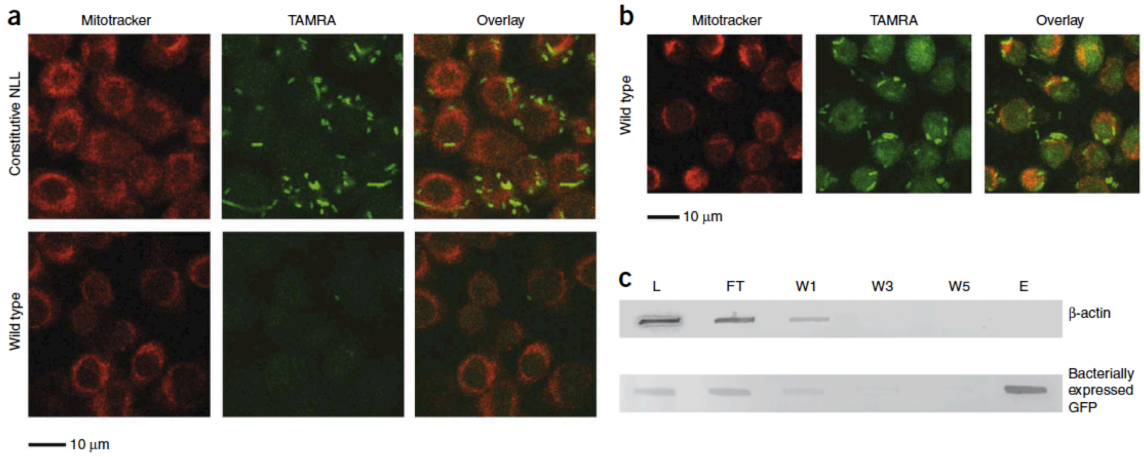


Figure 1.11. Cell-selective labeling in mixtures of bacterial and mammalian cells. (*a*) Fluorescence images of mixed cultures containing bacteria attached to or internalized by mouse alveolar macrophages. Infection was performed in medium containing azidonorleucine (**4**). Bacterial cells constitutively expressing the methionyl tRNA synthetase variant NLL-MetRS were labeled by TAMRA-alkyne (constitutive NLL), whereas cells lacking the NLL-MetRS (wild type) are visible only in the GFP channel (not shown). Macrophages were labeled with Mitotracker Deep Red (Invitrogen) and exhibited very low TAMRA background emission. In all cases, conjugation of TAMRA-alkyne was confined to bacterial cells expressing the NLL-MetRS. (*b*) Fluorescence images of macrophage infection with wild-type bacteria performed in the presence of azidohomoalanine. Protein synthesis by macrophages is indicated by strong TAMRA-alkyne emission from both bacterial cells and macrophages. (*c*) Macrophages were infected with bacterial cells that express GFP under induction with IPTG and that constitutively express the NLL-MetRS. Infection was performed in medium containing IPTG and azidonorleucine to facilitate bacterial synthesis and labeling of GFP. Total cell lysate from the infection was subjected to conjugation with alkyne-functionalized biotin; labeled proteins were enriched with streptavidin avidity. Bacterially expressed GFP and

mammalian β-actin were followed by immunoblots. Analyses of the lysate (L), unbound flow-through (FT), washes (W1, W3, W5) and eluent (E) reveal a separation of bacterial and mammalian representative proteins. Bacterially expressed GFP was labeled with azidonorleucine and thus subject to conjugation to biotin and enrichment with streptavidin. Proteins originating from macrophages, including β-actin, were not labeled with azidonorleucine and therefore were not conjugated to alkyne-functionalized biotin. Adapted from (318) with permission.

CHAPTER 2

Homoisoleucine: A Translationally Active Leucine Surrogate of Expanded Hydrophobic Surface Area

This chapter first appeared as a communication in *ChemBioChem*: James A. Van Deventer, John D. Fisk, and David A. Tirrell. *ChemBioChem* **2011,** 12, 700-702.

Abstract

Homoisoleucine (Hil) serves as an effective surrogate for leucine with respect to protein translation in bacterial cells. Replacement of Leu by Hil stabilizes coiled-coil peptides, as shown by the elevation of the thermal denaturation temperature. The increase in denaturation temperature is larger than that observed previously for replacement of Leu by trifluoroleucine.

Introduction

Noncanonical amino acids (ncAAs) provide useful tools for the investigation and control of protein behavior (1-4). Several laboratories have used ncAAs to explore the role of hydrophobic forces in stabilizing proteins, including prion proteins, T4 lysozyme, chloramphenicol acetyltransferase, green fluorescent protein, and coiled-coil and helix-bundle proteins (5-12). Here we examine the consequences of introducing the leucine surrogate (2*S*,4*S*)-2-amino-4-methylhexanoic acid (homoisoleucine, Hil, **2**; scheme 2.1) (5) into the coiled-coil peptide A1 (figure 2.1) (13). A1 contains six heptad repeats, designated (*abcdefg*), that mediate self-association of the peptide in aqueous solutions. Previous studies have shown that replacement of leucine by (2*S*,4*R*)-trifluoroleucine (Tfl, **3**) or hexafluoroleucine (Hfl, **4**) at the *d* positions of the heptad repeats leads to substantial stabilization of the coiled-coil structure of A1 (9, 10, 12), presumably through enhanced hydrophobic interactions between fluorinated peptide strands. Similar results have been obtained in other fluorinated coiled-coil and helix-bundle systems (14-16).

Whitesides and coworkers have pointed out that hydrocarbons and fluorocarbons exhibit equivalent “intrinsic” hydrophobicities when changes in molecular surface area are

taken into account (17). Marsh and coworkers have argued that the "efficient packing" of the bulkier fluorinated amino acids in helix bundle cores can be more important than fluorination per se (18). Because the molecular surface areas of Hil and Tfl are nearly identical (and larger than that of Leu by 14–19 $Å^2$) (19), we imagined that replacement of Leu by Hil might stabilize coiled-coil peptides such as A1.

Results and Discussion

We first focused our attention on the translational activity of Hil in bacterial cells. Schultz and coworkers reported incorporation of Hil into proteins via chemical misacylation of suppressor tRNA and *in vitro* translation (5), but we are unaware of previous studies of incorporation of Hil into cellular proteins. *E. coli* strain LAM1000 (a previously reported leucine auxotroph) was cotransformed with expression plasmid pA1EL and repressor plasmid pREP4 (10). pA1EL codes for both the protein A1 and a constitutively expressed copy of the *E. coli* leucyl-tRNA synthetase (LeuRS) gene. Protein expression was induced in minimal medium depleted of Leu and supplemented with Hil (see materials and methods for details). Electrophoretic analysis of whole-cell lysates indicated high-level protein expression in media containing as little as 0.25 mM L-Hil, a concentration comparable to the concentrations of canonical amino acids in minimal media.

ATP-PP_i exchange assays confirmed that Hil is activated by the *E. coli* LeuRS, albeit at a rate substantially lower than that characteristic of the natural substrate (table 2.1). The reduced rate of activation of Hil is consistent with our observation that Hil supports high-level protein expression only when LeuRS is overexpressed in the bacterial host.

A1 samples containing Leu and Hil were purified from 25 mL cultures in yields of 15.9 ± 2.5 mg L^{-1} and 10.2 ± 1.1 mg L^{-1}, respectively. Liter-scale expression and purification improved the yields of A1 sequence variants containing Hil two- to threefold. Liquid chromatography/mass spectrometry (LC/MS) indicated replacement of at least 97% of Leu by Hil (see "Determination of amino acid replacement levels" in materials and methods).

Figure 2.2A shows circular dichroism spectra of 10 μM solutions of the Leu- (Leu-A1) and Hil- (Hil-A1) forms of A1. Strong minima at 208 and 222 nm confirm that both proteins assume α-helical structures; K2D2 (20) analysis indicates helical contents of 60%–69%, consistent with the fact that the putative heptad repeats constitute 57% of the peptide sequence. A1 is expected to exist primarily as dimers under the conditions employed here (21).

Replacement of Leu by Hil increases the denaturation temperature of A1, as expected. Figure 2.2B shows the molar ellipticities at 222 nm of solutions of Leu-A1 and Hil-A1 as functions of temperature. Fitting the CD data to a model of a two-state transition between folded dimer and unfolded monomer states (22) yielded melting temperatures of 58.7 ± 0.2 °C and 75.8 ± 0.1 °C for Leu-A1 and Hil-A1, respectively.

Table 2.2 compares the extent to which the thermal denaturation temperature of A1 is elevated by replacement of Leu by bulkier hydrocarbon and fluorocarbon surrogates. Replacement of Leu by Hil raises T_m by 17 °C, as compared to 10 °C for replacement of Leu by (2*S*,4*R*)-Tfl and 22 °C for replacement by Hfl. Expansion of hydrophobic side chain volume at the *d*-position of the heptad repeat constitutes an effective strategy for

stabilization of coiled-coil peptides, irrespective of the hydrocarbon or fluorocarbon character of the side chain.

We do not mean to suggest similar molecular origins for the hydrophobic properties of hydrocarbons and fluorocarbons. Although Hil and Tfl behave similarly with respect to elevation of the melting temperature of A1, other experiments suggest important differences in the behavior of water adjacent to hydrocarbon and fluorocarbon side chains. Recent studies via ultrafast spectroscopy indicate a marked slowing of water motions upon replacement of Leu by Tfl at solvent-exposed sites (23). In contrast, replacement of Leu by Hil is accompanied by increased rates of solvent reorganization. Much remains to be done to elucidate the origins of hydrophobic effects in proteins and other molecular systems.

In conclusion, we find that Hil serves as an effective surrogate for Leu with respect to protein translation in bacterial cells, and that replacement of Leu by Hil leads to substantial stabilization of recombinant coiled-coil peptides. The results reported here also highlight the value of amino acid replacement at multiple sites in peptides and proteins; replacement of a single Leu residue by Hil in T4 lysozyme has been reported to cause an increase of just 1.9 °C in the melting temperature of the protein (5). In contrast, replacement of six Leu residues in the putative coiled-coil domain of A1 raises T_m by 17 °C.

Materials and Methods

Materials. All chemicals were purchased from Sigma-Aldrich and used as received unless otherwise noted. Dry solvents were obtained from commercial suppliers and used as received. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Gold

Biotechnologies; Ni-NTA resin and spin columns were purchased from Qiagen. Sequencing grade modified trypsin was purchased from Promega; C18 Zip Tips, and Amicon Ultra-4 and Amicon Ultra-15 concentration devices were purchased from Millipore. A BCA assay kit was obtained from Pierce Protein Research Products. The *E. coli* expression strain LAM1000 outfitted with the plasmids pA1EL and pREP4 has been described previously (10).

Protein expression. For 25 mL and 200 mL scale preparations of Hil-A1 and small-scale preparation of Leu-A1, a single colony of LAM1000 transformed with pA1EL and pREP4 was used to inoculate an overnight culture of M9 minimal medium (M9 salts containing glucose (0.4% w/v), thiamine HCl (35 mg L^{-1}), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), and 20 amino acids (40 mg L^{-1})) supplemented with ampicillin (200 mg L^{-1}), and kanamycin (35 mg L^{-1}). Overnight cultures were diluted into fresh M9 medium containing all 20 amino acids and allowed to grow with agitation at 37 °C until reaching an OD_{600} of approximately 0.9–1.0. Cells were pelleted at 6000 × *g* for 7 minutes at 4 °C, washed 3 times in ice-cold sodium chloride solution (0.9% w/v) and resuspended in fresh M9 minimal medium lacking leucine. Aliquots of the resuspended cultures were supplemented with Hil (0.5 mM) or Leu (0.3 mM), shaken at 37 °C for 15 minutes, and induced by addition of IPTG (1 mM final concentration). After 3 hours, cells were harvested by centrifugation at 6000 × *g* for 7 minutes. In the case of small-scale production, cells were frozen at −80 °C either before or after addition of Qiagen buffer B (8 M urea, pH 8.0, buffered with NaH_2PO_4 (100 mM) and Tris·Cl (10 mM)). For the case of large-scale production, cells were frozen at −80 °C before addition of Qiagen buffer B.

For 200 mL scale production of Leu-A1, a single colony of LAM1000 transformed with pA1EL and pREP4 was used to inoculate an overnight culture of 2×YT medium supplemented with ampicillin (200 mg L^{-1}) and kanamycin (35 mg L^{-1}). The overnight culture was diluted 1:100 into fresh 2×YT supplemented with ampicillin and kanamycin and allowed to grow at 37 °C with shaking until reaching an OD_{600} of approximately 0.9–1.1. The culture was induced by addition of IPTG to a final concentration of 1 mM. After 3 hours, cells were harvested by centrifugation at 6000 × *g* for 7 minutes and frozen at −80 °C.

Protein purification. At small scales, the cell pellets were thawed, Qiagen buffer B was added as necessary, and the resuspended pellets were incubated at room temperature with gentle agitation for at least 1 hour. The pellets were then subjected to 20–40 minutes of sonication in an immersion sonicator. In some cases, the pellets were frozen again at −80 °C and the above procedure was repeated. The sonicated pellets were then centrifuged for 20–30 minutes at 10000 × *g*, and the supernatants were saved. The clarified lysates were then subjected to purification using Qiagen Ni-NTA spin columns according to the manufacturer's protocols (2000 Ni-NTA Spin Handbook) or using Ni-NTA agarose according to the manufacturer's protocols with slight modifications. When Ni-NTA agarose was used, Qiagen buffer C was supplemented with 50 mM imidazole, and Qiagen buffer E was supplemented with 250 mM imidazole.

At large scales, the cell pellets were resuspended in Qiagen buffer B and subjected to sonication using a microtip on a Misonix Sonicator 3000. The total sonication process time was 10 minutes, with 5 second sonication pulses and 5 second wait periods. The

pellets were then frozen again at −80 °C and subjected to a second round of sonication. The lysates were clarified by centrifugation at approximately 75000 × *g* for 10 minutes at 25 °C. The clarified lysate was then subjected to purification using Ni-NTA agarose with buffers as described above.

Peptide mass spectrometry. Small (1–3 μL) samples of Leu-A1 and Hil-A1 eluents from the purifications were diluted 50- to 100-fold into sodium bicarbonate (50 mM, pH 7.8) and digested overnight by treatment with sequencing grade modified trypsin (5 μL) at 37 °C. Portions of these samples were desalted using C18 Zip Tips according to the manufacturer's instructions with one slight modification. Prior to column equilibration in 0.1% trifluoroacetic acid, columns were wetted with a 50/50 mixture of 0.1% trifluoroacetic acid and acetonitrile. The cleaned samples were then analyzed using matrix-assisted laser desorption ionization (MALDI) mass spectrometry on an Applied Biosystems Voyager DE Pro instrument at the mass spectrometry facility of the Caltech Division of Chemistry and Chemical Engineering. All peptide masses were found to lie within the mass ranges of manufacturer-specified instrument tolerances. See figure 2.3 for typical results.

For quantitative analysis, the remainders of the trypsinized samples (not desalted) were submitted to the Protein and Peptide Mass Analytical Laboratory (PPMAL) of the Beckman Institute at Caltech for liquid chromatography tandem mass spectrometry (LC/MS/MS) analysis. The samples were separated on a 6 cm long, 100 μm diameter C18 column using an Eksigent NanoLC-2D and then immediately injected into an Applied Biosystems QStar XL tandem mass spectrometer. Time-resolved data were analyzed using Analyst QS software provided by Applied Biosystems. Amino acid replacement levels

were determined by using information contained within extracted ion currents (XIC) of trypsin-digested protein samples. Peaks corresponding to peptides globally substituted with Hil were identified along with peaks corresponding to peptides substituted at only a fraction of the Leu positions. Assuming that substitution of Hil in place of Leu is a random event, the ratio of the peak areas yields the quantitative replacement level in a given sample. A full description of the method is provided below. Where possible, fragmentation of abundant peptide ions was used to confirm the sequences of substituted peptides.

Determination of amino acid replacement levels. Liquid chromatography tandem mass spectrometry (LC/MS/MS) was used to quantitate the Hil replacement levels in purified A1 protein samples. Figure 2.4 depicts results from a typical LC/MS/MS experiment. Figure 2.4A shows the total ion currents (TIC) from the trypsinized Hil-A1, and figure 2.4B–E show the mass spectra and mass-filtered extracted ion currents (XIC) from the doubly substituted peptide SXEDEAAEXEQK (X = Hil) and the mixture of singly substituted peptides SLEDEAAEXEQK and SXEDEAAELEQK. The unsubstituted peptide SLEDEAAELEQK was not detected in this experiment. The areas of the XICs permit an estimation of the substitution rate. Assuming that replacement of leucine is a random event, the distribution of peak areas should follow the binomial distribution,

$$A\left[(1-p)^2 + 2p(1-p) + p^2\right], \quad (2.1)$$

where A is a multiplication factor equal to the sum of the areas of the unsubstituted, singly substituted, and doubly substituted peaks. The quantity p is the probability of substitution at

a given site, and the three terms in the polynomial represent non-, singly-, and doubly-substituted peptides. The polynomial also has the property

$$(1-p)^2 + 2p(1-p) + p^2 = 1. \quad (2.2)$$

The ratios of the areas of two substituted peaks are more relevant in this particular case because the total area *A* cannot be determined from the available experimental data. The experimentally accessible peak area ratio of singly to doubly substituted peaks is

$$\frac{2p(1-p)}{p^2}, \quad (2.3)$$

which can be defined as *X*. Rearranging equation (2.2) yields the identity

$$\frac{2p-2p^2}{p^2} = \frac{2p(1-p)}{p^2} = X. \quad (2.4)$$

Solving for *p* yields the two roots

$$p = 0, \quad (2.5)$$

or

$$p = \frac{2}{X+2}. \quad (2.6)$$

Substituting the peak area ratio for *X* in the nonzero root yields the substitution rate.

The above analysis was used to determine the substitution of Hil in place of Leu in the peptides SLEDEAAELEQK and GSHHHHHHGSMASGDLENEVAQLER. A simpler calculation was also made to determine the incorporation of Hil based on the peptide AEIGDLNNTSGIR, which contains one leucine residue (The value *p* can be determined from the ratio of the XIC area of the substituted peptide to the sum of the substituted and unsubstituted peptide areas). When the XIC contained multiple peaks (often the case when searching for peptides of lower abundance), the areas of the clearly distinguishable peaks (generally having areas of 10 or more) were summed in order to ensure a conservative estimate of amino acid incorporation levels. The identities of the peptides in the SLEDEAAELEQK series were confirmed by tandem mass spectrometry. Two samples of Hil-A1 produced independently were analyzed to obtain the substitution levels summarized in table 2.3.

Protein characterization. To determine protein yields, portions of Leu-A1 and Hil-A1 produced at small scale and purified using Ni-NTA agarose resin were buffer exchanged into acetate buffer (100 mM NaCl, 10 mM sodium acetate, pH 4.0) using Amicon Ultra-4 concentration devices with a 3000 Da molecular weight cutoff. The concentrations of the exchanged samples were determined with a BCA assay kit.

Circular dichroism samples were prepared by dialyzing protein samples in Qiagen buffer E (containing 250 mM imidazole) against phosphate-buffered saline, pH 7.4. Concentrations were again determined using the BCA assay, but prior to performing the assay, samples were heated to 42 °C for approximately 20 minutes in order to ensure that

samples were fully dissolved. Samples were adjusted to 10 μM by concentration with Amicon Ultra-15 concentration devices (molecular weight cutoff = 3000 Da) as necessary and dilution in fresh PBS. Circular dichroism spectroscopy was performed on an Aviv 62DS spectropolarimeter in a 1 mm path length cell. All samples were heated to 42 °C and cooled on ice prior to measurement. Each experiment consisted of a wavelength scan performed at 1 °C followed by a temperature scan from 0 °C to 94.5 °C in 1.5 °C intervals with signal monitoring at 222 nm. During the temperature scans, the sample was allowed to equilibrate for one minute prior to performing readings at each temperature step. All data was referenced to background scans of PBS buffer acquired under identical conditions. Wavelength scans were analyzed using K2D2 (20), and temperature scans were analyzed using a Matlab implementation of a model for coiled-coil unfolding described previously (22).

ATP-PP_i exchange assays. *E. coli* LeuRS was expressed and purified as previously described and its concentration was determined from its absorbance at 280 nm under native conditions (24). Assays were run at room temperature in buffer containing HEPES (30 mM, pH 6.8), $MgCl_2$ (10 mM), dithiothreitol (1 mM), ATP (2 mM), and [^{32}P]-PP_i (2 mM, 3 μCi in 200 μL rxn volume). Activation of leucine was performed in solutions containing LeuRS (75 nM) and varying Leu concentrations (1.6 to 50 μM), while activation of homoisoleucine was measured using a higher concentration of LeuRS (300 nM) and varying Hil concentrations (16 to 500 μM). Aliquots (30 μL) were taken every 2 (Leu) or 6 (Hil) minutes and quenched in a suspension of activated charcoal (3% w/v) containing $HClO_4$ (7% w/v) and inorganic pyrophosphate (200 mM). The

charcoal was washed twice in a solution of $HClO_4$ (0.5% w/v) and inorganic pyrophosphate (10 mM), and added to 20 mL scintillation vials. Scintillation fluid (5 mL) was added to each vial and samples were counted using a Beckman Coulter liquid scintillation counter. Data were fitted with nonlinear regression using the program Igor.

Synthesis of homoisoleucine (2-amino-4-methylhexanoic acid, CAS # 3570-21-6). Silica chromatography was performed using 230-400 mesh silica gel 60 (EMD). TLC was run on Baker-flex silica gel IB-F plates, R_fs are reported under the same solvent conditions as columns unless otherwise noted. TLC was examined under UV light for fluorescent compounds or alternatively stained with $KMnO_4$, ceric ammonium molybdenate or *p*-anisaldehyde. NMR spectra were recorded on Varian spectrometers (300 MHz for ^1H) and processed with NUTS NMR software. NMR spectra were referenced to internal standards; proton spectra were referenced to tetramethylsilane and carbon spectra were referenced to solvent peaks. FAB mass spectrometry was performed at the Caltech Division of Chemistry and Chemical Engineering Mass Spectrometry Facility.

The following procedure was adapted from O'Donnell and Eckrich (25): aminoacetonitrile benzophenone imine (1.21 g, 5.5 mmol), benzyltriethylammonium chloride (0.1 g, 0.4 mmol), 50% aq NaOH (0.75 mL, 14 mmol) and toluene (1 mL) were combined in a 10 mL round bottom flask which contained a magnetic stir bar. The flask was chilled in an ice-water bath and stirred vigorously (~1200 rpm). (*S*)-2-methyl bromobutane (1.15 g, 0.94 mL, 7.6 mmol, 1.4 eq) was added portion-wise via syringe over 1 hour to the stirred solution. The reaction mixture was stirred for an additional 2 hours at 0 °C and then allowed to come to room temperature. Stirring was continued at room

temperature for 96 hours. The reaction mixture was transferred to a separatory funnel and diluted with H_2O (20 mL) and CH_2Cl_2 (40 mL). The aqueous layer was washed 3 times with CH_2Cl_2 (10 mL × 3) and the combined organic layers were washed 3 times with H_2O (10 mL × 3) and once with saturated NaCl (10 mL). The organic layers were dried over Na_2SO_4 and filtered, and the solvent was removed to yield a yellow oil (1.80 g). Purification by silica chromatography (eluting with 6% ethyl acetate in hexanes; R_f 0.18 in 10% ethyl acetate/hexanes) gave a yellow oil (1.40g, 88%); a mixture of diastereomers. ^{1}H NMR (300 MHz, $CDCl_3$) δ 0.73 (dd, 3H J = 4.5, 6.5 Hz), 0.82 (dd, 3H J = 7.5, 14.0 Hz), 1.00-1.37(m, 2H), 1.41-1.81(m, 2H), 1.82-2.14 (m, 1H), 4.27 (dd, 0.5H, J = 6.3, 8.0 Hz), 4.31 (dd, 0.5H, J = 6.3, 8.1 Hz), 7.15-7.26 (m, 2H), 7.27-7.37 (m, 2H), 7.37-7.58 (m, 4H), 7.58-7.82 (m, 2H) ^{13}C NMR (75 MHz, $CDCl_3$) δ 10.724, 10.891, 18.416, 18.699, 28.865, 28.923, 30.680, 31.086, 41.266, 41.347, 51.037, 51.606, 119.747, 119.899, 127.108, 127.238, 128.039, 128.789, 129.205, 130.933, 135.007, 135.051, 138.243, 138.294, 172.259, 172.582. FAB MS calculated for $C_{20}H_{22}N_2$ (M+) 290.1783, observed 290.1787.

The following procedure was adapted from Dorizon et al. (26): HCl (1 M, 14 mL) was added dropwise to a solution of alkylated aminoacetonitrile benzophenone imine (290 mg, 1.00 mmol) in diethyl ether (7 mL). The mixture was stirred vigorously at room temperature for 24 hours. The aqueous phase was extracted twice with diethyl ether and evaporated to give 1-amino-1-cyano-3-methylpentane hydrochloride (180 mg, ~100%). This material was dissolved in 6 M HCl (3 mL) and heated to reflux for 72 h. The solution was evaporated under reduced pressure to give an off-white solid (168 mg, 93%). ^{1}H NMR (300 MHz, CD_3OD) δ 0.929 (t, 3 H J = 7.5 Hz), 0.936 (t, 3 H J = 7.5 Hz), 0.979 (d, 3 H J = 3.6), 0.999 (d, 3 H J = 3.4), 1.148-1.548 (m, 2 H), 1.554-2.010 (m, 3 H), 3.979 (m, 1H);

^{13}C NMR (75 MHz, CD_3OD) δ 10.036, 10.265, 17.665, 18.004, 28.846, 29.254, 30.649, 30.659, 37.601, 37.656, 51.117, 51.133, 171.117, 171.244; FAB MS calculated for $C_7H_{16}NO_2$ (M+H) 146.1176, observed 146.1205.

Acknowledgements

We thank Shelly Tzlil for assistance in implementing the model for thermal unfolding of coiled-coil peptides in Matlab. This research was supported by NIH grant GM62523. J. A. V. was supported in part by an N. D. S. E. G. Fellowship; J. D. F. was supported in part by an NIH Postdoctoral Fellowship.

References

1. Liu CC & Schultz PG (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413-444.
2. 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.
3. Budisa N (2006) *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins* (Wiley-VCH, Weinheim, Germany).
4. Connor RE & Tirrell DA (2007) Non-canonical amino acids in protein polymer design. *Polymer Reviews* **47,** 9-28.
5. Mendel D*, et al.* (1992) Probing protein stability with unnatural amino acids. *Science* **256,** 1798-1802.

6. 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.

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

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

9. Tang Y*, et al.* (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. *Angew. Chem. Int. Ed.* **40,** 1494-1496.

10. 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.

11. Son S, Tanrikulu IC, & Tirrell DA (2006) Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues. *ChemBioChem* **7,** 1251-1257.

12. Montclare JK, Son S, Clark GA, Kumar K, & Tirrell DA (2009) Biosynthesis and stability of coiled-coil peptides containing (2S,4R)-5,5,5-trifluoroleucine and (2S,4S)-5,5,5-trifluoroleucine. *ChemBioChem* **10,** 84-86.

13. Petka WA, Harden JL, McGrath KP, Wirtz D, & Tirrell DA (1998) Reversible hydrogels from self-assembling artificial proteins. *Science* **281,** 389-392.

14. Tang Y*, et al.* (2001) Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine. *Biochemistry* **40,** 2790-2796.

15. Bilgicer B, Fichera A, & Kumar K (2001) A coiled coil with a fluorous core. *J. Am. Chem. Soc.* **123,** 4393-4399.

16. Lee KH, Lee HY, Slutsky MM, Anderson JT, & Marsh ENG (2004) Fluorous effect in proteins: De novo design and characterization of a four-alpha-helix bundle protein containing hexafluoroleucine. *Biochemistry* **43,** 16277-16284.

17. Gao JM, Qiao S, & Whitesides GM (1995) Increasing binding constants of ligands to carbonic-anhydrase by using greasy tails. *J. Med. Chem.* **38,** 2292-2301.

18. Buer BC, de la Salud-Bea R, Hashimi HMA, & Marsh ENG (2009) Engineering protein stability and specificity using fluorous amino acids: the importance of packing effects. *Biochemistry* **48,** 10810-10817.

19. Molecular surface areas of Leu, Hil, (*S*,*S*)-Tfl, and (*S*,*R*)-Tfl were determined using Spartan '08 (Wavefunction, Inc. Irvine, CA). Structures of amino acids were minimized using Hartree-Fock calculations with a 6-31G* basis in vacuum, followed by determination of the molecular surface areas.

20. Perez-Iratxeta C & Andrade-Navarro MA (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol.* **8**.

21. Tang et al. confirmed the predominance of dimers in dilute solutions of A1 containing Leu or fluorinated analogs of Leu at neutral pH; the oligomerization state at low (10 μM) concentrations was found to be insensitive to fluorination (9, 10).

22. Schneider JP, Lear JD, & DeGrado WF (1997) A designed buried salt bridge in a heterodimeric coiled coil. *J. Am. Chem. Soc.* **119,** 5742-5743.

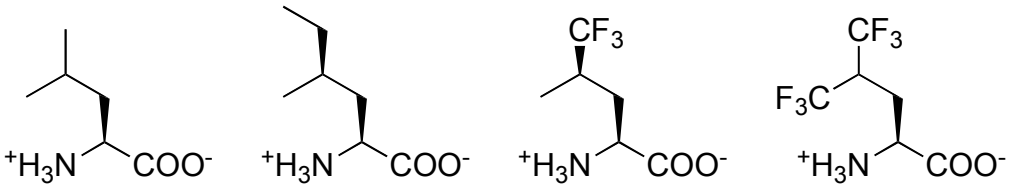
23. Kwon OH*, et al.* (2010) Hydration dynamics at fluorinated protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17101-17106.

24. Tang Y & Tirrell DA (2002) Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **41,** 10635-10645.

25. Odonnell MJ & Eckrich TM (1978) Synthesis of amino acid derivatives by catalytic phase transfer alkylations. *Tetrahedron Lett.* 4625-4628.

26. Dorizon P*, et al.* (1999) Stereoselective synthesis of highly functionalized cyclopropanes. Application to the asymmetric synthesis of (1S,2S)-2,3-methanoamino acids. *J. Org. Chem.* **64,** 4712-4724.

27. Boniecki MT, Vu MT, Betha AK, & Martinis SA (2008) CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 19223-19228.



1 2 3 4

Scheme 2.1. Amino acids used in study. **1**, leucine (Leu). **2**, (2*S*,4*S*)-2-amino-4-methylhexanoic acid (homoisoleucine, Hil). **3**, (2*S*,4*R*)-trifluoroleucine (Tfl). **4**, hexafluoroleucine (Hfl).

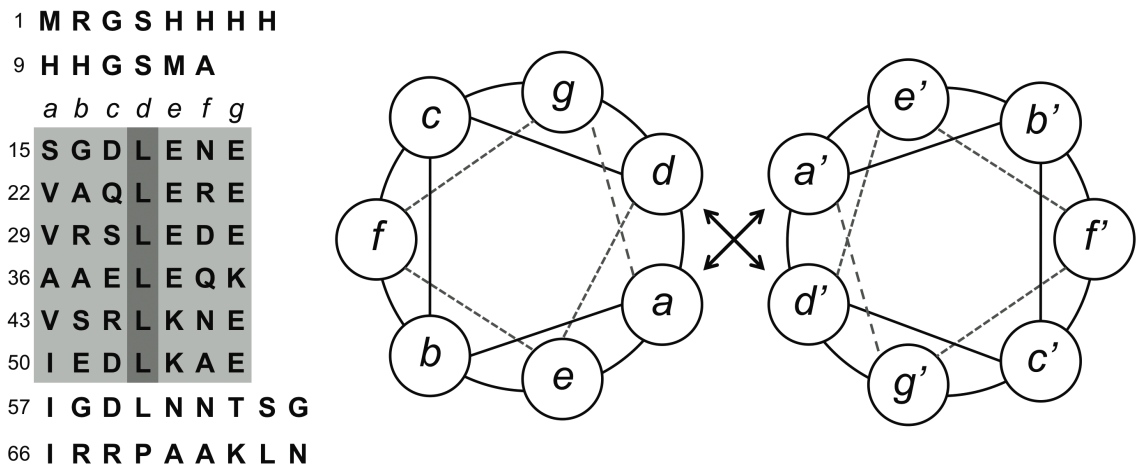


Figure 2.1. A1 peptide sequence and helical wheel representation of A1 homodimers. The amino acids that comprise the putative heptad repeats are highlighted in gray, with additional emphasis on the leucine residues at the *d* positions.

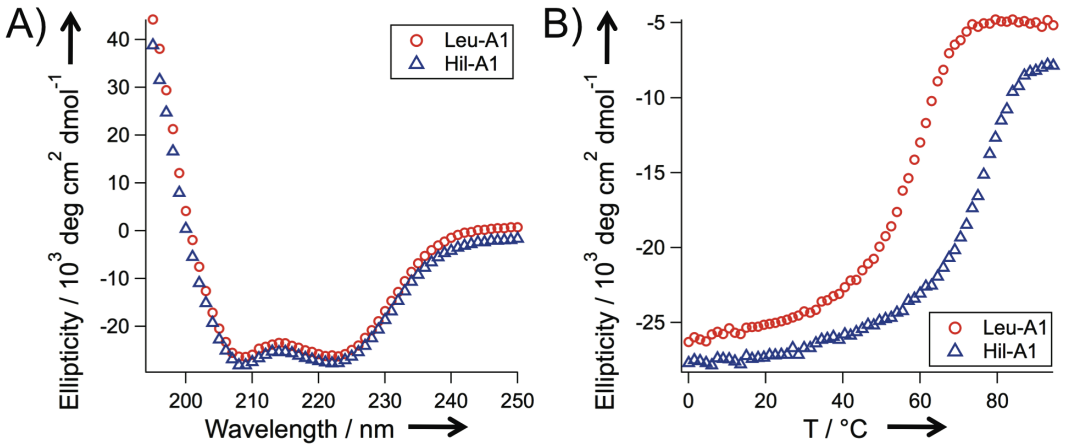


Figure 2.2. Circular dichroism spectra of Leu-A1 and Hil-A1. A) Wavelength scans performed at 1 °C. B) Ellipticity at 222 nm as a function of temperature. All experiments were performed with 10 μM peptide in PBS, pH 7.4.

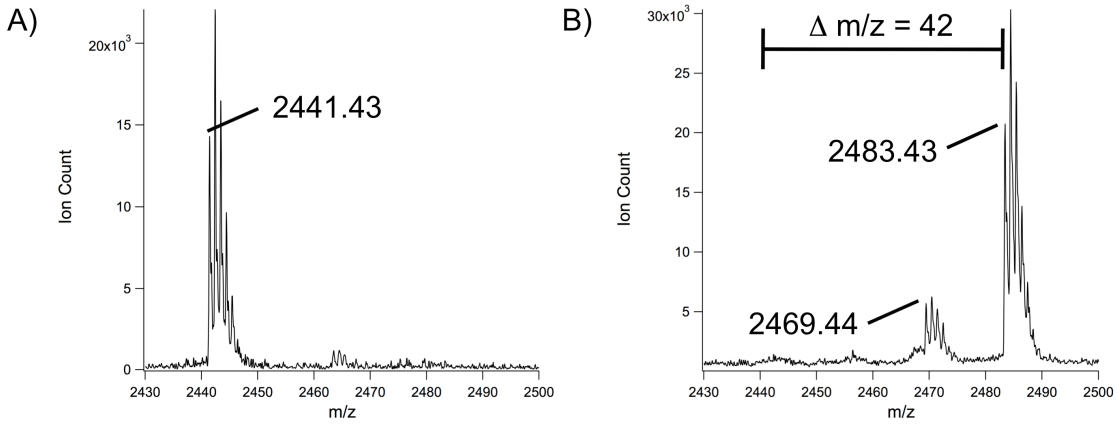


Figure 2.3. MALDI spectra of tryptic fragments of A) Leu-A1 and B) Hil-A1. The portions of the spectra shown encompass the m/z region in which the peptide LKNEIEDLKAEIGDLNNTSGIR appears with or without Hil substitution. In the Leu-A1 sample, the major peak appears at 2441.43 Da (calculated mass: 2442.28 Da). In the Hil-A1 sample, the largest peak appears at 2483.43 Da (calculated mass: 2484.32 Da), 42 mass units away from the unsubstituted peak, corresponding to complete replacement of Leu by Hil. A smaller peak at 2469.44 Da (calculated mass: 2470.31 Da) indicates the presence of some peptides that contain only two out of three Hil substitutions.

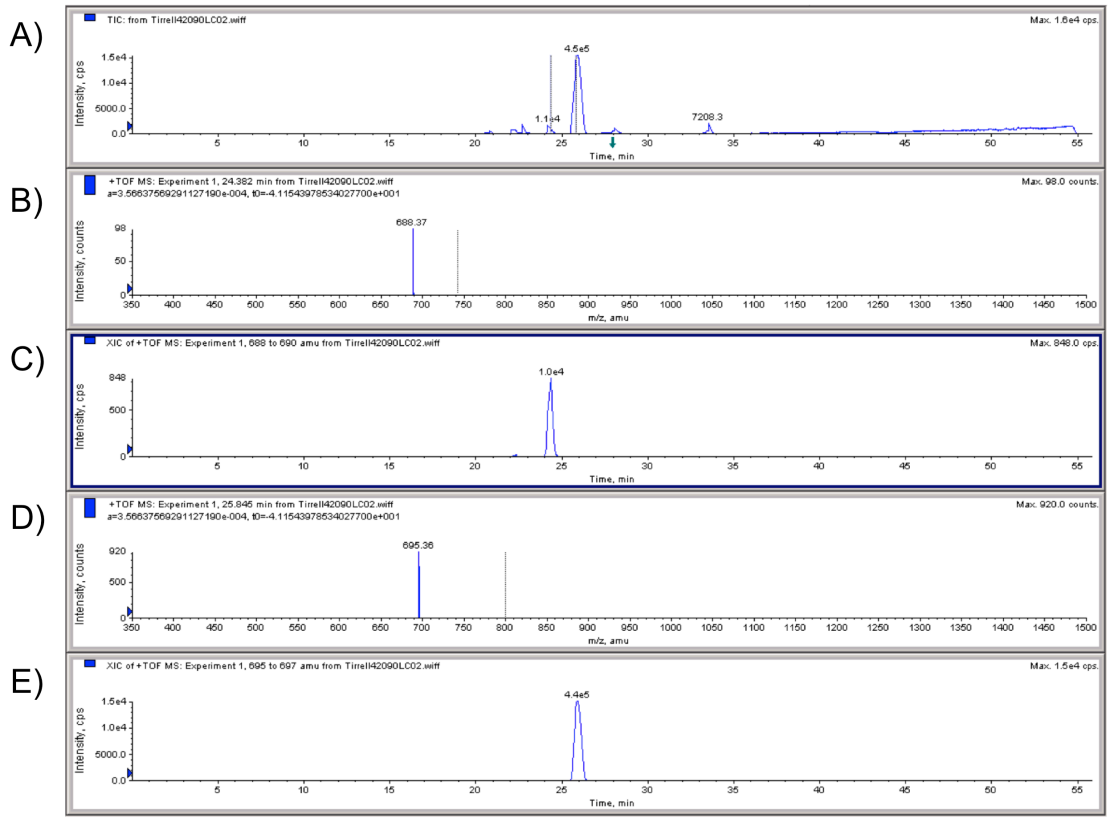


Figure 2.4. LC/MS results obtained on trypsinized sample of Hil-A1 (Sample A in table 2.3). A) Total ion current (TIC) of sample. B) Mass spectrum at time = 24.382 min, revealing a doubly charged ion having the mass corresponding to the masses of the singly substituted peptides SXEDEAAELEQK and SLEDEAAEXEQK (X = Hil). Calculated doubly charged ion 688.33 Da, observed 688.37 Da. C) Extracted ion current (XIC) of masses ranging from 688 to 690 amu. The mass filtering reveals a single, well-defined peak of integrated area 1.0×10^4 arbitrary units. D) Mass spectrum at time = 25.845 min, revealing a doubly charged peak having a mass corresponding to the mass of the doubly substituted peptide SXEDEAAEXEQK. Calculated doubly charged ion 695.34 Da, observed 695.36 Da. E) XIC of masses ranging from 695 to 697 amu. Again, a single, well-defined peak is observed and has an integrated area of 4.4×10^5 arbitrary units. The sequences of these peptides were confirmed with LC/MS/MS fragmentation.

Table 2.1. Kinetic parameters for activation of Leu and Hil by LeuRS

Substrate[a]	k_{cat} [s^{-1}]	K_m [μM]	k_{cat}/K_m [rel]
Leu[b]	15.1 ± 2.2	3.7 ± 1.9	1
Hil	0.4 ± 0.1	77 ± 65	1/690

[a] Leu was used as the L-isomer; Hil as a mixture of the D- and L- isomers. The concentrations of Hil reported here are those of the L-isomer. Kinetic parameters are reported as averages determined from three independent experiments with errors reported as averages of the 95% confidence intervals. [b] Parameters determined for activation of Leu are consistent with previous reports; the value of k_{cat} measured in this work is within the range of reported values, while the value of K_m reported here is lower than literature values by a factor of 2–10 (10, 12, 24, 27).

Table 2.2. Stabilization of A1 by replacement of Leu with noncanonical amino acids

Amino acid at *d* position	Leu	(2*S*,4*R*)-Tfl	Hil	Hfl[b]
ΔT_m[a]	0	10 (ref 12)	17	22 (ref 10)

[a] Increase in melting temperature (as compared to Leu-A1) determined from CD spectroscopy of 10 μM solutions of peptide in PBS, pH 7.4. [b] 74% replacement of Leu.

Table 2.3. Incorporation levels of Hil in A1 samples determined from multiple series of substituted peptides

Peptide series	Sample A	Sample B
SLEDEAAELEQK	98.9%	97.1%
GSHHHHHHGSMASGDLENEVAQLER	99.5%	99.0%
AEIGDLNNTSGIR	98.5%	ND[a]
[a] Not determined due to insufficient signal of unsubstituted ion.		

CHAPTER 3

Hydration Dynamics at Fluorinated Protein Surfaces

This chapter first appeared as an article in *Proceedings of the National Academy of Sciences, USA*: Oh-Hoon Kwon, Tae Hyeon Yoo, Christina M. Othon, James A. Van Deventer, David A. Tirrell, and Ahmed H. Zewail. *Proc. Natl. Acad. Sci. USA* **2010,** 107, (40) 17101-17106.

Abstract

Water-protein interactions dictate many processes crucial to protein function including folding, dynamics, interactions with other biomolecules, and enzymatic catalysis. Here we examine the effect of surface fluorination on water-protein interactions. Modification of designed coiled-coil proteins by incorporation of 5,5,5-trifluoroleucine or (4*S*)-2-amino-4-methylhexanoic acid enables systematic examination of the effects of side-chain volume and fluorination on solvation dynamics. Using ultrafast fluorescence spectroscopy, we find that fluorinated side chains exert electrostatic drag on neighboring water molecules, slowing water motion at the protein surface.

Introduction

The past decade has witnessed substantial expansion in the number and diversity of noncanonical amino acids that can be incorporated into recombinant proteins expressed in bacterial cells (1-3). Fluorinated amino acids have drawn special attention (4-16) because of the unusual solubility properties of fluorinated hydrocarbons. Several independent studies have shown that fluorination of coiled-coil and helix-bundle proteins leads to enhanced stability with respect to thermal or chemical denaturation (6-12), an effect attributed to the hyperhydrophobic and fluorophilic character of fluorinated amino acid side chains.

Although both classes of compounds are hydrophobic, hydrocarbons and fluorocarbons differ in important ways (17-22). The high electronegativity of fluorine renders the C-F bond both strongly polar and weakly polarizable (17, 21, 22). The dipole associated with the C-F bond exerts strong inductive effects on neighboring bonds (23)

and can form reasonably strong electrostatic interactions with ionic or polar groups when the two moieties are appropriately positioned. The hydrophobic character of fluorinated compounds has been described as “polar hydrophobicity (17),” and is believed to play important roles in organic and medicinal chemistry. Furthermore, the C-F bond is significantly longer than the C-H bond, and the calculated volume of the trifluoromethyl group is about twice that of a methyl group (20). The studies described here constitute an attempt to understand more fully the interaction of water with fluorinated molecular surfaces, and to provide a sound basis for the use of fluorinated amino acids in the engineering of proteins with unique and useful physical properties.

The hydration layer adjacent to protein surfaces exhibits properties different from those of bulk water; the more rigid and denser structure of the hydration layer plays a crucial role in protein structure, folding, dynamics, and function (24-26). Elucidation of the dynamic features of this region, on the timescales of atomic and molecular motion, is essential in understanding protein hydration. In the past decade, the knowledge of hydration on protein surfaces has been extensively expanded by studying the dynamic properties of biological water for various proteins containing tryptophan (Trp) or synthetic fluorescent amino acids as local probes; the results have revealed multicomponent relaxation dynamics spanning a wide range of timescales (25, 27-29). The nature of the protein hydration layer can be affected not only by the topographic and electrostatic properties of the protein surface (24), but also by the physical and chemical properties of individual surface-exposed residues (27, 30). In view of the unique properties of the C-F bond and of fluorocarbon–water interfaces (23, 31), we anticipated that fluorinated amino acid side chains might exhibit unusual hydration behavior. Here

we report studies of local hydration dynamics at fluorinated protein surfaces by monitoring the time-dependent fluorescence Stokes shifts of surface-exposed Trp residues in coiled-coil proteins with 5,5,5-trifluoroleucine (Tfl, **1**; scheme 3.1) residues adjacent to the probe. The results are compared to the hydration dynamics at hydrogenated protein surfaces with Leu (**2**) or (4*S*)-2-amino-4-methylhexanoic acid (homoisoleucine, Hil, **3**) adjacent to the Trp probe. Hil has approximately the same volume as Tfl (20, 21), and although the shapes of the residues differ, the nearly identical side-chain volumes of Tfl and Hil allow us to differentiate changes due to fluorination from those that result from the increase in side-chain volume that accompanies replacement of Leu (scheme 3.1).

Results

Coiled-coil protein system. The coiled-coil protein A1 (figure 3.1 A and B) was used as a model system to examine the effects of fluorinated amino acids on local hydration dynamics. The primary structure of A1 contains six copies of a heptad repeat $(abcdefg)_n$, where positions *a* and *d* are occupied by hydrophobic amino acids. Self-association of the peptide juxtaposes the *a* and *d* positions and results in the formation of a hydrophobic core. Fluorinated Leu analogues have previously been incorporated into the *d* positions of A1; the resulting proteins exhibited improved resistance to thermal and chemical denaturation with minimal differences in secondary structure (9, 11, 12). In this work, the surface-exposed Asp residue at the *f* position of the third heptad (position 34) was replaced by Trp, which serves as a fluorescence probe (figure 3.1C). The Trp variant of A1 was designated A1m. In order to examine the effects of fluorinated analogues on the local hydration

dynamics, a Leu codon was introduced at one of two positions within A1m. Mutation of a serine residue at the *c* position of the third heptad (position 31) yielded a variant of A1 designated S31L (figure 3.1D), while replacement of an alanine residue at position *b* of the fourth heptad (position 37) gave the A1 variant A37L (figure 3.1E). Each protein was expressed in Tfl, Leu, and Hil form, yielding a total of nine different proteins that were examined in detail (see (32) for nomenclature).

Characterization of global structure. Analysis of each protein showed that the overall structural properties of the molecules were generally insensitive to genetic mutations and incorporation of noncanonical amino acids. Circular dichroism spectroscopy indicated that all nine proteins were helical, as determined from the molar ellipticity at 222 nanometers (figure 3.2) (33); an analysis with K2D2 software showed that the helicities of individual proteins range between approximately 40% and 48% (34). These results are consistent with the design of the A1 protein (35), in which approximately half of the amino acids are located within the heptad repeats expected to form α-helical secondary structure. The oligomerization states of the protein samples were determined by sedimentation velocity analysis (figure 3.3). Although A1 forms dimers and tetramers at neutral pH (11), the variants examined in this study form trimers or hexamers under mildly acidic conditions (pH 4). We suggest that protonation of Glu side chains at the *e* and *g* positions (figure 3.1A) of the proteins decreases the density of negative charges adjacent to the hydrophobic core and promotes formation of larger helical aggregates at pH 4. A1m, in which the single Trp residue occupies a surface-exposed position, is predominantly trimeric in Leu-, Tfl-, and Hil-forms, with a small fraction of hexamers

(see figure 3.3). The majority of the S31L samples are present as hexamers, while the A37L samples appear to contain mixtures of trimers and hexamers.

Characterization of local structure. The steady-state fluorescence emission spectrum of Trp depends on the extent of exposure of the Trp side chain to water (36). All nine protein samples showed emission maxima between 349 and 352 nm, close to that of free Trp at 353 nm (table 3.1 and figure 3.4). These observations indicate that the Trp residues are exposed to the aqueous environment (consistent with the original design), and not involved in oligomerization of the proteins. In addition, the steady-state UV-visible absorption and steady-state fluorescence emission spectra of each mutant containing Leu were nearly identical to the spectra of the corresponding mutant when it contained Tfl or Hil (figure 3.4), further confirming that perturbation of the protein structure upon replacement of Leu by Tfl or Hil was minimal.

The mobility of the probe residue was explored in each protein by measuring fs-resolved depolarization dynamics (figure 3.5). The anisotropic dynamics were found to consist of three components: ultrafast (≤ 500 fs), intermediate (20–80 ps), and slow (≥2 ns) decays. The ultrafast decays are attributed to fast internal conversion between the first two excited singlet states (^1L_a and ^1L_b) of Trp, the intermediate decays to local wobbling motions of Trp, and the slow decays to tumbling motions of the proteins (28, 37). Similar values for the wobbling motions (ϕ_{Trp}) and their cone semiangles (θ) were observed for each series of S31L and A37L proteins (table 3.1); see material and methods for details.

Both mutation of residues around Trp and fluorination of the protein hydrophobic core can affect the environment of the probe and change the protein structure and/or the

dynamic properties of the hydration layer. In many cases these properties are related to one another. The minimal change in the steady-state fluorescence spectrum caused by replacement of Leu by Tfl or Hil suggests similar features of the hydration region probed by Trp (e.g., the effective number of water molecules in the hydration shell). In addition, the similarity of the Trp wobbling angle of the Leu-, Tfl-, and Hil-forms of the proteins suggests similar organization and flexibility of neighboring residues around the probe (28). All these features make it possible to compare the dynamic properties of protein hydration for most of the proteins in the A1m, S31L, and A37L proteins containing Leu, Tfl, and Hil. For A1m-H, we note that the wobbling angle of Trp was found to be 33°, which is significantly higher than the 17°–21° wobbling angles determined for all of the other proteins. This result indicates that the organization of local residues or the flexibility of the local Trp environment in A1m-H differs from that in the other proteins, despite the lack of global structural changes observed by circular dichroism or sedimentation velocity measurements. The abnormal behavior of the A1m-H variant is also observed in the fluorescence lifetime measurements. Every protein except A1m-H displayed a short-lifetime component of a few hundred picoseconds, present at all wavelengths. These types of quenching processes have been attributed to Trp interactions with nearby charged residues (38-40), and the absence of such a feature in A1m-H again indicates that this protein has a local structure different from those of the other eight proteins. The perturbation of local structure and, thus, local solvent exposure can result in different hydration dynamics, making it unreliable to compare the dynamics of A1m-H to those of the other A1m proteins. Accordingly, the dynamics obtained for A1m-H were not used in the analysis that follows. Small shifts in the fluorescence emission maximum

and Trp wobbling angle were observed for A37L-L as compared to A37L-T and A37L-H (see table 3.1 and figure 3.4). These differences may be significant enough to alter the local environment surrounding the Trp probe, potentially complicating assignment of changes observed in the hydration dynamics to a particular effect (e.g., changes in an amino acid close to Trp). Despite these concerns, the dynamics results for A37L-L remain consistent with the conclusions of the paper (see below).

Our stringent standards for comparison of hydration dynamics between modified proteins require that there be (i) no global change in protein structure as measured by circular dichroism spectroscopy and sedimentation velocity measurements; (ii) no change in solvent exposure as measured by steady-state fluorescence maximum (± 1 nm); and (iii) no change in local protein structure or flexibility as measured by fluorescence anisotropy (± 1°). Seven of the nine proteins prepared in this study met all of these criteria, and an additional protein, A37L-L, displayed changes just outside the margin of error. Only one protein, A1m-H, showed changes significant enough to require us to disregard the hydration measurements observed. Given the subtle effects of the chemical environment on hydration dynamics, we will compare hydration results only within protein families. Thus, our strongest conclusions will be drawn from observations made on the S31L protein variants, and the data for A1m and A37L will be used as corroborating evidence.

Ultrafast hydration dynamics. To investigate hydration dynamics at the protein surfaces, we utilized a methodology developed by Zhong and coworkers for the reconstruction of femtosecond-resolved fluorescence spectra (28, 41). As an example,

figure 3.6A shows several representative femtosecond-resolved fluorescence transients recorded for A1m-T. The overall decay dynamics is retarded compared to that of free Trp in buffer solution. Details of the results for all the protein samples are presented in table 3.1. The hydration dynamics of the proteins were well represented by triple-exponential decays with distinctive timescales of 0.2–0.8, 1.4–6.1, and 10–61 picoseconds. Relaxation occurring on a time scale of a few hundred femtoseconds to several picoseconds is attributed to fast librational/rotational motions of bulk-type and local water molecules around Trp. Observation of the fs component suggests that the Trp probe in the test proteins is neither crowded by neighboring residues nor protected from exposure to water (27, 28, 42). The slowest phase of hydration dynamics (on the timescale of tens of picoseconds) is collective water network rearrangement coupled to protein fluctuation dynamics (27, 42-44).

Several key features of the results (figure 3.7 and table 3.1) are summarized as follows. First, S31L-T and A37L-T, in which Tfl lies close to Trp as well as in the hydrophobic core, showed slower hydration dynamics than their Hil and Leu counterparts, indicating that the fluorinated surface of the protein slows down the hydration dynamics. For S31L-T, the timescales of local and collective water motions (τ_2 and τ_3, respectively) are increased by 2–5 times (3.0 and 48 ps) from those of S31L-H (1.4 and 10 ps). The overall solvation of S31L-T is slower than that of S31L-L as well. However, this difference is manifested as an increase in the contribution of the τ_3 component to the overall solvation of Trp. The 70% increase of the relaxation energy (333 cm^{-1}) of the slowest component (E_3) compared to that of the nonfluorinated S31L-L (194 cm^{-1}) is an indication of the dramatic slowing of hydration dynamics near the

fluorinated surface. For A37L-T, τ_2 and τ_3 are also retarded to a greater extent (3–5 times slower) than those for A37L-H. These results suggest that replacement of Leu by Tfl increases the residence time of water molecules near the Trp probe and/or the number of water molecules influenced by the amino acid side chain.

Second, S31L-H and A37L-H showed similar or even faster hydration than their Leu counterparts. This result indicates that in the comparison between hydrocarbon residues, increasing the hydrophobic surface area results in faster motion of water molecules around the residue. It should be noted that, for the S31L series, hydration is greatly accelerated when Leu is replaced with Hil. This pronounced hydrophobic effect (due to the increase in the size of the residue) on the hydration is counteracted by fluorination of leucine, resulting in slowing dynamics for S31L-T. On the other hand, for A37L, increasing the size of the hydrophobic surface does not appear to affect the hydration dynamics as greatly. Therefore, the retardation of the dynamics upon fluorination is much more pronounced for A37L-T than for the corresponding fluorination of S31L. Finally, A1m-L and A1m-T, which differ from one another only in the nature of their hydrophobic cores, exhibited almost identical hydration dynamics. Fluorination of the hydrophobic core of a helix-bundle protein can affect protein dynamics (45). However, the nearly identical hydration dynamics for A1m-L and A1m-T spanning a few hundred picoseconds indicates that modification of the hydrophobic core of A1m does not affect protein motions that are coupled to local hydration dynamics at the surface of the protein on the timescales examined here (46).

Discussion

Protein surfaces are heterogeneous, consisting of polar, hydrophilic, and hydrophobic residues, and it is intriguing to consider how the heterogeneous surface chemistry affects the behavior of water molecules in the protein hydration layer. Head-Gordon and coworkers have reported heterogeneous water dynamics in the first hydration shell of model peptides (*N*-acetyl-leucinemethylamide and *N*-acetyl-glycinemethylamide), with faster water motions near the hydrophobic side chains than near the hydrophilic backbone (47, 48). Similar results have been reported for molecular dynamics simulation studies of a folded β-hairpin peptide (30). In addition, Qiu et al. showed that mutation of charged or polar residues of the enzyme staphylococcal nuclease into more hydrophobic residues (Ala), resulted in faster hydration dynamics; this result was attributed to the lack of strong interaction between the charges (or dipoles) of the mutated protein and the surrounding water (27). This observation can be understood in that the elimination of specific interactions between hydrophobic residues and water molecules causes a lower number of hydrogen bonds between water and a hydrophobic surface compared to those near a hydrophilic surface, thus allowing water molecules to reorient more freely. Computational studies have suggested that water layers adjacent to extended hydrophobic surfaces of low curvature are of lower density than those around hydrophilic and small hydrophobic molecules, and are dynamic rather than static (49-56). X-ray reflectivity experiments indicate submonolayer water depletion at hydrocarbon and fluorocarbon surfaces (57). Our observation of accelerated hydration dynamics around larger Hil residues as compared to the smaller Leu is consistent with these experimental and computational results, supporting the idea that water molecules neighboring hydrophobic

side chains in the hydration layer of proteins are more dynamic than those around polar or hydrophilic residues.

Even though a simple comparison suggests that Tfl should be more hydrophobic than Leu by virtue of its larger surface area, introduction of a Tfl residue adjacent to the Trp probe caused retardation of the local hydration dynamics, in contrast to the results obtained when Leu was replaced with Hil (figure 3.7). These results suggest that hydration dynamics around fluorinated amino acid side chains cannot be explained exclusively by the increase in residue size. The C-F bond is assumed not to be involved in hydrogen bonding with liquid water, largely because of its low polarizability (17). However, replacement of Leu by Tfl introduces a strong dipole at the fluorinated protein surface. Lee and coworkers have shown that introduction of CF_3 groups reduces the contact angle of water on self-assembled alkanethiol monolayers (23), an effect that they attribute to dipolar interactions. Our results suggest that such dipolar interactions can also slow water motions at fluorinated molecular surfaces.

Fluorinated compounds are more hydrophobic than hydrogenated compounds of equal carbon number (4, 5, 17-21), and the increase in hydrophobic character of fluorocarbons has been ascribed to their increased molecular size (18, 58). This interpretation appears to be consistent with the observation that the melting temperature of A1-Tfl is 13 °C higher than that of A1-Leu (11), while T_m for A1-Hil is increased by 17 °C in comparison to A1-Leu. However, the results reported here clearly indicate that the chemical nature of the protein surface dictates the dynamics of solvent-protein interactions, and that size effects alone cannot explain the altered solvation dynamics observed at fluorinated protein surfaces.

Conclusions

The results reported here show that fluorinated amino acids influence hydration dynamics at protein surfaces in a manner quite different from their hydrogenated counterparts. In general, water-protein interactions dictate many processes crucial to protein function including folding, dynamic motions, interactions with other biomolecules, and enzymatic catalysis (26). The slower timescales of hydration dynamics observed near fluorinated residues in proteins suggest that some of the water-mediated processes processes listed above may be changed upon fluorination. Tailoring the dynamics of protein-water interactions by the introduction of fluorinated residues may yield proteins with functional properties, such as binding, molecular recognition, or catalytic activities, that cannot be achieved with the canonical amino acids. Understanding hydration dynamics at fluorinated molecular surfaces is a critical step toward exploiting the properties of fluorine in biological systems.

Materials and Methods

Summary of protein expression and characterization. A1 variants A1m, S31L, and A37L were expressed in 2×YT medium (which contains Leu) to yield proteins A1m-L, S31L-L, and A37L-L, respectively, in Leu-free M9 minimal medium (12.8 g/L $Na_2HPO_4 \cdot 7H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl) supplemented with 19 canonical amino acids plus Tfl to give A1m-T, S31L-T, and A37L-T, and in Leu-free M9 medium supplemented with 19 canonical amino acids plus Hil to give A1m-H, S31L-H, and A37L-H. The proteins were purified under denaturing conditions and dialyzed against 10 mM acetate (pH 4)/100 mM NaCl. The extent of replacement of Leu by Tfl in A1m-T,

S31L-T, and A37L-T, was determined by amino acid analysis to be 90%–91%. Leu replacement by Hil in A1m-H, S31L-H, and A37L-H was analyzed by liquid chromatography tandem mass spectrometry and determined to be at least 90% (figure 3.8). See below for further details of expression, purification, and incorporation analysis.

Summary of steady-state measurements. Circular dichroism spectra were recorded on an Aviv 62DS spectropolarimeter (Lakewood, NJ). Absorption spectra were collected using a Cary 500 UV-Vis spectrophotometer and a 0.05 mm path length cuvette. Steady-state fluorescence emission spectra were measured using a FluoroMax-2 fluorimeter (ISA-Spex).

Summary of time-resolved fluorescence measurements. The experimental apparatus for time-resolved measurements are detailed below. All fluorescence spectra and transients were obtained by the excitation of samples (~550 μM) at 295 nanometers. The lifetime components were obtained by global analysis of fluorescence transients collected using a time-correlated single photon counting spectrometer. All transients show additional multiple-exponential decay (at the blue side) and rise (at the red side) with time constants spanning from a few hundred femtoseconds to several tens of picoseconds. In order to extract hydration dynamics precisely, we reconstructed apparent and lifetime-associated time-resolved fluorescence spectra with eight or nine transients at different wavelengths covering the blue and the red sides (figure 3.6B). By fitting these spectra to lognormal functions, we traced the time-dependent apparent emission maxima (ν_s) and lifetime-associated emission maxima (ν_l) as plotted in figure 3.6C. Using $\Delta\nu(t) = \nu_s(t) - \nu_l(t)$, we

correlated the extracted time-dependent spectral shift, $\Delta\nu(t)$, to the hydration energy relaxation, ΔE_s (figure 3.7).

Materials. All restriction enzymes were purchased from New England Biolabs (Beverly, MA). D,L-5,5,5-trifluoroleucine (Tfl) was purchased from Oakwood Products (West Columbia, SC). DNA oligomers were synthesized at Qiagen (Valencia, CA) or Integrated DNA Technologies (Coralville, IA). (4*S*)-2-amino-4-methylhexanoic acid (homoisoleucine, Hil) was prepared according to the methods of O'Donnell and Eckrich (59) and Dorizon and coworkers (60).

Plasmid construction. An EcoRI/HindIII fragment of pQEA1 (11) containing the A1 coding sequence was ligated into EcoRI/HindIII-digested pQE-80L (Qiagen) to yield pQE-80L/A1. The Asp residue at position 34 of A1 was mutated to Trp by site-directed mutagenesis. The resulting plasmid was designated pQE-80L/A1m. A Leu codon was introduced into either position 31 or position 37, yielding pQE-80L/S31L and pQE-80L/A37L, respectively. Plasmid pA1EL (12), which encodes both the protein A1 and a constitutively expressed leucyl-tRNA synthetase (LeuRS), was mutated using similar site-directed mutagenesis techniques. A Trp codon was introduced first at position 34 of A1, yielding pA1mEL. Introduction of leucine codons into either position 31 or position 37 resulted in the plasmids pS31LEL and pA37LEL, respectively.

Expression of fluorinated proteins. M9 medium supplemented with 0.4% glucose, 3.5 mg/L thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 amino acids (40 mg/L), and 200 mg/L

ampicillin was inoculated 1:50 with an overnight culture (M9) of *Escherichia coli* strain DH10B transformed with pQE-80L/A1m, pQE-80L/ S31L, or pQE-80L/A37L and grown at 37 °C with shaking. After each culture reached OD_{600} = 0.9–1.0, the cells were harvested by centrifugation (6000 × *g*, 4 °C, 6 min) and washed twice with cold 0.9% NaCl. The cell pellets were resuspended in M9 medium containing 19 amino acids (no Leu) and 1 mM Tfl. Protein expression was induced 10 min after the medium shift by addition of IPTG to a final concentration of 1 mM. After 3 h, the cells were harvested by centrifugation (6000 × *g*, 4 °C, 10 min), and the cells were stored at −20 °C at least 12 h before purification. In the case of the protein S31L-T, one sample was made using the procedure for the production of proteins containing Hil described below.

Expression of proteins containing homoisoleucine. M9 medium supplemented with 0.4% glucose, 35 mg⁄L thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 amino acids (40 mg⁄L), 200 mg⁄L ampicillin, and 35 mg⁄L kanamycin was inoculated 1:50 with an overnight culture (M9) of *E. coli* strain LAM1000 transformed with pREP4 and pA1mEL, pS31LEL, or pA37LEL at 37 °C with shaking. After each culture reached OD_{600} = 0.9–1.1, the cells were harvested by centrifugation (5000 × *g*, 4 °C, 15 min) and washed three times with cold 0.9% NaCl. The cell pellets were resuspended in M9 medium containing 19 amino acids (no Leu) and 0.5 mM Hil. Protein expression was induced 15 min after the medium shift by the addition of IPTG to a final concentration of 1 mM. Cells were harvested by centrifugation (5000 × *g*, 4 °C, 15 min), resuspended in Qiagen buffer B (8M urea, 100 mM NaH_2PO_4, 10 mM TrisCl, pH 8.0) and sonicated for 10 min total process time

with a pulse duration of 5 s and a wait duration of 5 s. The sonicated lysates were frozen at −80 °C for at least 12 h before proceeding with purification.

Expression of hydrogenated proteins. Rich (2×YT) medium was used instead of supplemented M9 medium. When the culture reached OD_{600} = 0.9–1.0, IPTG was added to a final concentration of 1 mM. After 3 h, the cells were harvested by centrifugation (6000 × *g*, 4 °C, 10 min), and the cells were stored at −20 °C at least 12 h before purification. In the case of S31L-L, one sample was expressed using the cell strain LAM1000 containing pREP4 and pS31LEL and harvested using the procedure used for the production of proteins containing Hil.

Protein purification. N-terminally histidine-tagged A1 variants were purified under denaturing conditions by affinity chromatography using Ni-NTA resin (Qiagen, Chatsworth, CA) according to the manufacturer's instructions. For proteins containing Hil and for one batch each of S31L-L and S31L-T, the lysates in Qiagen buffer B were thawed, sonicated, and then clarified using centrifugation (~75,000 × *g*, 25 °C, 10 min). Imidazole was added to Qiagen buffer C (50 mM) and Qiagen buffer E (250 mM) in order to improve purification efficiency. The purified protein solutions were dialyzed against 10 mM sodium acetate (pH 4)/100 mM NaCl, and were concentrated by ultrafiltration (Amicon Ultra-15 devices, mwco: 10,000 or 3000, Millipore, Billerica, MA). The protein concentration was determined as measured by the absorbance at 280 nanometers of solutions, assuming extinction coefficients of 5500 M^{-1} cm^{-1} (61).

Amino acid analysis and sedimentation velocity analysis. Amino acid analysis of fluorinated proteins was performed at the W. M. Keck Facility at Yale University (New Haven, CT) on a Hitachi L-8900 amino acid analyzer (San Jose, CA) after hydrolysis at 115 °C in 70% formic acid. Sedimentation velocity analysis was performed at the National Analytical Ultracentrifugation Facility at the University of Connecticut (Storrs, CT) by using a Beckman XL-I Analytical Ultracentrifuge at 20 °C. The rotor was accelerated to 55,000 rpm, and interference scans were acquired at 1 min intervals for 7 h. The data were analyzed by using the program Sedfit (62) to obtain normalized $c(s)$ versus sedimentation coefficient plots (figure 3.3).

Mass spectrometry. Liquid chromatography tandem mass spectrometry (LC/MS/MS) of proteins containing Hil or Tfl was performed at the Caltech Protein and Peptide Mass Analytical Laboratory. Trypsinized samples were subjected to liquid chromatography on an Eksigent (Dublin, CA) NanoLC-2D using a 6 cm long, 100 μm diameter C18 column, followed by MS/MS on an Applied Biosystems (Foster City, CA) QStar XL instrument. Data were analyzed using Analyst QS software provided by Applied Biosystems. Hil or Tfl incorporation levels were estimated using information contained within extracting ion currents (XIC) of trypsin-digested protein samples. For a given sample, a peak corresponding to a peptide globally substituted with the noncanonical amino acid and coding for multiple leucines was identified, and the related peak corresponding to replacement at a fraction of the leucine positions was also identified. Determination of the ratio of the partially substituted to globally substituted peak areas allowed for the estimation of amino acid incorporation levels assuming that leucines in the fragment were

replaced statistically. An example calculation is shown in the subsection "LC/MS/MS amino acid incorporation estimates," which can be found below.

Time-correlated single-photon counting (TCSPC). The protein samples were prepared at 55 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution. The TCSPC measurements were performed by using femtosecond pulses (<100 fs) from a Ti-sapphire oscillator (Spectra-Physics, Mai Tai HP). Laser output, of which the repetition rate was attenuated from 80 to 8 MHz utilizing a pulse picker (Spectra-Physics, Model 3980-5), was tuned to 885 nanometers and frequency-tripled to 295 nanometers using a time-plate tripler (Minioptic Technology, TP-2000B) for selective excitation of Trp. The UV beam, vertically polarized using a half waveplate, was introduced to a sample chamber and focused onto the sample cell. The residual frequency-doubled beam from the tripler was directed to a photodiode to trigger a TCSPC system (PicoQuant GmbH, FluoTime 200). Typically, the energy of the excitation pulse (attenuated) at the sample was ~10 picojoules. In a right-angle geometry, the emitted fluorescence was collected at a magic angle (54.7°) with respect to the vertically polarized excitation beam and focused into a MCP-PMT (Hamamatsu, R3809U), which is attached to a double monochromator. The photomultiplier tube signal was routed to a time-to-amplitude converter as a start signal followed by a constant fractional discriminator (PicoQuant GmbH, SPC 630). To avoid possible photobleaching and photodegradation, samples were kept stirring using a micro magnetic stirrer. In this configuration, the instrument response has a full width at half maximum of ~30 ps. Multiexponential decays convoluted with instrumental response functions were analyzed using the FluoFit software package (PicoQuant).

Femtosecond fluorescence upconversion. The protein samples were prepared at 550 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution. An amplified Ti-sapphire laser system (Spectra-Physics, Hurricane X) was used, which produces ~110 fs pulses centered at 805 nanometers (fundamental), with a 1 kHz repetition rate and a 0.8 millijoule energy. The output beam was split into equal parts to generate the pump and the gate pulse trains. For the pump, the fundamental light was used to pump an optical parametric amplifier (Spectra-Physics, OPA-800C), the infrared idler output of which was sum-frequency mixed with the residual fundamental in a 0.5 mm thick β-barium borate (BBO) crystal (type I), recompressed with a prism pair, and frequency-doubled to provide the 295-nanometers pulses in a 1.0 mm thick BBO crystal. The pump pulses were focused, with a 24 cm focal length lens, on the rotating circular cell (1 mm thickness) containing the sample. Typically, the energy of the pump pulse (attenuated) at the sample was ~200 nJ. At these energies, the fluorescence signals from samples were linearly dependent on the pump energy. To check for sample degradation during experiments, fluorescence spectra were periodically measured right after the rotating cell by using a fiber-optic-coupled spectrometer (Acton Research, SpectraPro-300i) coupled to a charge-coupled device (Princeton Instruments, SpectruMM-256HB) before and after the collection of averaged transients for each sample. No difference between the spectra was observed.

The forward-scattered fluorescence from excited samples was collected and focused by two off-axis parabolic mirrors into a 0.5 mm thick BBO crystal. Cutoff filters were placed between the mirrors to reject scattered laser light and pass the desired fluorescence wavelengths. The gate pulses, attenuated to 23 μJ/pulse, passed through a computer-controlled optical delay line and were noncollinearly overlapped with the

fluorescence in the BBO crystal. After the crystal, the upconverted signal was separated from the gate beam and the fluorescence by using an iris, and was focused on the entrance slit of a 0.25 m double-grating monochromator (Jobin Yvon, DH10) equipped with a photomultiplier tube at the exit slit. Upconversion efficiency was maximized by angle-tuning of the BBO crystal. The upconverted fluorescence transients were taken at the magic angle (54.7°) of the pump polarization relative to the gate polarization, parallel to the acceptance axis of the upconversion crystal, in order to eliminate the influence of induced sample anisotropy on the signal. The photomultiplier output was amplified (Stanford Research Systems (SRS), SR445) and processed by a gated integrator (SRS, SR250). The temporal response of the instrument was typically 350–450 fs. The observed fluorescence transients were fit to theoretical functions, using a Scientist nonlinear least-squares fitting program (Micromath), for the convolution of the Gaussian instrument response function with a sum of exponentials. All experiments were carried out at an ambient temperature of ~24 °C, and all fluorescence transients were obtained by the excitation of samples at 295 nanometers.

For fluorescence anisotropy measurements, the pump-beam polarization was rotated either parallel or perpendicular to the acceptance axis of the upconversion crystal to collect the parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) signals, respectively. These transients were used to construct time-resolved anisotropy: $r(t) = (I_{\parallel} - I_{\perp})/(I_{\parallel} + 2I_{\perp})$. The results of the time-resolved anisotropy are shown in figure 3.5. The ultrafast depolarization time constant, ϕ_1, attributed to fast internal conversion between the first two excited singlet states (^1L_a and ^1L_b) of Trp, varies dramatically with the time resolution. This process has a timescale of ~100 fs (63, 64). The limited resolution of our current apparatus (350–450 fs) does not

allow us to fully resolve these dynamics and gives rise to a large uncertainty in the value of ϕ_1. The variability of ϕ_1 will impact the fit of ϕ_{Trp}. The uncertainty in the amplitude of the anisotropy, r_{Trp} and r_∞, is not however affected by the limited time resolution of our data. Therefore we use the wobbling cone angle to reveal details about the local crowding near Trp. The wobbling cone angle is given by $1 - r_{Trp}/(r_{Trp} + r_\infty) = [(3\cos^2 q - 1)/2]^2$ (37), and only depends upon the amplitude of the tryptophan wobbling motion and the anisotropy due to the rotation of the molecule.

LC/MS/MS amino acid incorporation estimates. LC/MS/MS was used to estimate the replacement levels of leucine in some protein samples. Figure 3.8 depicts the total ion currents (TIC, figure 3.8A) and three extracted ion currents (XICs, figure 3.8B–D) from a digested A1m-H sample. The large peak in figure 3.8B corresponds to a peptide in which all of the Leu residues are replaced by Hil, the smaller peak in figure 3.8C corresponds to a mixture of two peptides containing one Leu and one Hil residue, and the very small peak in figure 3.8D corresponds to a peptide containing only Leu residues. The areas in the three XICs allow determination of the extent of incorporation of noncanonical amino acids in place of leucine. Assuming that there is a probability p of homoisoleucine substitution in place of leucine, the distribution of peak areas should correspond to the binomial distribution,

$$A\left[(1-p)^2 + 2(1-p)p + p^2\right], \quad (3.1)$$

where A is a multiplication factor equal to the total area of the three peaks and the three terms of the polynomial correspond to nonsubstituted, singly substituted, and doubly substituted peptides, respectively (the term for singly substituted peaks takes into account both positional isomers of singly substituted peptides). Because these three terms represent the only combinations of substitutions possible in the peptide, the relationship

$$(1-p)^2 + 2(1-p)p + p^2 = 1 \quad (3.2)$$

also holds. The ratio between two peaks in a peptide series depends only on the probability of incorporation and not on the value of A. Therefore, the ratio of two peaks from experimental data can be used in order to get an estimate of p. The ratio of the peak areas of singly substituted to doubly substituted peptides is

$$\frac{2(1-p)p}{p^2}. \quad (3.3)$$

Rearranging the above expression,

$$\frac{2p-2p^2}{p^2} = \frac{2(1-p)p}{p^2} = X, \quad (3.4)$$

where X is the experimentally observable ratio of singly substituted to doubly substituted peptides. Solving for p gives

$$p = 0, \quad (3.5)$$

or

$$p = \frac{2}{X + 2}, \quad (3.6)$$

with the root of interest being the nonzero root. Substituting for the ratio of peak areas gives an estimate of the incorporation level p.

In some cases, peaks corresponding to peptides containing three leucine or leucine analogs were observed and used to quantify incorporation levels. In these cases, the peak area distribution is represented by

$$A\left[(1-p)^3 + 3(1-p)^2 p + 3(1-p)p^2 + p^3\right], \quad (3.7)$$

with

$$(1-p)^3 + 3(1-p)^2 p + 3(1-p)p^2 + p^3 = 1. \quad (3.8).$$

Defining the ratio of doubly substituted to triply substituted peak area as X, substituting X into equation (3.8), and solving for p yields the nonzero root

$$p = \frac{3}{X + 3}, \quad (3.9)$$

again enabling an estimation of the incorporation level of noncanonical amino acids in place of leucine. In some cases, the tandem mass spectrometry did not enable positive identification of all possible positional isomers of a peptide. For example, in some cases, only two out of three of the possible doubly substituted positional isomers containing three possible substitution locations were identified in the tandem mass spectrometry data. In these cases, *X* was multiplied by an appropriate factor to account for peptides that were not observed (again invoking the assumption of completely random incorporation). Using the above example, when only two out of three doubly substituted peptides could be identified, X was multiplied by a factor of 1.5 in order to estimate what the peak area ratio would have been with all three peaks present in equal weights. Using this methodology, the homoisoleucine-containing proteins used were found to have 90% or greater Hil in place of Leu, and the sample of S31L-T that was analyzed in this fashion contained approximately 99% Tfl in place of Leu. These results were obtained by examining three separate series of peptides from each protein sample. These peptides had the following sequences: AEIGDLNNTSGIR, GSHHHHHHGSMASGDLENEVAQLER, and SLEWEAAELEQK (A1m), LLEWEAAELEQK (S31L), or SLEWEALELEQK (A37L).

Acknowledgements

We thank Professor Thomas Miller for helpful discussion, J. D. Fisk for synthesis of homoisoleucine, and the referees for their thoughtful comments on the original manuscript. This work is supported by National Institutes of Health (NIH) Grant GM62523, National Science Foundation (NSF) Grant DMR-0964886, Office of Naval

Research (ONR) Grant N00014-03-1-0793, a Samsung Scholarship (to T. H. Y.), and a National Defense Science and Engineering Graduate Fellowship (to J. A. V.).

References

1. Budisa N (2006) *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins* (Wiley-VCH, Weinheim, Germany).
2. Link AJ, Mock ML, & Tirrell DA (2003) Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **14,** 603-609.
3. Xie JM & Schultz PG (2006) Innovation: a chemical toolkit for proteins - an expanded genetic code. *Nature Reviews: Molecular Cell Biology* **7,** 775-782.
4. Marsh ENG (2000) Towards the nonstick egg: designing fluorous proteins. *Chem. Biol.* **7,** R153-R157.
5. Yoder NC, Yuksel D, Dafik L, & Kumar K (2006) Bioorthogonal noncovalent chemistry: fluorous phases in chemical biology. *Curr. Opin. Chem. Biol.* **10,** 576-583.
6. Bilgicer B, Fichera A, & Kumar K (2001) A coiled coil with a fluorous core. *J. Am. Chem. Soc.* **123,** 4393-4399.
7. Jackel C, Salwiczek M, & Koksch B (2006) Fluorine in a native protein environment - How the spatial demand and polarity of fluoroalkyl groups affect protein folding. *Angew. Chem. Int. Ed.* **45,** 4198-4203.
8. Lee KH, Lee HY, Slutsky MM, Anderson JT, & Marsh ENG (2004) Fluorous effect in proteins: De novo design and characterization of a four-alpha-helix bundle protein containing hexafluoroleucine. *Biochemistry* **43,** 16277-16284.

9. Montclare JK, Son S, Clark GA, Kumar K, & Tirrell DA (2009) Biosynthesis and stability of coiled-coil peptides containing (2S,4R)-5,5,5-trifluoroleucine and (2S,4S)-5,5,5-trifluoroleucine. *ChemBioChem* **10,** 84-86.

10. Son S, Tanrikulu IC, & Tirrell DA (2006) Stabilization of bzip peptides through incorporation of fluorinated aliphatic residues. *ChemBioChem* **7,** 1251-1257.

11. Tang Y*, et al.* (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. *Angew. Chem. Int. Ed.* **40,** 1494-1496.

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

14. Wang P, Tang Y, & Tirrell DA (2003) Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc.* **125,** 6900-6906.

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

16. 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.

17. Biffinger JC, Kim HW, & DiMagno SG (2004) The polar hydrophobicity of fluorinated compounds. *ChemBioChem* **5,** 622-627.

18. Gao JM, Qiao S, & Whitesides GM (1995) Increasing binding constants of ligands to carbonic-anhydrase by using greasy tails. *J. Med. Chem.* **38,** 2292-2301.

19. Jackel C & Koksch B (2005) Fluorine in peptide design and protein engineering. *Eur. J. Org. Chem.* 4483-4503.

20. Leroux F (2004) Atropisomerism, biphenyls, and fluorine: A comparison of rotational barriers and twist angles. *ChemBioChem* **5,** 644-649.

21. Muller K, Faeh C, & Diederich F (2007) Fluorine in pharmaceuticals: Looking beyond intuition. *Science* **317,** 1881-1886.

22. Dunitz JD (2004) Organic fluorine: Odd man out. *ChemBioChem* **5,** 614-621.

23. Graupe M, Takenaga M, Koini T, Colorado R, & Lee TR (1999) Oriented surface dipoles strongly influence interfacial wettabilities. *J. Am. Chem. Soc.* **121,** 3222-3223.

24. Merzel F & Smith JC (2002) Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci. U. S. A.* **99,** 5378-5383.

25. Pal SK & Zewail AH (2004) Dynamics of water in biological recognition. *Chem. Rev.* **104,** 2099-2123.

26. Levy Y & Onuchic JN (2006) Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 389-415.

27. Qiu WH*, et al.* (2006) Protein surface hydration mapped by site-specific mutations. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 13979-13984.

28. Qiu WH*, et al.* (2006) Ultrafast solvation dynamics of human serum albumin: Correlations with conformational transitions and site-selected recognition. *J. Phys. Chem. B* **110,** 10540-10549.

29. Cohen BE*, et al.* (2002) Probing protein electrostatics with a synthetic fluorescent amino acid. *Science* **296,** 1700-1703.

30. Daidone I, Ulmschneider MB, Di Nola A, Amadei A, & Smith JC (2007) Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 15230-15235.

31. Moore FG & Richmond GL (2008) Integration or segregation: How do molecules behave at oil/water interfaces? *Acc. Chem. Res.* **41,** 739-748.

32. Protein variants containing Tfl, Leu, and Hil, are identified by the addition of "-T," "-L," and "-H," respectively. For example, S31L containing Leu is designated "S31L-L."

33. Chen YH, Yang JT, & Martinez HM (1972) Determination of secondary structures of proteins by circular-dichroism and optical rotary dispersion. *Biochemistry* **11,** 4120-4131.

34. Perez-Iratxeta C & Andrade-Navarro MA (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol.* **8**.

35. Petka WA, Harden JL, McGrath KP, Wirtz D, & Tirrell DA (1998) Reversible hydrogels from self-assembling artificial proteins. *Science* **281,** 389-392.

36. Vivian JT & Callis PR (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.* **80,** 2093-2109.

37. Steiner RF (1991) *Top. Fluoresc. Spectrosc.*, ed Lakowicz JR (Plenum Press, New York), Vol 2, pp 1-52.

38. Qiu WH*, et al.* (2008) Ultrafast quenching of tryptophan fluorescence in proteins: Interresidue and intrahelical electron transfer. *Chem. Phys.* **350,** 154-164.

39. Siemiarczuk A, Petersen CE, Ha CE, Yang JS, & Bhagavan NV (2004) Analysis of tryptophan fluorescence lifetimes in a series of human serum albumin mutants with substitutions in subdomain 2A. *Cell Biochem. Biophys.* **40,** 115-122.

40. Xu JH & Knutson JR (2009) Quasi-static self-quenching of Trp-X and X-Trp dipeptides in water: ultrafast fluorescence decay. *J. Phys. Chem. B* **113,** 12084-12089.

41. Lu WY, Kim J, Qiu WH, & Zhong DP (2004) Femtosecond studies of tryptophan solvation: correlation function and water dynamics at lipid surfaces. *Chem. Phys. Lett.* **388,** 120-126.

42. Li TP, Hassanali AAP, Kao YT, Zhong DP, & Singer SJ (2007) Hydration dynamics and time scales of coupled water-protein fluctuations. *J. Am. Chem. Soc.* **129,** 3376-3382.

43. Zhang LY, Yang Y, Kao YT, Wang LJ, & Zhong DP (2009) Protein hydration dynamics and molecular mechanism of coupled water-protein fluctuations. *J. Am. Chem. Soc.* **131,** 10677-10691.

44. Golosov AA & Karplus M (2007) Probing polar solvation dynamics in proteins: A molecular dynamics simulation analysis. *J. Phys. Chem. B* **111,** 1482-1490.

45. Lee HY, Lee KH, Al-Hashimi HM, & Marsh ENG (2006) Modulating protein structure with fluorous amino acids: Increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine. *J. Am. Chem. Soc.* **128,** 337-343.

46. We anticipated that A1m-H would also show hydration dynamics identical to those of A1m-L and A1m-T. However as discussed before, the local protein packing and

flexibility near the Trp probe appear to have been altered by substitution of Hil for Leu, reflected in the large Trp wobbling angle and total Stokes shift, ΔE_s. Because the local structure has been changed, we would expect to observe altered hydration dynamics.

47. Russo D, Hura G, & Head-Gordon T (2004) Hydration dynamics near a model protein surface. *Biophys. J.* **86,** 1852-1862.

48. Russo D, Murarka RK, Copley JRD, & Head-Gordon T (2005) Molecular view of water dynamics near model peptides. *J. Phys. Chem. B* **109,** 12966-12975.

49. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* **437,** 640-647.

50. Richmond GL (2002) Molecular bonding and interactions at aqueous surfaces as probed by vibrational sum frequency spectroscopy. *Chem. Rev.* **102,** 2693-2724.

51. Shen YR & Ostroverkhov V (2006) Sum-frequency vibrational spectroscopy on water interfaces: Polar orientation of water molecules at interfaces. *Chem. Rev.* **106,** 1140-1154.

52. Miller TF, Vanden-Eijnden E, & Chandler D (2007) Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 14559-14564.

53. ten Wolde PR & Chandler D (2002) Drying-induced hydrophobic polymer collapse. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6539-6543.

54. Ashbaugh HS & Paulaitis ME (2001) Effect of solute size and solute-water attractive interactions on hydration water structure around hydrophobic solutes. *J. Am. Chem. Soc.* **123,** 10721-10728.

55. Huang DM & Chandler D (2002) The hydrophobic effect and the influence of solute-solvent attractions. *J. Phys. Chem. B* **106,** 2047-2053.

56. Jensen TR*, et al.* (2003) Water in contact with extended hydrophobic surfaces: Direct evidence of weak dewetting. *Phys. Rev. Lett.* **90**.

57. Mezger M*, et al.* (2010) On the Origin of the Hydrophobic Water Gap: An X-ray Reflectivity and MD Simulation Study. *J. Am. Chem. Soc.* **132,** 6735-6741.

58. Rossky PJ & Dalvi VH (2010) Molecular origins of fluorocarbon hydrophobicity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13603-13607.

59. Odonnell MJ & Eckrich TM (1978) Synthesis of amino acid derivatives by catalytic phase transfer alkylations. *Tetrahedron Lett.* 4625-4628.

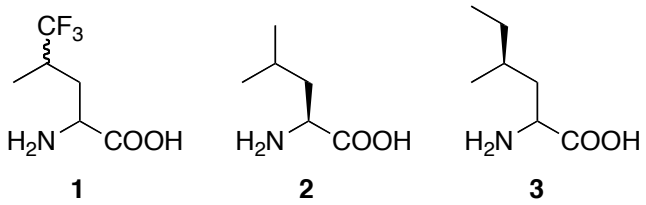
60. Dorizon P*, et al.* (1999) Stereoselective synthesis of highly functionalized cyclopropanes. Application to the asymmetric synthesis of (1S,2S)-2,3-methanoamino acids. *J. Org. Chem.* **64,** 4712-4724.

61. Swiss Institute of Bioinformatics. Protparam tool. 2003, http://ca.expasy.org/tools/protparam.html.

62. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606-1619.

63. Shen XH & Knutson JR (2001) Subpicosecond fluorescence spectra of tryptophan in water. *J. Phys. Chem. B* **105,** 6260-6265.

64. Zhong DP, Pal SK, & Zewail AH (2001) Femtosecond studies of protein - DNA binding and dynamics: Histone I. *Chemphyschem* **2,** 219-227.



Scheme 3.1. Amino acids used in study. **1**, 5,5,5-trifluoroleucine (Tfl). **2**, leucine (Leu). **3**, (4*S*)-2-amino-4-methylhexanoic acid (homoisoleucine, Hil).

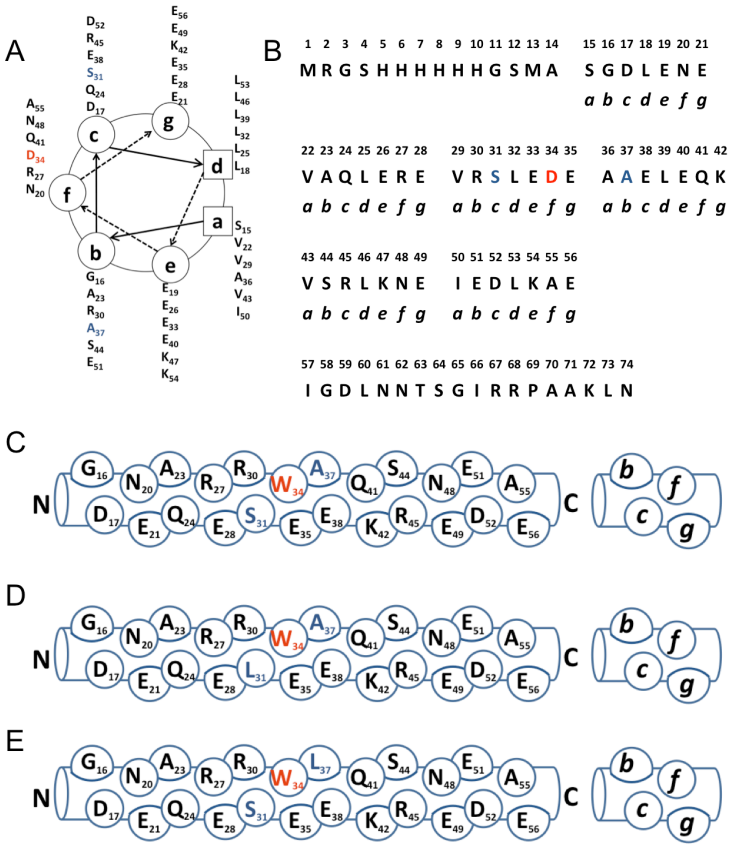


Figure 3.1. Protein sequence and structure. A) Helical wheel diagram and B) amino acid sequence of the A1 protein. The Asp residue at the *f* position of the third heptad (position 34) was replaced by Trp to yield a variant of A1 designated Alm. A Leu codon was introduced at the *c* position of the third heptad (position 31, dark blue) or at the *b* position of the fourth heptad (position 37, dark blue) to yield S31L and A37L, respectively. Side views of C) A1m, D) S31L, and E) A37L.

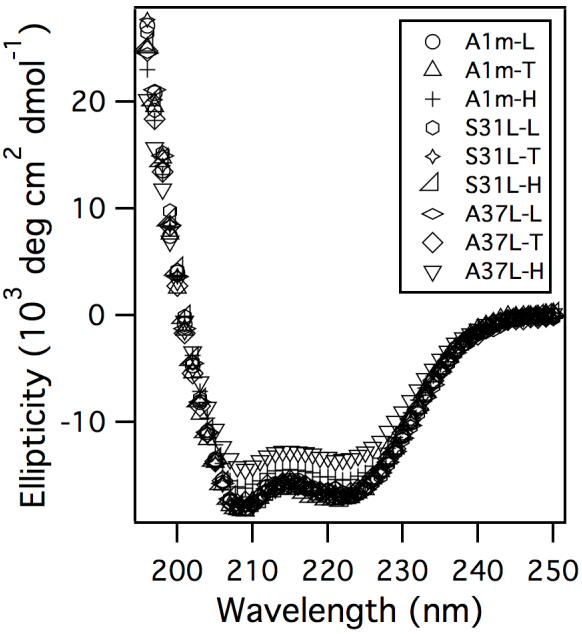


Figure 3.2. Circular dichroism. Wavelength scans of A1m and variant proteins at 25 ºC. The protein samples were prepared at 20 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solutions.

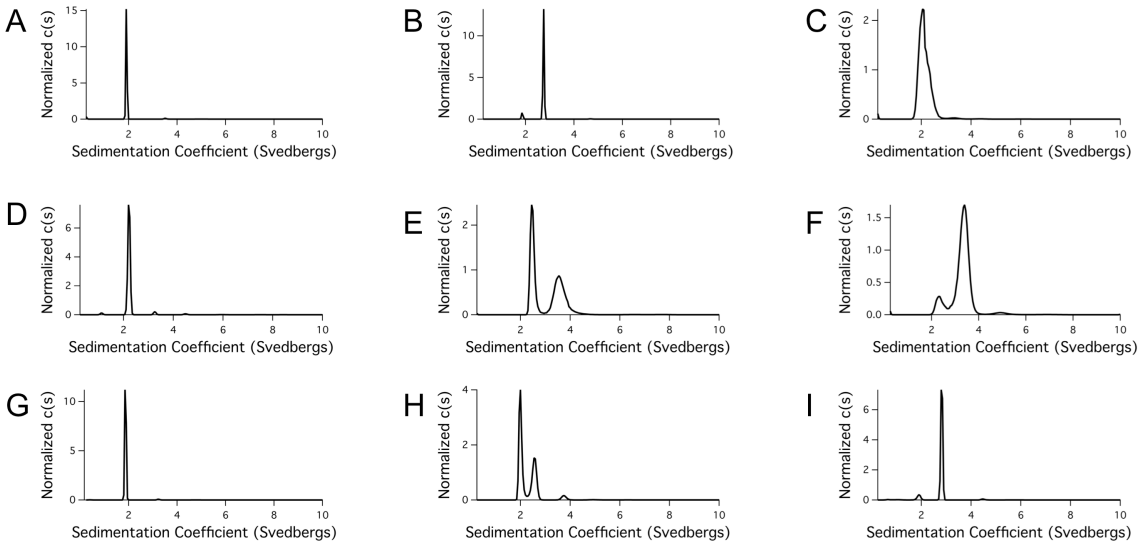


Figure 3.3. Normalized plots from the Sedfit *c*(*s*) analysis for A) A1m-L, B) S31L-L, C) A37L-L, D) A1m-T, E) S31L-T, F) A37L-T, G) A1m-H, H) S31L-H, and I) A37L-H. The protein samples were prepared at 550 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution. The broad, single peak of the A37L-L trace may be the result of an equilibrium mixture between trimeric and higher-order species. It seems likely that the breadth of the traces derived from sedimentation velocity analysis of A1m-T, S31L-T, and A37L-T is a consequence of incomplete replacement of Leu with Tfl.

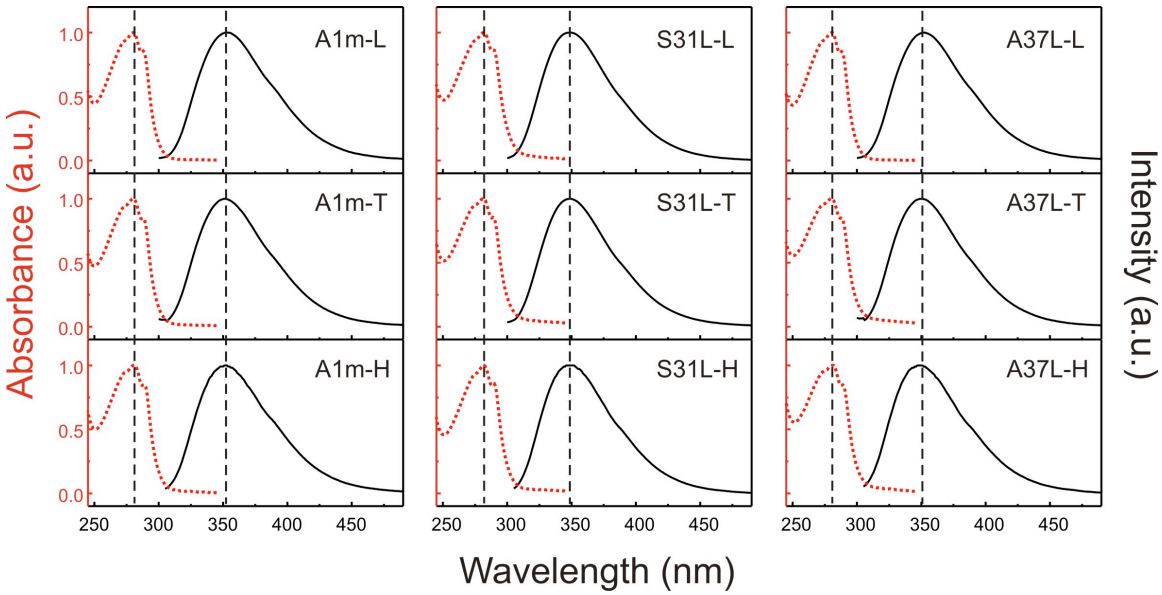


Figure 3.4. Steady-state UV-visible absorption (red) and fluorescence emission spectra (black) of proteins excited at 295 nanometers. The protein samples were prepared at 550 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution.

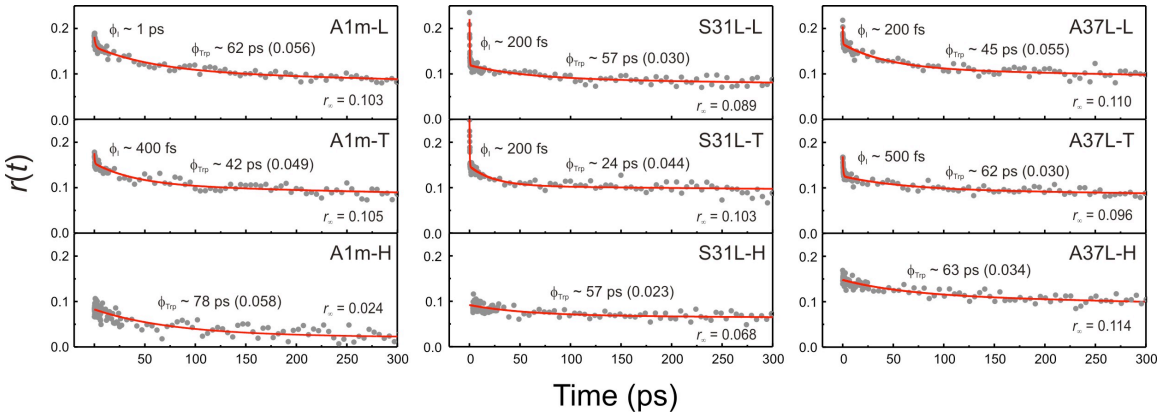


Figure 3.5. Time-resolved anisotropy, $r(t)$, of the proteins. All anisotropy decays were fitted to $r(t) = r_1 \exp(-t/f_1) + r_{Trp} \exp(-t/f_{Trp}) + r_\infty$, where r_1 is the initial ultrafast anisotropy, r_{Trp} is the Trp motion-related anisotropy (value given in parentheses in each panel), r_∞ is the offset anisotropy, f_1 is the initial ultrafast internal-conversion time constant of Trp (≤1 ps), and f_{Trp} is the Trp-rotational correlation time constant.

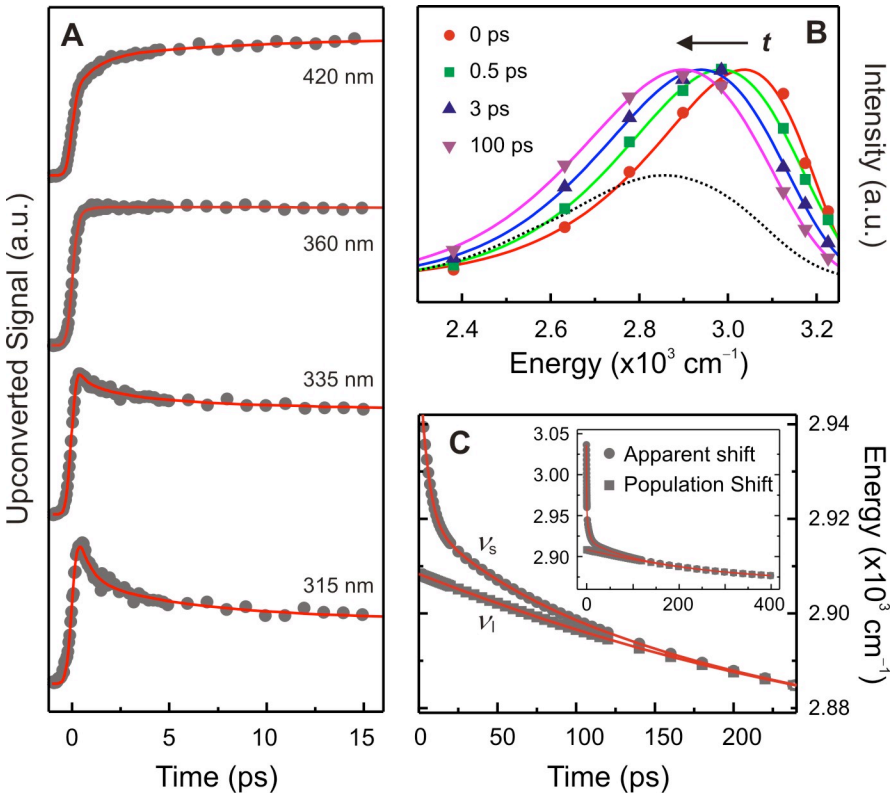


Figure 3.6. Hydration dynamics. Experimental determination of local hydration dynamics at the surface of A1m-T, excited at 295 nm. A) Representative femtosecond-resolved fluorescence up-conversion transients. B) Normalized time-resolved fluorescence spectra at different time delays. The steady-state emission spectrum is also depicted (dotted line). C) Time-dependent spectral shift of the apparent emission maxima (ν_s) and the lifetime-associated (population) emission maxima (ν_l). Inset: entire evolution of ν_s and ν_l.

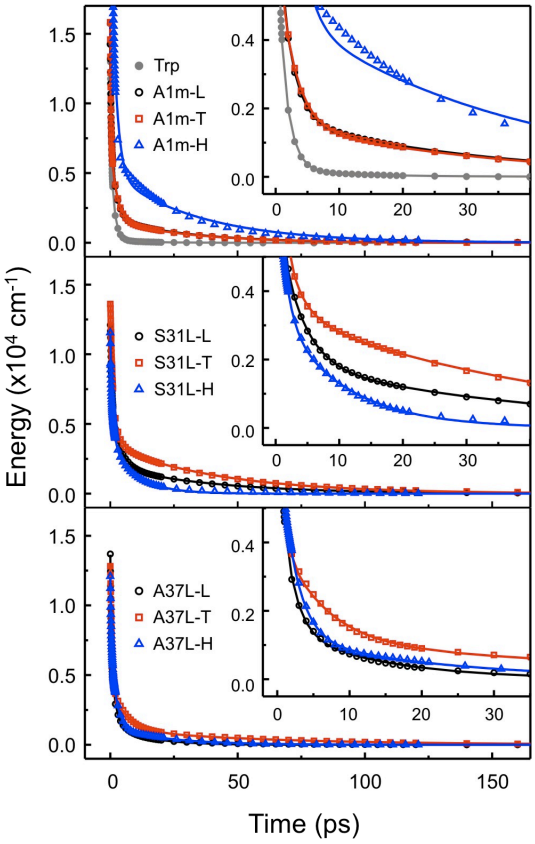


Figure 3.7. Hydration energy relaxation. Comparison of the hydration-correlated energy relaxation, $\Delta E_s(t)$, probed by Trp emission. Top panel: The solvation-energy relaxation data for A1m proteins. Data for free Trp in the same buffer is also depicted for comparison. Middle panel: Solvation energy relaxation data for S31L analogs. Bottom panel: solvation energy relaxation data for A37L analogs. Insets: enlargement of the early-time hydration behavior.

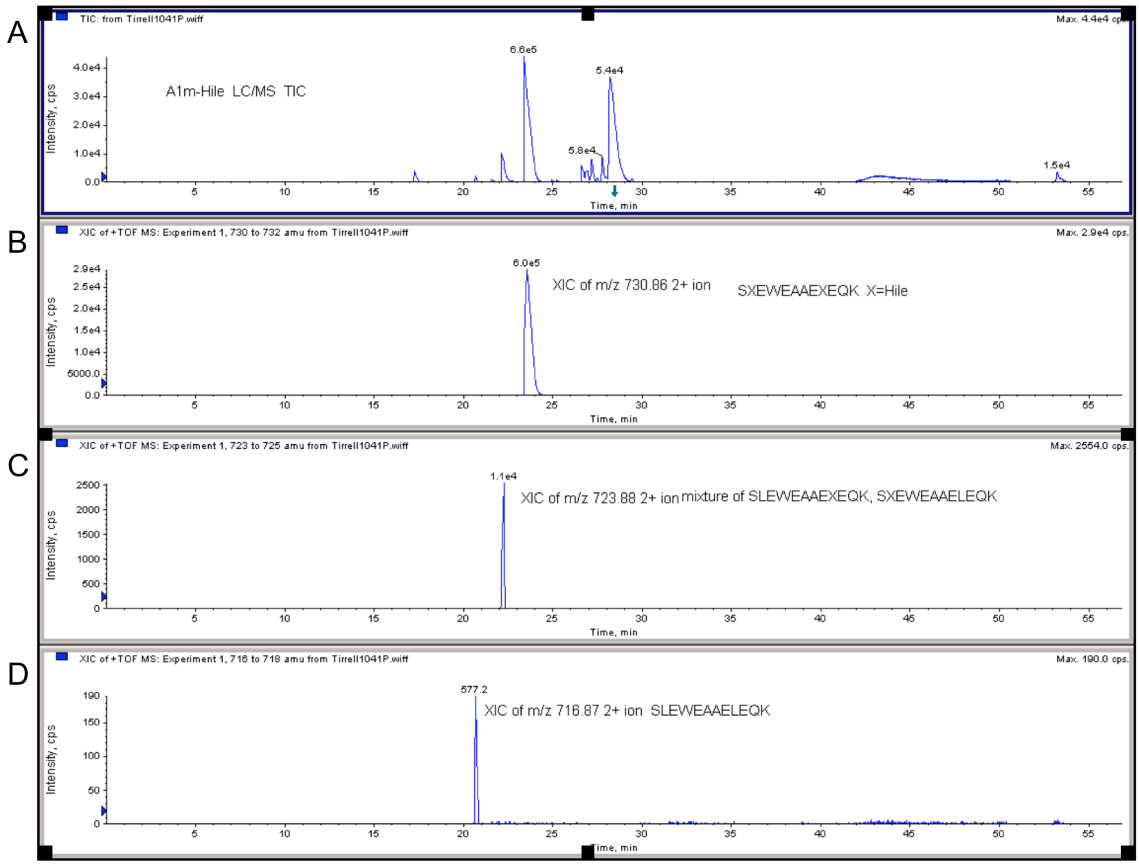


Figure 3.8. LC/MS/MS of trypsinized A1m-H. A) TIC of digested protein sample. B)–D) XICs of peptides containing B) two, C) one, or D) no leucine to homoisoleucine substitutions in the peptide SLEWEAAELEQK. The ratios of the peak areas obtained in the XICs can be used to estimate the extent of leucine replacement in the protein sample. Peptide masses: SXEWEAAEXEQK: 2+ ion: 730.86 Da observed, 730.87 Da expected. SLEWEAAEXEQK, SXEWEAAEXEQK: 2+ ion: 723.88 Da observed, 723.86 Da expected. SLEWEAAELEQK 2+ ion: 716.87 Da observed, 716.85 expected.

Table 3.1. Fluorescence emission maxima (λ_{max}), hydration-correlated energy relaxation [$\Delta E_s(t)$], and depolarization dynamics [$r(t)$]

Sample	λ_{max}, nm	$\Delta E_s(t)$*						$r(t)$ †	
		τ_1, ps	τ_2, ps	τ_3, ps	E_1, cm^{-1}	E_2, cm^{-1}	E_3, cm^{-1}	r_{Trp}	θ, °
Trp	353	0.30	1.5	13	883	682	18	0.196	
A1m-L	352	0.30	2.1	31	610	646	171	0.056	21
A1m-T	352	0.28	2.5	31	877	568	161	0.049	20
A1m-H	352		1.9	34	0	2138	500	0.058	33
S31L-L	349	0.53	3.6	40	580	450	194	0.030	18
S31L-T	349	0.79	3.0	48	607	375	333	0.044	19
S31L-H	349	0.21	1.4	10	308	492	344	0.023	18
A37L-L	352	0.31	1.7	13	685	522	157	0.055	20
A37L-T	350	0.56	6.1	61	685	324	108	0.030	17
A37L-H	349	0.34	2.2	21	443	625	128	0.034	17

* All hydration-correlated energy relaxation dynamics were fitted to $\Delta E_s(t) = E_1\exp(-t/\tau_1) + E_2\exp(-t/\tau_2) + E_3\exp(-t/\tau_3)$.

† Refer to materials and methods for anisotropy analysis detail.

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.

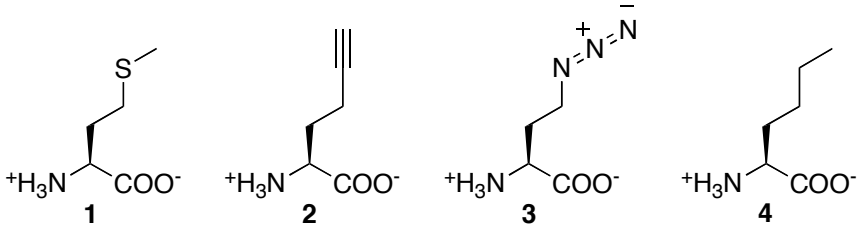
60. Link AJ*, et al.* (2006) Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10180-10185.

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.

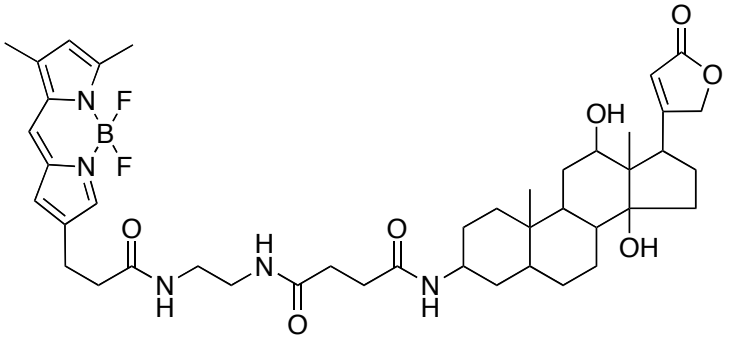
62. Kiick KL, Saxon E, Tirrell DA, & Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 19-24.

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.

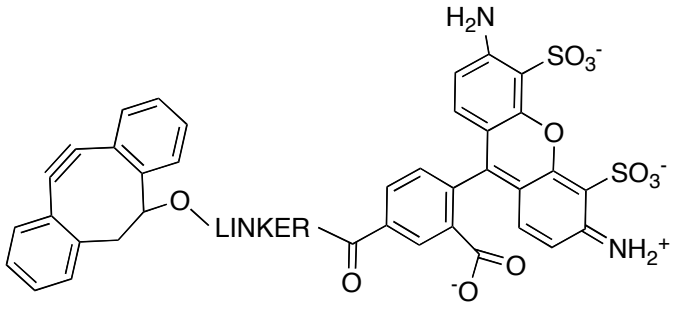
64. Onell A & Andersson K (2005) Kinetic determinations of molecular interactions using Biacore - minimum data requirements for efficient experimental design. *J. Mol. Recognit.* **18,** 307-317.



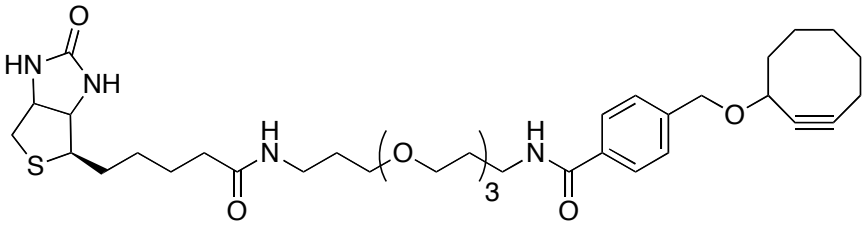
1 **2** **3** **4**



5

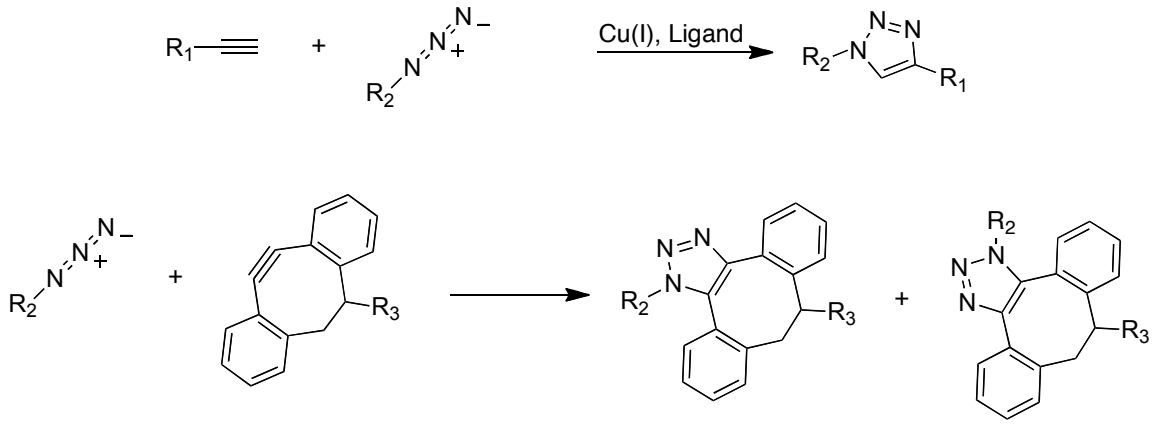


6



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).

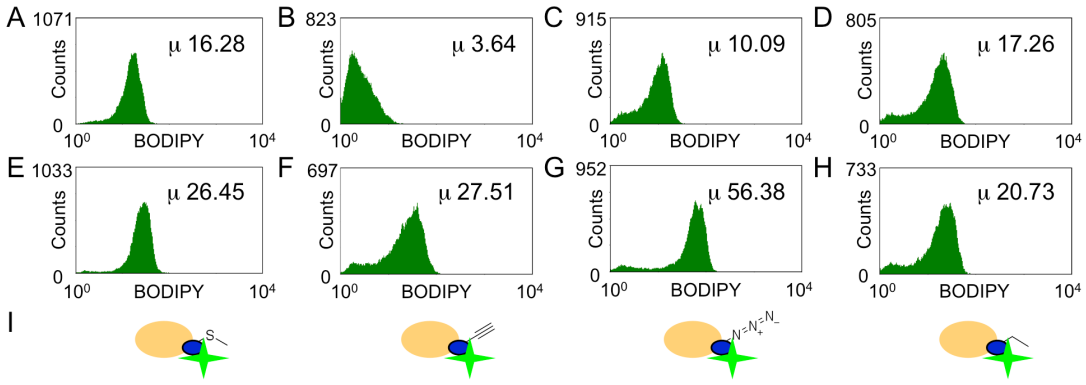


Figure 4.1. Flow cytometry studies of the binding properties of cell surface displayed scFvs. (*A–D*) Base construct expression in Met (*A*), Hpg (*B*), Aha (*C*), and Nrl (*D*) contexts, followed by exposure of cells to fluorescently labeled antigen **5**. (*E–H*) Mut2 construct expression in Met (*E*), 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.

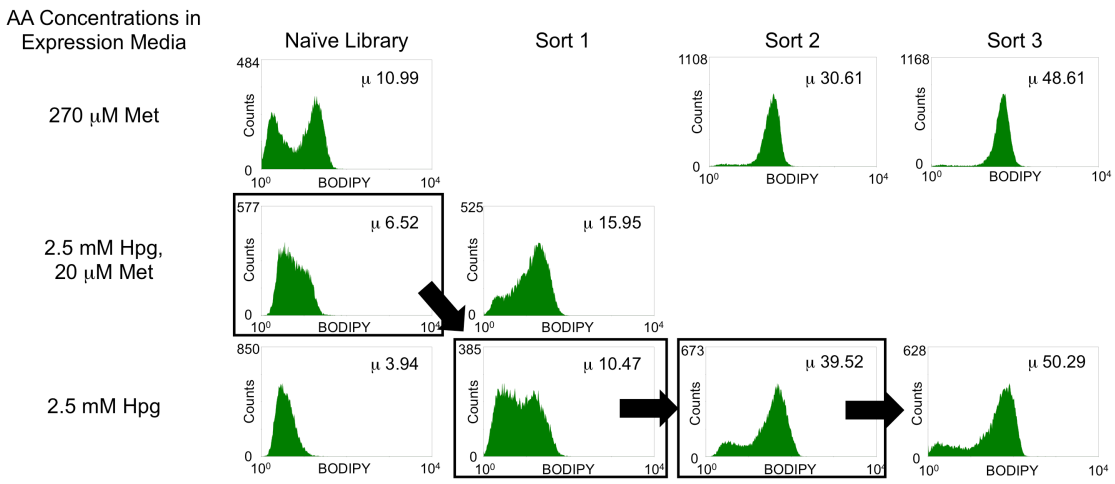


Figure 4.2. Fluorescence activated cell sorting of Lib1_1a for clones that function when Met is replaced by Hpg. Expression of cells harboring the naïve library and sorted populations was performed in minimal media containing concentrations of Met and Hpg as indicated. After treatment of induced cells 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.

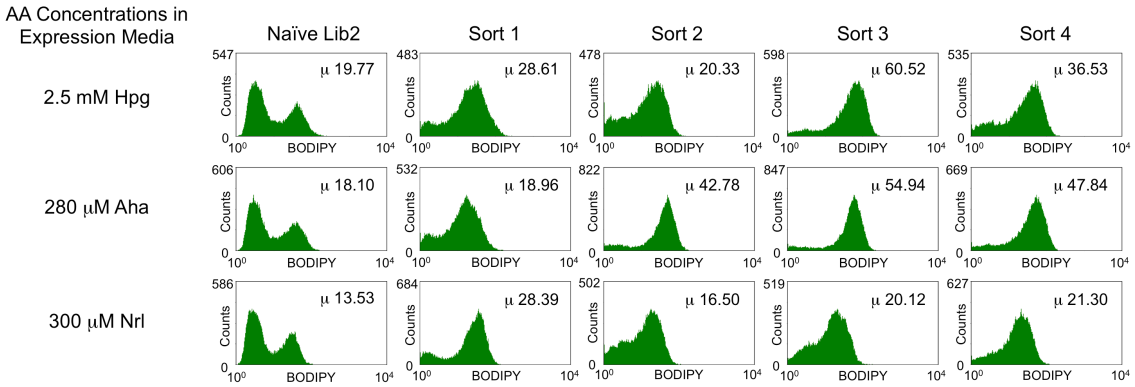


Figure 4.3. Fluorescence activated cell sorting of Lib2 for clones that function when Met is replaced by ncAAs. Expression of the naïve library and sorted populations was performed in minimal media containing concentrations of ncAAs as indicated. The fluorescence of each population was measured on the 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.

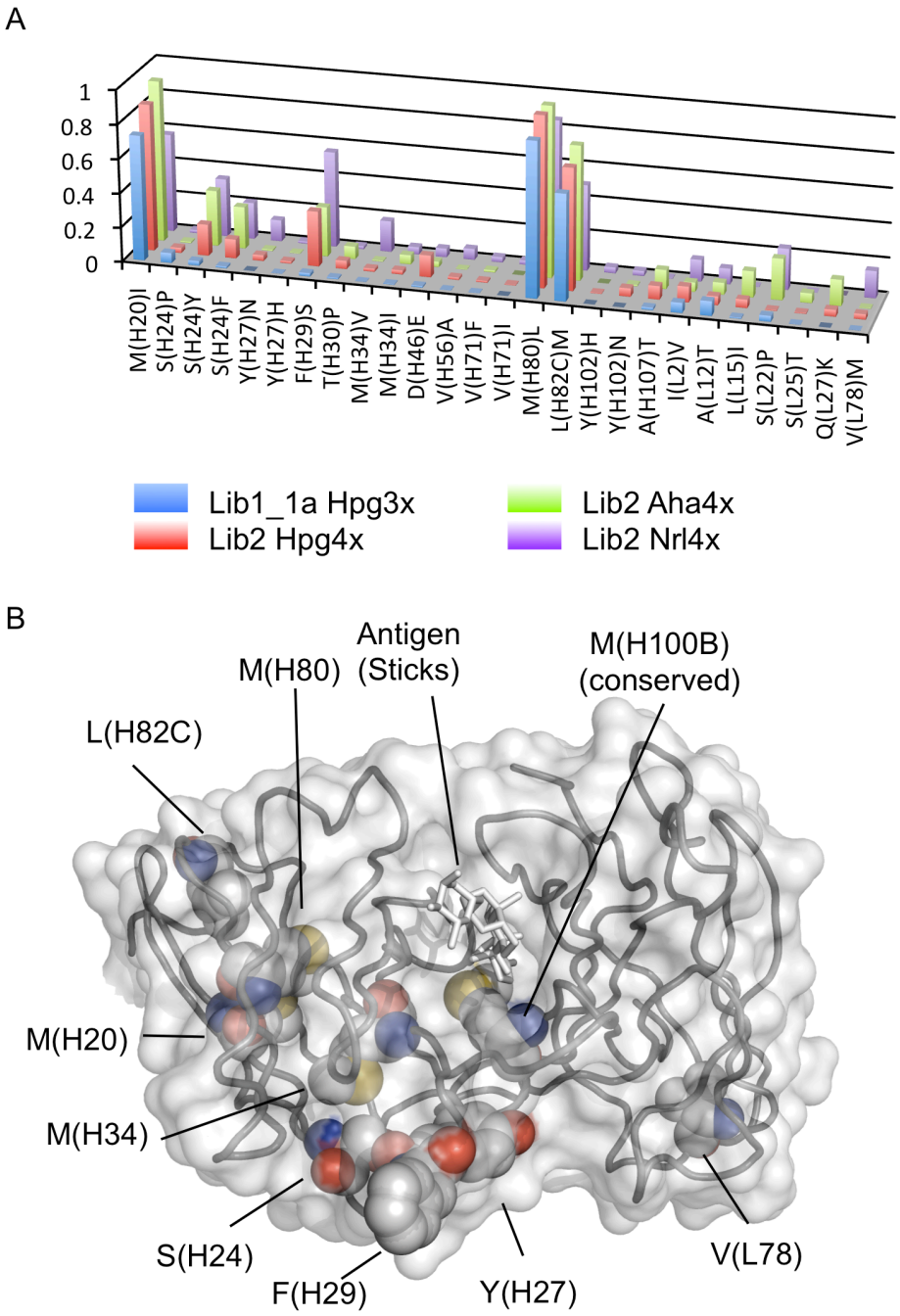


Figure 4.4. Population-level sequence characterization of scFv mutants using high-throughput sequencing. (*A*) Frequently mutated amino acids in sorted scFv populations (Kabat numbering). (*B*) Structural positions of mutations to scFv in crystal structure of 26-10 Fab-digoxin complex solved by Jeffrey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

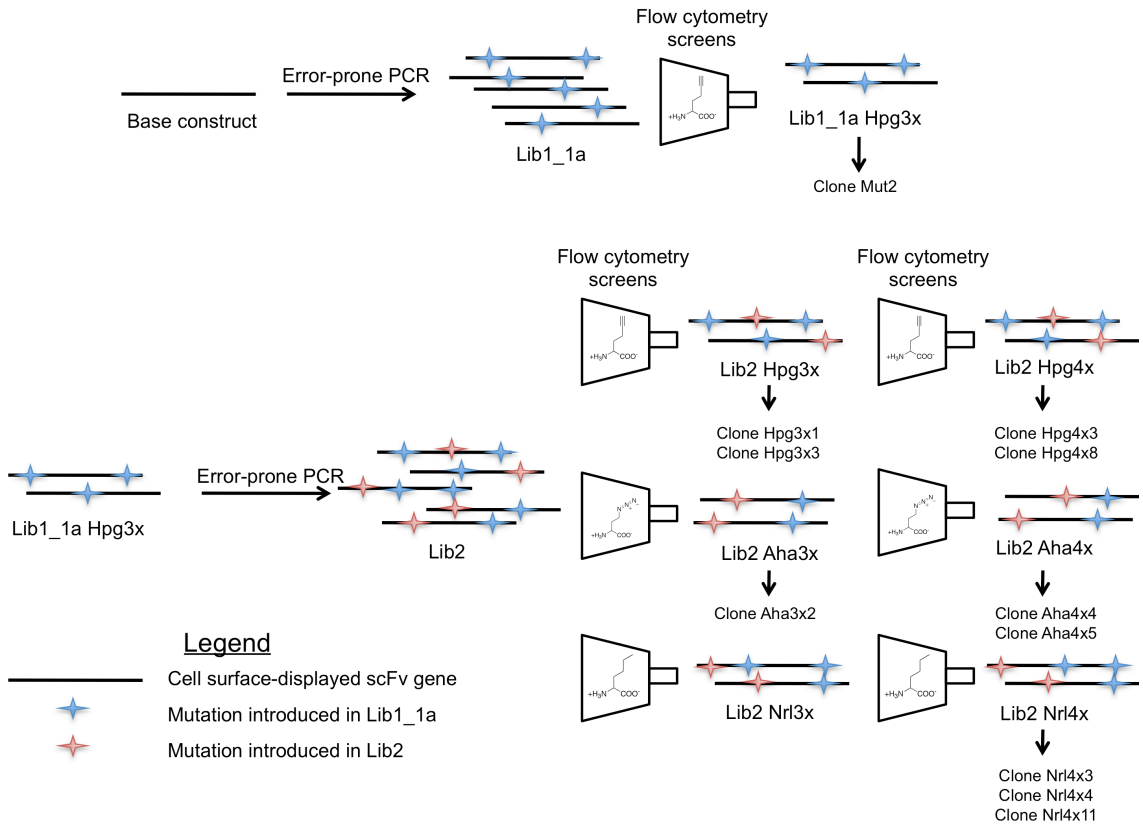


Figure 4.5. Summary of directed evolution of cell surface-displayed scFvs. Error-prone PCR using the DNA of Base (both display anchor and scFv) enabled construction and expression of Lib1_1a on the surface of *E. coli* cells. The library was screened for clones retaining binding function after the replacement 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.

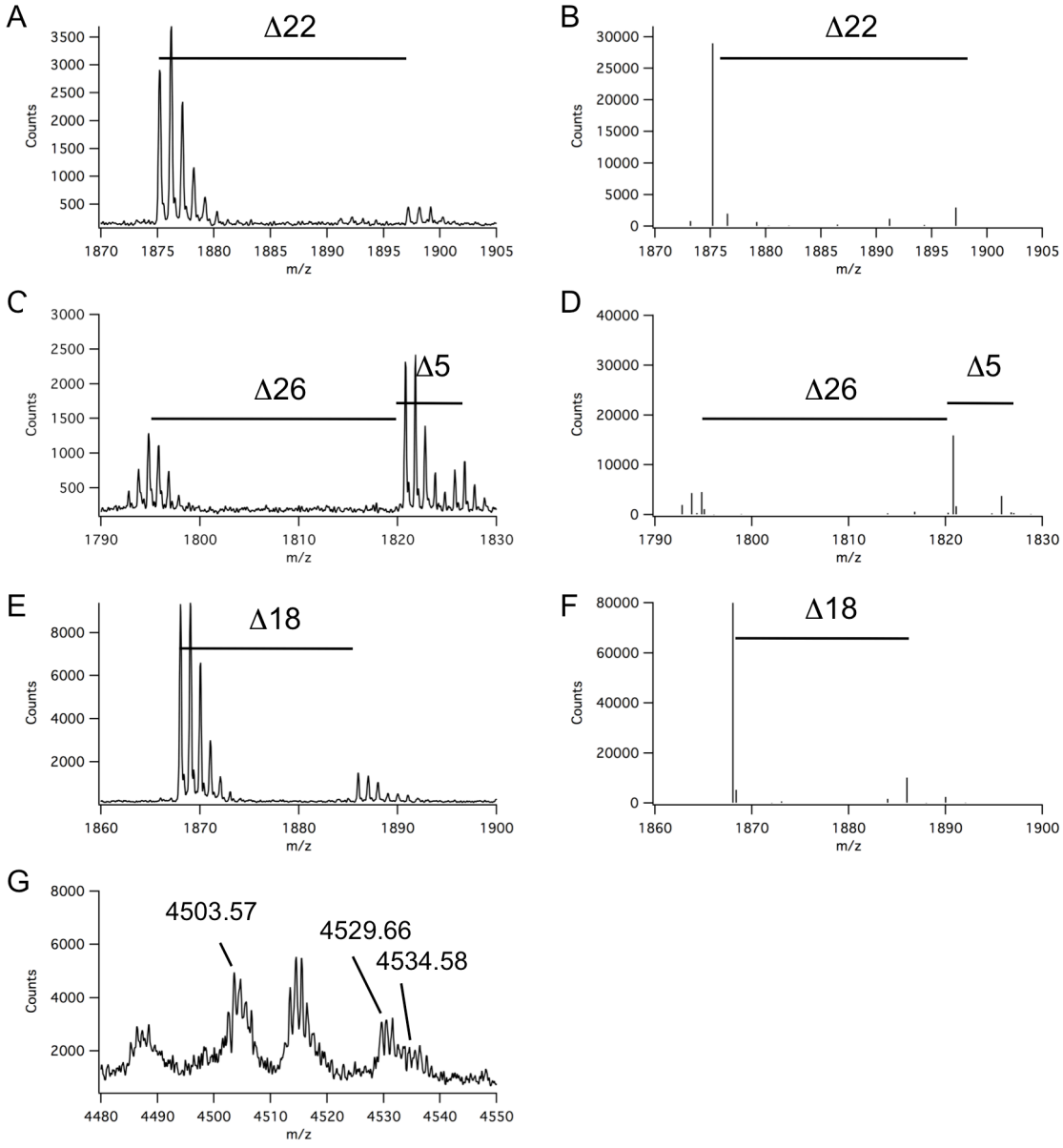


Figure 4.6. Examples of data used in estimating ncAA incorporation levels in scFvs with matrix-assisted laser desorption ionization (MALDI) mass spectrometry. (*A*, *B*) Portion of MALDI spectrum of trypsinized Hpg4x3 containing Hpg before (*A*) and after (*B*) processing that includes the peptide FSGNIFTDFY**M**NWVR. 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.

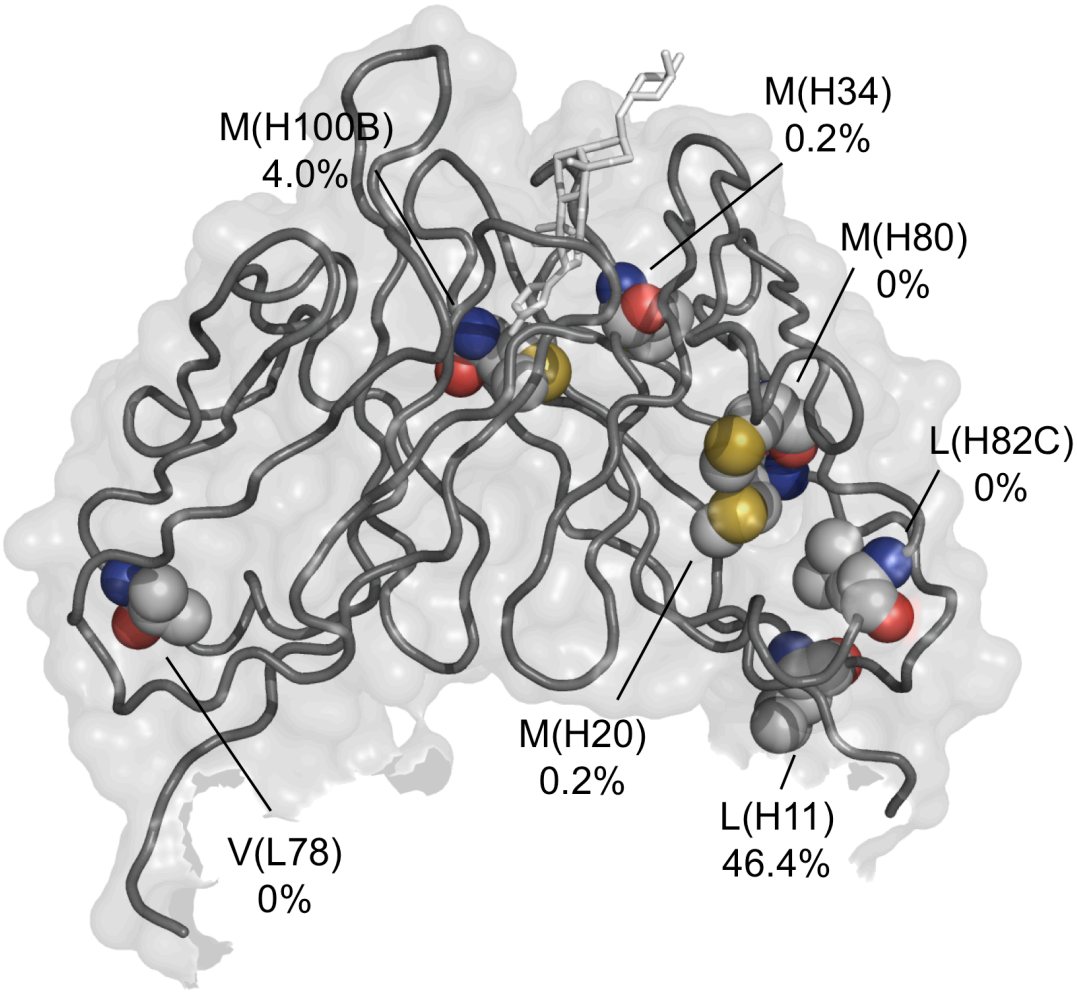


Figure 4.7. Positions of Met residues and residues mutated to Met in scFv. Each residue in the protein that exists as Met in at least one solubly produced variant is shown in space filling, with calculated surface accessibilities from GetArea (40) listed beneath the residue of interest. The accessibility 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.

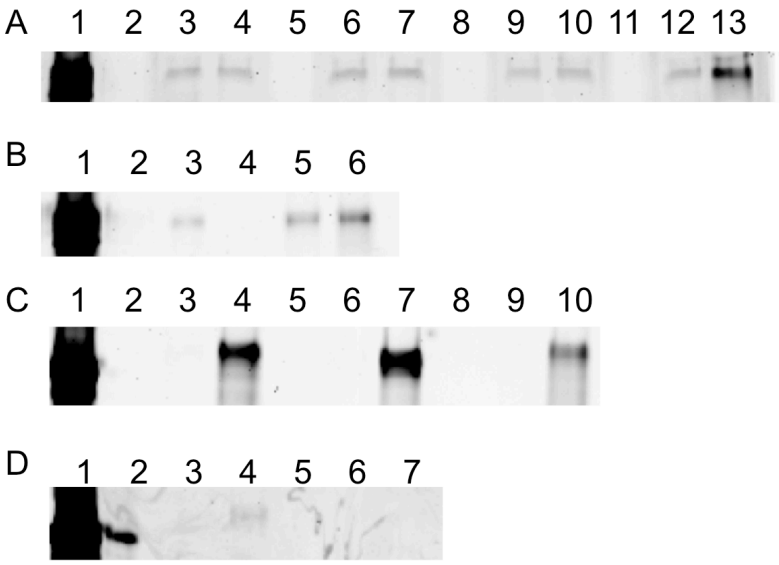


Figure 4.8. Copper-catalyzed azide-alkyne cycloadditions (CuAAC). All reactions were performed with solutions of 1.0–1.75 μM protein using reaction conditions as recommended by Hong et al. (58). After one hour, room temperature reactions, all reactions were quenched using 10 mM 3-azidopropanol. Reacted 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.

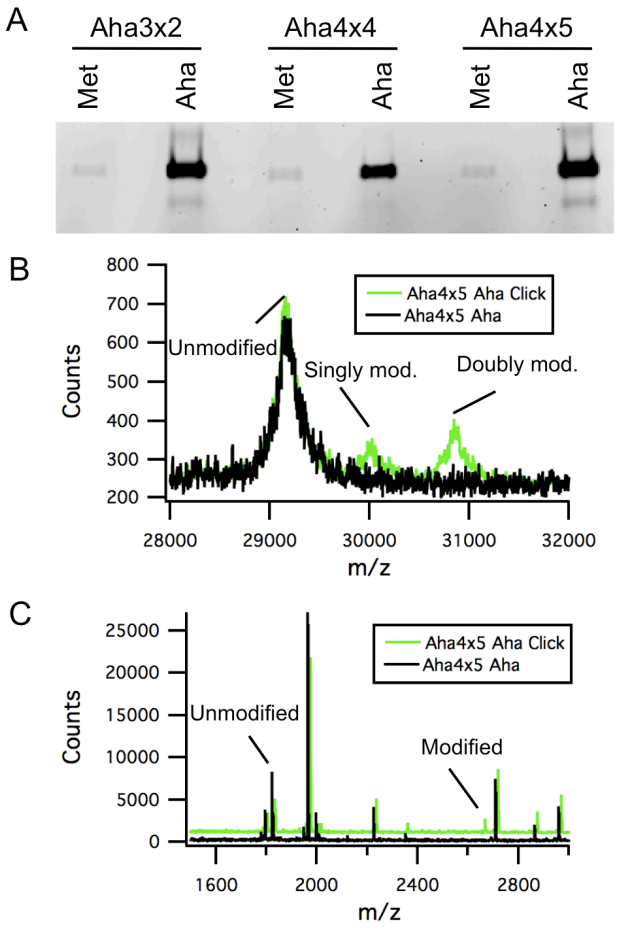


Figure 4.9. Strain-promoted click chemistry on scFvs using fluorescently labeled compound **6**. (*A*) SDS-PAGE on Met and Aha forms of scFv variants after reactions with **6**. (*B*) MALDI mass spectrometry on the Aha form of intact Aha4x5 before and after reaction with **6**. The ladders of signals 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).

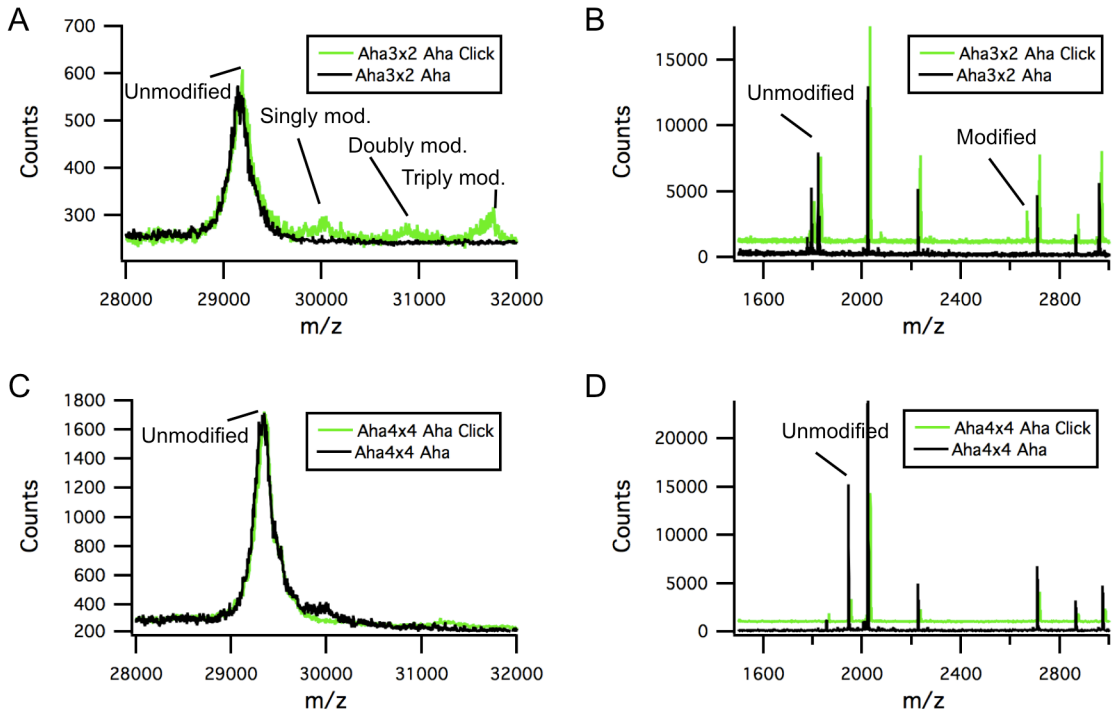


Figure 4.10. MALDI mass spectrometry on scFvs before and after strain-promoted click chemistry. (*A*) MALDI mass spectrometry on the Aha form of intact Aha3x2 before and after reaction with **6**. The ladders of signals detected are spaced apart by approximately 834 Da, corresponding to the molecular 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.

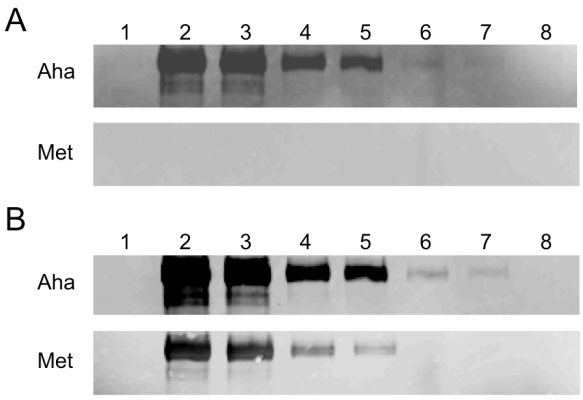


Figure 4.11. Western blotting using fluorescently labeled scFvs. (*A*) Probing nitrocellulose membranes for digoxigenin-labed BSA (BSA-Dig) using Aha and Met forms of Aha4x5 subjected to reaction with **6** and detection of Alexa-Fluor 488 fluorescence. Each lane of the SDS-PAGE gel transferred onto 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*).

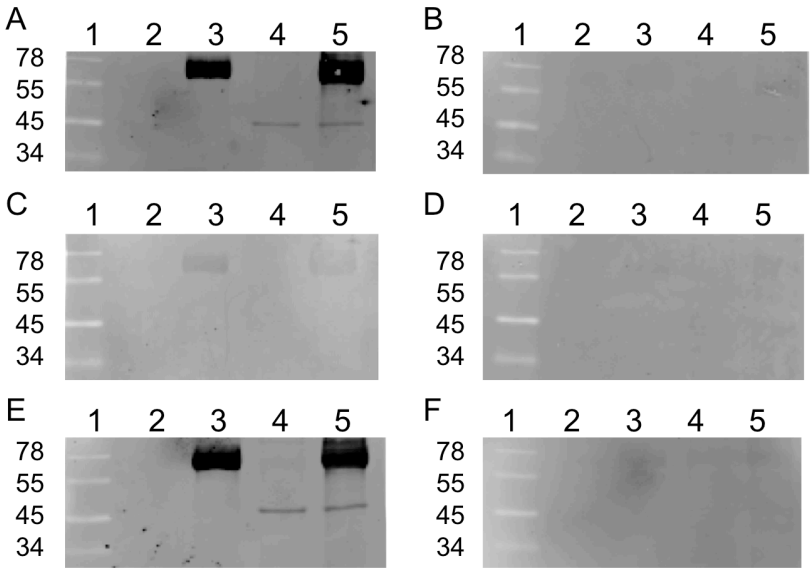


Figure 4.12. Western blotting using fluorescently labeled scFvs, part 2. 1000 ng of BSA or digoxigenin-labeled BSA (BSA-Dig) were run on SDS-PAGE gels, with 5.5 μg *E. coli* lysates added to protein samples as specified. Gels were transferred to nitrocellulose membranes, blocked, and probed for the presence 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.

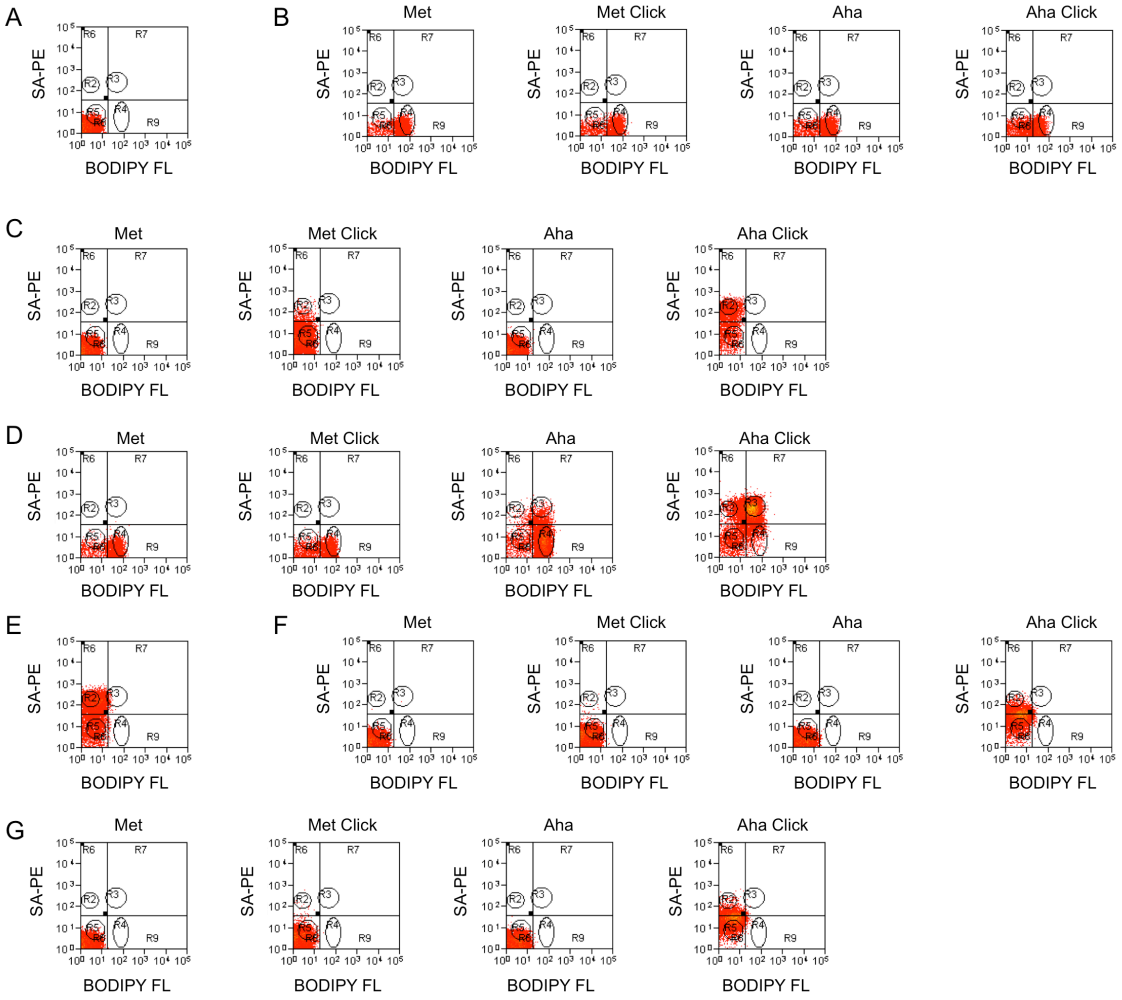


Figure 4.13. Flow cytometry of cell surface-displayed Aha4x5 to probe binding function and chemical modification with strained alkynes. The two-dimensional dot plots show the simultaneous measurements of the amount of **5** bound to cells (a measurement of binding function) and streptavidin-phycoerythrin (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.

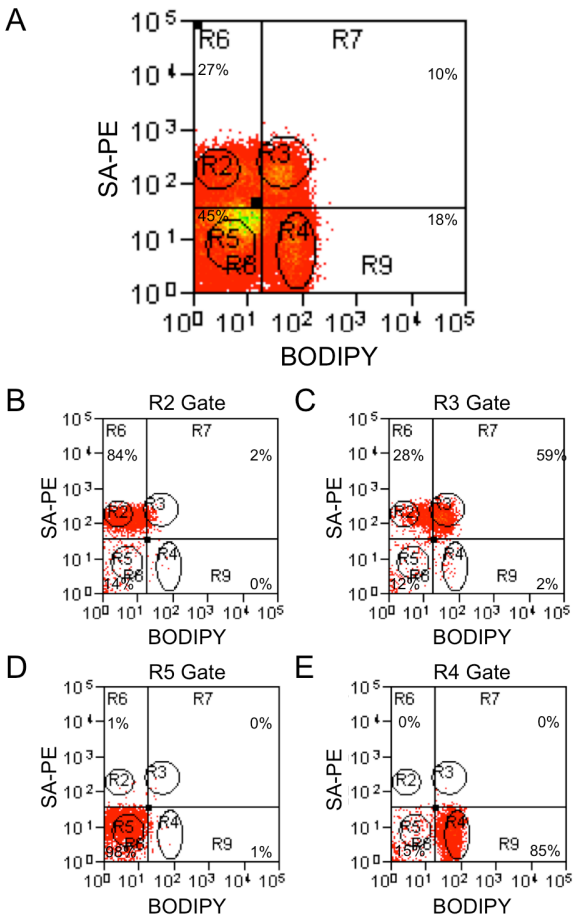


Figure 4.14. Fluorescence activated cell sorting for isolation of functional, modified proteins. (*A*) Two-dimensional dot plot of a mixture of four populations of cells. Each cellular population was reacted with **7** and exposed to streptavidin-phycoerythrin (SA-PE) and **5** for detection of chemical 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'